Fifty-first Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
October 2 - 3, 2013

Nonstationary Resource Sharing with Imperfect Binary Feedback: An Optimal Design Framework for Cost Minimization

Yuanzhang Xiao and Mihaela van der Schaar[1]

***Abstract*— We develop a novel design framework for decentralized resource sharing among self-interested users, who adjust their resource usage levels to minimize the costs of resource usage (e.g. energy consumption or payment) while fulfilling minimum payoff (e.g. throughput) requirements. We model the users' interaction as a *repeated resource sharing game with imperfect monitoring*, which captures the following features of the considered interaction. First, the users are decentralized and self-interested, i.e. they aim to minimize their own costs based on their locally available information and will not "blindly" follow the prescribed resource sharing rules unless it is in their self-interests to do so. Second, the users coexist in the system for some time and interact with each other *repeatedly*. Finally, the players receive a *binary* feedback informing them about the *imperfectly* measured interference/congestion level.**

The key feature of our proposed policy is that it is nonstationary, namely the users choose time-varying resource usage levels. This is in contrast with *all* existing policies, which are stationary and dictate users to choose constant resource usage levels. The proposed nonstationary policy is also deviation-proof, in that the self-interested users find it in their self-interests to comply with the policy, and it can be constructed by a low-complexity online algorithm that is run by each user in a distributed fashion. Moreover, our proposed policy only requires the users to have imperfect binary feedback, as opposed to existing solutions based on repeated game models which require a large amount of feedback. The proposed design framework applies to many resource sharing systems, such as power control, medium access control (MAC), and flow control. As a motivating example, we investigate the performance improvement of our nonstationary policy over state-of-the-art policies in power control, and show that significant performance gain (up to 90% energy saving) can be achieved.

I. Introduction

Resource sharing systems are ubiquitous. Examples of such systems range from classic resource sharing problems such as power control [1][7][12]–[15], medium access control [8], flow control [9]–[11], workload and task scheduling etc., to emerging new problems such as resource allocation in cloud data centers and demand-side management in smart grids.

In this paper, we model the resource sharing systems as repeated resource sharing games with imperfect monitoring, which characterize the following important features of such systems:

- *Decentralized and self-interested users.* The users are decentralized and self-interested. Hence, we model their interaction as a game.
- *Repeated interaction.* The users stay in the system for some time. Hence, we model their interaction using a repeated game rather than a one-shot game. As a result, unlike existing works which assume constant resource usage levels, in our model users may choose time-varying resource usage levels.
- *Imperfect monitoring.* The users can never perfectly observe the resource usage status in the system. In particular, we allow the users to have *very limited* monitoring of the resource usage status. Specifically, they only receive a *binary* feedback signal, which is quantized from the *erroneous* measurement on the interference/congestion level.

Some or all of the aforementioned features have been studied in numerous past works [1]–[14]. However, these works proposed *stationary* resource sharing policies, which dictate the users to choose constant resource usage levels over the time horizon in which they interact. In contrast, our previous works [21][22] are the first to propose *nonstationary* resource sharing policies, which allow the users to choose time-varying resource usage levels based on the history of past (imperfect) observations on the resource usage status. We have shown the significant performance improvement by using nonstationary policies (e.g. up to 200% improvement of spectrum efficiency in cognitive radio networks [22]).

However, our previous works [21][22] formulated and solved the *payoff maximization* problem, in which each user aims to maximize its long-term payoff (e.g. throughput). In this paper, we consider the *cost minimization* problem, in which each user aims to minimize its long-term cost (e.g. energy consumption or payment for resource usage) subject to the minimum requirement for the achieved long-term payoff. The cost minimization problem is much harder to solve, because 1) we need to consider both payoffs and costs, and 2) the users are coupled in the constraint domain (i.e. the minimum payoff requirements) instead of the objective function domain as in the payoff maximization problem. We will describe the challenges of the cost minimization problem compared to the payoff maximization problem in more details in Sec. II-C. But we would like to briefly illustrate the difficulty of cost minimization problems under simpler scenarios. When the users are obedient and use stationary policies, the network utility maximization (NUM) framework [23] commonly used for payoff maximization problems does not apply here, because the constraints on the payoffs are coupled in such a complicated way, that the problem cannot be decomposed into uncoupled subproblems by using dual

[1]The authors are with the Electrical Engineering Department, UCLA, Los Angeles, CA 90095, USA. Email: {yxiao,mihaela}@ee.ucla.edu

TABLE I

COMPARISONS AGAINST EXISTING WORKS.

	Formulation	Nonstationary	Feedback (Overhead)	Hetero. Selfish users
[1]–[7]	Min. Cost	No	Error-free, unquant. (Large)	Yes
[8]–[14]	Max. Payoff	No	Error-free, unquant. (Large)	Yes
[15]	Max. Payoff	Yes	Error-free, unquant. (Large)	Yes
[16]	Max. Payoff	Yes	Erroneous, limited (Medium)	Yes
[17][18]	Min. Cost	Yes	Erroneous, binary (One-bit)	No
[19][20]	Max. Payoff	Yes	Erroneous, binary (One-bit)	No
[21][22]	Max. Payoff	Yes	Erroneous, binary (One-bit)	Yes
This work	Min. Cost	Yes	Erroneous, binary (One-bit)	Yes

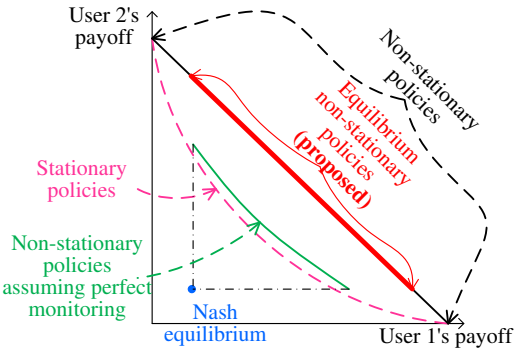

Fig. 1. An illustration of the operating points achievable by different policies.

decomposition. When the users are self-interested and use stationary policies, the Nash equilibrium (NE) is not the correct solution concept any more. Due to the coupled constraints, the equilibrium is defined as the generalized NE (GNE) [5]. Even establishing the existence of GNE is a non-trivial task. We can imagine that designing nonstationary policies in cost minimization problems should be much more challenging than designing stationary policies.

In summary, we propose in this paper a novel design framework for cost minimization in resource sharing among decentralized users. We consider the problems in which each player aims to minimize the cost associated with resource usage while fulfilling its minimum payoff requirement. We propose nonstationary resource sharing policies, which significantly outperform existing stationary policies. Although the nonstationary policies themselves are complicated, we are able to come up with a low-complexity online algorithm, which can be run by each user in a decentralized manner, to construct the policy. In other words, while the design of the nonstationary policies is mathematically complicated, their implementation is, fortunately, of low complexity. In fact, our proposed policy is significantly easier to implement at run-time than well-known stationary policies such as those in [1]–[7]. Moreover, the proposed policy only requires the users to have imperfect and binary feedback on the resource usage status. Note that our design framework can be easily adapted to the case of obedient users, which will be discussed in Sec. IV-D.

In the rest of this paper, we will first compare our work with existing works in Sec. II. Then we describe the system model and formulate the design problem in Sec. III. We solve the design problem in Sec. IV and demonstrate the performance improvement through simulations in Sec. V. Finally, Sec. VI concludes the paper.

II. RELATED WORKS

We summarize the major differences between the existing works and our work in Table I. Detailed explanations are as follows.

A. Stationary Policies

Most existing works [1]–[14] focused on *stationary* policies, which restrict the users to consume resources at *constant* levels over the time horizon in which they interact[1]. Our proposed nonstationary policies significantly outperform the stationary policies proposed in the existing works [1]–[14], especially in the presence of strong interference/congestion. This can be illustrated in Fig. 1 in a two-user case. We can see that due to strong interference/congestion, the set of payoffs achievable by stationary policies is *not convex*, and each payoff in it is Pareto dominated by some payoffs achievable only by nonstationary policies.

Note that we list [12]–[14] in the category of stationary policies, although they design policies in a repeated game framework. This is because in the equilibrium where the system operates, the policies in [12]–[14] use constant resource usage levels. The repeated game formulation is used only to provide incentives for self-interested users to cooperate.

B. Nonstationary Policies

1) Nonstationary Policies Based on Repeated Games: The major limitation of the works based on repeated games with perfect monitoring [15] is the assumption of perfect monitoring, which requires error-free and unquantized feedback of the resource usage status. The theory of repeated games with imperfect monitoring [16] allows erroneous and limited feedback, but requires that the amount of feedback increases with the number of resource usage levels that the users can choose. In contrast, we only require binary feedback regardless of the number of resource usage levels, which significantly reduces the feedback overhead.

2) Nonstationary Policies Based on Constrained MDP: The theory of constrained Markov decision processes (MDP) or constrained partially-observable MDP (POMDP) [17] has been used to solve resource sharing problems (see representative works [18]). However, most of the approaches based on constrained MDP or POMDP solve only *single*-user decision problems (or multi-user problems with *homogeneous* users), and cannot be easily extended to the case where *multiple heterogeneous* users compete for a

[1]Although some resource sharing policies [1]–[10] go through a transient period of adjusting the resource usage levels before the convergence to the optimal resource usage levels, the players maintain constant resource usage levels after the convergence.

TABLE II

DIFFERENCES BETWEEN THE DESIGN FRAMEWORK FOR PAYOFF MAXIMIZATION AND THAT FOR COST MINIMIZATION.

	Payoff Maximization	Cost Minimization
Optimization	Resource usage schedule only	Resource usage schedule and resource usage levels
Feasible operating points	Part of a hyperplane	Part of each one of infinite number of hyperplanes

single resource. In addition, they assume that the users are cooperative.

3) Nonstationary Policies Based on Multi-arm Bandit: Nonstationary policies based on multi-arm bandit (MAB) have been proposed in [19]–[20]. First, [19]–[20] focused on the payoff (i.e. throughput) maximization problem, while our work studies the cost minimization problems. In addition, [19][20] assumed that the users are *homogeneous* and *cooperative*, while our work considers heterogeneous and self-interested users.

C. Comparison With Our Previous Works

Most related to this work are our previous works [21][22]. However, the design frameworks proposed in [21][22] and in this work are significantly different because the design objectives are different. In [21][22], we aimed to design TDMA (time-division multiple access) resource sharing policies that maximize the users' sum payoff without considering cost minimization. In TDMA policies, only one user is active and consumes resources in each time slot. To maximize its payoff, each user will choose the maximum resource usage level when it is active. Hence, what we optimized was *only the resource usage schedule of the users*. In this work, since we aim to minimize the cost subject to the minimum payoff requirements, and there may be many resource usage levels that fulfill the payoff requirement, we also need to optimize *the users' resource usage levels in addition to the resource usage schedule*, which makes the design problem more challenging.

Next, we explain the differences in the design frameworks in detail. Both design frameworks include three steps: characterization of the set of feasible operating points, selection of the optimal operating point, and the distributed implementation of the policy. The fundamental difference is in the first step, which is the most important step in the design. In [21][22], since each user chooses the maximum resource usage level when it is active, we know that the set of feasible operating points lies in the hyperplane determined by each user's maximum achievable payoff. Hence, we only need to determine which portion of this particular hyperplane is achievable. On the contrary, in this work, since the users may not choose the maximum resource usage levels when active, the feasible operating points lie in *infinite number of hyperplanes*, each of which goes through the vector of minimum payoff requirements (see Fig. 2 for illustration). Hence, it is more difficult to characterize the set of feasible operating points in this work. Due to the more complicated characterization of the feasible operating points, the selection of the optimal operating point (the second step) also becomes a more complicated optimization problem in this work (although we can prove that it can be converted to a convex optimization problem under reasonable assumptions). In summary, in this work, the first two steps in the design framework are fundamentally different from those in [21][22], and are more challenging. We summarize the major differences between the design frameworks for payoff maximization and for cost minimization in Table II.

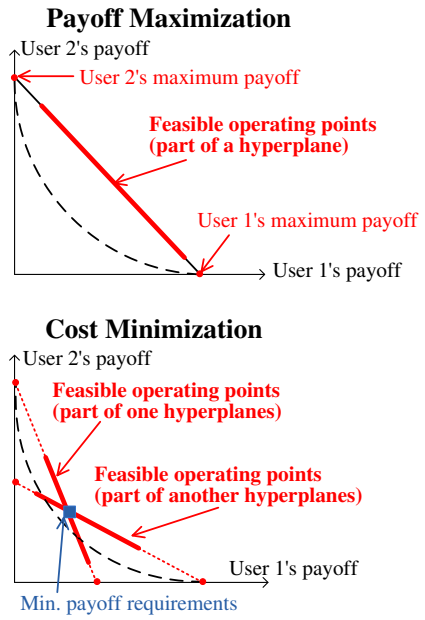

Fig. 2. Illustration of feasible operating points in the design framework for the payoff maximization problem and that for the cost minimization problem.

III. SYSTEM MODEL

The Stage Game: Consider a system with N players sharing a common resource. Denote the set of the players by $\mathcal{N} \triangleq \{1, 2, \ldots, N\}$. Each player i chooses its action a_i (i.e., its resource usage level) from its action set $A_i \subset \mathbb{R}_+$. The joint action profile of all the players is denoted by $\mathbf{a} = (a_1, \ldots, a_N) \in \mathcal{A} \triangleq \times_{i \in \mathcal{N}} A_i$, and the action profile of all the players other than player i is denoted by $\mathbf{a}_{-i}$. Given the joint action profile $\mathbf{a}$, each player i receives a payoff $u_i(\mathbf{a})$, where $u_i : \mathcal{A} \to \mathbb{R}_+$ is player i's utility function. The interaction among the players is characterized by the game tuple $\mathcal{G} = \langle \mathcal{N}, \{A_i\}_{i \in \mathcal{N}}, \{u_i\}_{i \in \mathcal{N}} \rangle$. We define the resource sharing game $\mathcal{G}$ as follows.

Definition 1: $\mathcal{G}$ is a resource sharing game, if 1) each player i's action set A_i is compact and includes 0 as an element (i.e. $0 \in A_i$); 2) each player i's utility function u_i is decreasing in player j's action a_j, $\forall j \neq i$, and is 0 when $a_i = 0$, $\forall \mathbf{a}_{-i}$; 3) the set of feasible payoffs $\mathcal{V} = \{\mathbf{u}(\mathbf{a}) = (u_1(\mathbf{a}), \ldots, u_N(\mathbf{a})) : \mathbf{a} \in \mathcal{A}\}$ has $N + 1$ extremal points[2]: $(0, \ldots, 0) \in \mathbb{R}^N$, $\mathbf{u}(\tilde{\mathbf{a}}^1), \ldots, \mathbf{u}(\tilde{\mathbf{a}}^N)$.

[2]The extremal points of a convex set are those that are not convex combinations of other points in the set.

The above definition captures the main characteristics of a resource sharing system. The first property on the action sets indicates that each user's action set is closed and bounded, which is consistent with the users' inability to consume unlimited amount of resources. In addition, each user can choose not to use any resource by taking action 0. The second property on the utility function reflects the interference/congestion among the players: increasing resource usage by one user results in a decrease in the other users' payoffs. Moreover, when a user does not use the resource by choosing action 0, it will receive zero payoff. The third property captures the strong interference/congestion among the users: the increase of one player's payoff comes at such an expense of the other players' payoffs that the set of feasible payoffs is *nonconvex*. We are particularly interested in the scenarios where the interference/congestion is strong, because efficient resource sharing policies are essential to mitigate the strong interference/congestion. On the contrary, if the interference/congestion among users is weak, efficient resource sharing policies are less important, because the users can just choose their optimal resource usage levels individually without affecting the other users. We illustrate the payoffs in a two-user resource sharing game in Fig. 1.

***The Repeated Game**:* Since players interact with each other repeatedly, we model their interaction as a repeated game with the stage game $\mathcal{G}$ played in every period $t = 0, 1, 2, \ldots$. At the beginning of each period t, players choose the action profile $\mathbf{a}^t$, leading to an interference/congestion level $f(\mathbf{a}^t)$. We assume that $f(\mathbf{a})$ is increasing in a_i, $\forall i \in \mathcal{N}$. The interference/congestion level is observed by the players with errors. We denote the noisy observation at period t by z^t, which is characterized by the conditional probability density function $\eta(z^t|f(\mathbf{a}^t))$. The players also quantize the noisy observation for limited feedback. We denote the feedback signal at the end of period t by y^t. In this paper, we will focus on binary feedback, which has minimal overhead. The noisy observation z^t is quantized as 0 if it is below a threshold z_0, and quantized as 1 otherwise. Then the conditional probability distribution of the feedback signal is $\rho(y^t = 0|\mathbf{a}^t) = \int_{z^t < z_0} \eta(z^t|f(\mathbf{a}^t))dz^t$ and $\rho(y^t = 1|\mathbf{a}^t) = 1 - \rho(y^t = 0|\mathbf{a}^t)$.

A resource sharing policy specifies the action profile $\mathbf{a}^t$ the players choose at each period t based on the past history. The history up to period t is the collection of past feedback signals, namely $h^t = \{y^0, \ldots, y^{t-1}\}$ for $t \geq 1$ and $h^0 = \varnothing$. Hence, player i's policy π_i is a mapping from the set of all possible histories $\mathcal{H} \triangleq \sqcup_{t=0}^{\infty}\{0,1\}^t$ to its action set A_i. We define the resource sharing policy as the joint policy profile $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_N)$.

We classify all the spectrum sharing policies into two categories, stationary and nonstationary policies. A spectrum sharing policy $\boldsymbol{\pi}$ is *stationary* if and only if for all $i \in \mathcal{N}$, for all $t \geq 0$, and for all $h^t \in \{0,1\}^t$, we have $\pi_i(h^t) = a_i^{\text{stat}}$, where $a_i^{\text{stat}} \in A_i$ is a constant. A spectrum sharing policy is *nonstationary* if it is not stationary. In this paper, we restrict our attention to a special class of nonstationary polices, namely TDMA policies (with constant resource usage levels). A spectrum sharing policy $\boldsymbol{\pi}$ is a TDMA policy if at most one user consumes resources in each time slot, and each user i chooses the same resource usage level $a_i^{\text{TDMA}} \in A_i$ when it is active.

***Costs, (Long-term) Payoffs, and Equilibrium**:* Each user i has a cost $c_i(a_i)$ in choosing the resource usage level a_i. The cost can model, for example, the payment or energy consumption associated with using the resources.

Each user i's long-term payoff is defined as the expected discounted average payoff. Assuming, as in [12]–[16], the same discount factor $\delta \in [0, 1)$ for all the players, player i's payoff can be written as

$$U_i(\boldsymbol{\pi}) = \mathbb{E}_{h^0,h^1,\ldots}\left\{(1-\delta)\left[u_i(\boldsymbol{\pi}(h^0)) + \sum_{t=1}^{\infty}\delta^t u_i(\boldsymbol{\pi}(h^t))\right]\right\}.$$

Similarly, we can define player i's long-term cost as

$$C_i(\boldsymbol{\pi}) = \mathbb{E}_{h^0,h^1,\ldots}\left\{(1-\delta)\left[c_i(\pi_i(h^0)) + \sum_{t=1}^{\infty}\delta^t c_i(\pi_i(h^t))\right]\right\}.$$

Each user i aims to minimize its long-term cost $C_i(\boldsymbol{\pi})$ while fulfilling a minimum payoff requirement $U_i^{\min}$. From one user's perspective, it has the incentive to deviate from a given resource sharing policy, if by doing so it can fulfill the minimum payoff requirement with a lower cost. Hence, we can define the equilibrium as follows.

Definition 2: A resource sharing policy $\boldsymbol{\pi}$ is an equilibrium if for all $i \in \mathcal{N}$, we have

$$\pi_i = \arg\min_{\pi_i'} C_i(\pi_i', \boldsymbol{\pi}_{-i}), \text{ subject to } U_i(\pi_i', \boldsymbol{\pi}_{-i}) \geq U_i^{\min},$$

where $\boldsymbol{\pi}_{-i}$ is the joint policy of all the users except user i.

IV. The Design Framework

The goal of the designer is to minimize certain cost criterion while fulfilling all the users' minimum payoff requirements at the equilibrium. The cost criterion can be represented by a function $W(C_1(\boldsymbol{\pi}), \ldots, C_N(\boldsymbol{\pi}))$. An example cost criterion can be the weighted sum of all the users' costs, i.e. $W = \sum_{i\in\mathcal{N}} w_i \cdot C_i(\boldsymbol{\pi})$ with $w_i \geq 0$ and $\sum_{i\in\mathcal{N}} w_i = 1$. We can formally define the policy design problem as

$$\begin{aligned} \min_{\boldsymbol{\pi}} \quad & W(C_1(\boldsymbol{\pi}), \ldots, C_N(\boldsymbol{\pi})) \\ s.t. \quad & \boldsymbol{\pi} \text{ is an equilibrium,} \\ & U_i(\boldsymbol{\pi}) \geq U_i^{\min}, \ \forall i \in \mathcal{N}. \end{aligned} \tag{1}$$

We outline the proposed design framework to solve the policy design problem (illustrated in Fig. 3), which consists of three steps. First, we characterize the set of feasible operating points that can be achieved at equilibria. Then, given this set, we select the optimal operating point based on the cost criterion. Finally, we construct the optimal resource sharing policy that achieves the optimal operating point.

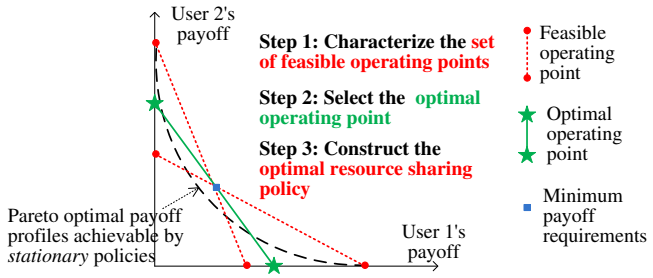

Fig. 3. The design framework.

A. Characterize The Set of Feasible Operating Points

The first step in solving the design problem (1) is to quantify the set of feasible operating points that can be achieved at equilibria. We define an operating point as $\bar{\mathbf{u}} = (\bar{u}_1, \ldots, \bar{u}_N)$, which is a collection of each user i's instantaneous payoff $\bar{u}_i$ when user i is the only one that uses the resources. An operating point $\bar{\mathbf{u}}$ is *feasible*, if there exists an equilibrium $\boldsymbol{\pi}$ such that 1) $u_i(\boldsymbol{\pi}(h^t)) = \bar{u}_i$ for all h^t such that $\pi_i(h^t) > 0$; 2) each user i achieves its minimum payoff requirement, i.e. $U_i(\boldsymbol{\pi}) = U_i^{\min}$.

Before quantifying the set of feasible operating points, we define $b_{ij}(\bar{u}_i)$ as

$$b_{ij}(\bar{u}_i) = \sup_{a_j \in A_j, a_j > 0} \frac{\rho(y=1|\tilde{\mathbf{a}}^i(\bar{u}_i)) - \rho(y=1|a_j, \tilde{\mathbf{a}}^i_{-j}(\bar{u}_i))}{u_j(a_j, \tilde{\mathbf{a}}^i_{-j}(\bar{u}_i))/\bar{u}_j},$$

where $\tilde{\mathbf{a}}^i(\bar{u}_i)$ is the action profile that satisfies $\tilde{a}^i_j = 0, \forall j \neq i$ and $u_i(\tilde{\mathbf{a}}^i) = \bar{u}_i$.

Now we state Theorem 1, which characterizes the set of feasible operating points.

Theorem 1: An operating point $\bar{\mathbf{u}}$ is feasible, if

- Condition 1: the discount factor δ satisfies $\delta \geq \underline{\delta} \triangleq 1/\left(1 + \frac{1-\sum_{i\in\mathcal{N}} \underline{\mu}_i}{N-1+\sum_{i\in\mathcal{N}}\sum_{j\neq i}(-\rho(y=1|\tilde{\mathbf{a}}^i(\bar{u}_i))/b_{ij}(\bar{u}_i))}\right)$, where $\underline{\mu}_i \triangleq \max_{j\neq i} \frac{1-\rho(y=1|\tilde{\mathbf{a}}^i(\bar{u}_i))}{-b_{ij}(\bar{u}_i)}$.
- Condition 2: $\sum_{i\in\mathcal{N}} U_i^{\min}/\bar{u}_i = 1$, and $\bar{u}_i \leq U_i^{\min}/\underline{\mu}_i$.

Proof: See [24, Appendix A]. ■

B. Select The Optimal Operating Point

Given the set of feasible points obtained in Theorem 1, we need to select the optimal operating point $\bar{\mathbf{u}}^\star$ based on the cost criterion W. The following proposition formulates the problem of finding the optimal operating point.

Proposition 1: The optimal operating point $\bar{\mathbf{u}}^\star$ can be solved by the following optimization problem

$$\bar{\mathbf{u}}^\star = \arg\min_{\bar{\mathbf{u}}} \quad W(\bar{C}_1(\bar{u}_i), \ldots, \bar{C}_N(\bar{u}_i)), \qquad (2)$$
$$\text{subject to} \quad \sum_{i\in\mathcal{N}} U_i^{\min}/\bar{u}_i = 1, \ \bar{u}_i \leq U_i^{\min}/\underline{\mu}_i,$$

where $\bar{C}_i(\bar{u}_i) = \frac{U_i^{\min}}{\bar{u}_i} \cdot \tilde{a}^i_i(\bar{u}_i)$. In particular, if $W(\bar{C}_1, \ldots, \bar{C}_N)$ is jointly convex in $\bar{C}_1, \ldots, \bar{C}_N$, the optimization problem (2) is convex.

Proof: See [24, Appendix B]. ■

TABLE III

THE ALGORITHM RUN BY EACH USER i.

Require: $\{U_j^{\min}/\bar{u}_j^\star\}_{j\in\mathcal{N}}$ and $\bar{u}_i^\star$

Initialization: Sets $t = 0$, $u'_j(0) = U_j^{\min}/\bar{u}_j^\star$ for all $j \in \mathcal{N}$.

repeat

Calculates $d_j(t) = \frac{u'_j(t) - \underline{\mu}_j}{1 - u'_j(t) + \sum_{k\neq j}(-\rho(y=1|\tilde{\mathbf{a}}^j)/b_{jk})}, \forall j$

Finds $i^* \triangleq \arg\max_{j\in\mathcal{N}} d_j(t)$

if $i = i^*$ **then**

Chooses resource usage level $\tilde{a}^{i^*}_{i^*}(\bar{u}_i^\star)$

else

Does not consume resources (i.e. chooses action 0)

end if

Updates $u'_j(t+1)$ for all $j \in \mathcal{N}$

if $y^t = 0$ **then**

$u'_{i^*}(t+1) = \frac{1}{\delta} u'_{i^*}(t) - (\frac{1}{\delta} - 1) \cdot (1 + \sum_{j\neq i^*} \frac{\rho(y=1|\tilde{\mathbf{a}}^{i^*})}{-b_{i^*j}})$

$u'_j(t+1) = \frac{1}{\delta} u'_j(t) + (\frac{1}{\delta} - 1) \cdot \frac{\rho(y=1|\tilde{\mathbf{a}}^{i^*})}{-b_{i^*j}}, \forall j \neq i^*$

else

$u'_{i^*}(t+1) = \frac{1}{\delta} u'_{i^*}(t) - (\frac{1}{\delta} - 1) \cdot (1 - \sum_{j\neq i^*} \frac{\rho(y=0|\tilde{\mathbf{a}}^{i^*})}{-b_{i^*j}})$

$u'_j(t+1) = \frac{1}{\delta} u'_j(t) - (\frac{1}{\delta} - 1) \cdot \frac{\rho(y=0|\tilde{\mathbf{a}}^{i^*})}{-b_{i^*j}}, \forall j \neq i^*$

end if

$t \leftarrow t + 1$

until $\varnothing$

TABLE IV

COMPARISON OF DESIGN FRAMEWORKS FOR SELF-INTERESTED AND OBEDIENT USERS.

	Conditions	Boundary	Algorithm
Obedient	$\underline{\delta} = \frac{N-1}{N}$	$\underline{\mu}_i = 0, \forall i$	$b_{ij} = -\infty, \forall i, j$
Self-interested	$\underline{\delta} > \frac{N-1}{N}$	$\underline{\mu}_i > 0, \forall i$	$b_{ij} \in (-\infty, 0), \forall i, j$

C. Construct The Optimal Policy

Given the optimal operating point obtained in the second step, each user i runs the algorithm in Table III in a decentralized manner, and achieves its minimum payoff requirements. The resulting policy is an equilibrium, in that if a user does not follow the algorithm, it will either achieve a lower payoff or achieve the same payoff with a higher cost.

The optimal policy can be viewed as implementing a simple "largest-distance-first" scheduling, namely the user farthest away from the optimal operating point (i.e. has the largest $d_j(t)$) occupies the resources by itself. However, it is not trivial to define the "distance" from the optimal operating point $d_j(t)$. We can prove that the distance $d_j(t)$ used in our proposed algorithm is the correct one.

Theorem 2: If each user $i \in \mathcal{N}$ runs the algorithm in Table III, then each user i can achieve its minimum payoff requirement $U_i^{\min}$ with a cost $\bar{C}_i$ that minimizes the cost criterion $W(\bar{C}_1, \ldots, \bar{C}_N)$. The policy implemented by the algorithm is an equilibrium: if a user does not follow the algorithm, it will either fail to achieve its minimum payoff requirement, or achieve it with a higher cost.

Proof: See [24, Appendix C]. ■

D. Obedient Users

Obedient users will follow any resource sharing policy as long as their minimum payoff requirements are achieved.

Without requiring the policy to be deviation-proof, the design framework can be greatly simplified. We summarize the differences in the design frameworks for self-interested users and obedient users in Table IV.

First, the sufficient conditions for feasible operating points are relaxed (i.e. a smaller critical discount factor $\underline{\delta}$). Second, the boundaries of the feasible operating points $\underline{\mu}_i$ become zero. In other words, the operating points $\bar{u}_i$ can be arbitrarily large. Third, in the algorithm to compute the resource sharing policy, since $b_{ij} = -\infty$, the terms related to b_{ij} vanish, which makes the algorithm simpler.

E. Implementation

Our proposed design framework can be implemented in two phases: an initial information exchange phase in which the optimal operating point is calculated, followed by a decentralized implementation phase in which users run the algorithm in Table III in a decentralized manner. In the following, we first specify what information needs to be exchanged in the initial information exchange phase. Then we show that the total overhead of initial information exchange and feedback in the proposed framework is much smaller than those in existing works.

1) Overhead of initial information exchange and feedback: To better illustrate the overhead, we compare the overhead of initial information exchange and feedback in the proposed framework with that in [1]–[7], which proposed energy-efficient stationary power control algorithms. The comparison is summarized in Table V. In the initial information exchange phase, the proposed framework has an additional overhead of $N^2 + N$ compared to [1]–[7]. This additional overhead mainly comes from the information exchange of b_{ij}, which is used for deviation-proofness[3]. However, in the run time, the feedback overhead of the proposed policy is significantly lower than that of [1]–[7]. Specifically, in [1]–[7], each user's receiver needs to feedback the interference temperature in *each time slot*, which has an overhead of N *real numbers* per time slot. In contrast, in our proposed framework, each user only needs to feedback a one-bit binary signal. This feedback can be further reduced: they can send a distress signal (possibly just a probe signal) only when the feedback signal is $y = 1$.

2) Computational complexity: As we can see from Table III, the computational complexity of each user in constructing the optimal policy is very small. In each time slot, each user only needs to compute N distances $\{d_j(t)\}_{j\in\mathcal{N}}$, and N normalized values $\{u'_j(t)\}_{j\in\mathcal{N}}$, all of which are determined by analytical expressions. In addition, although the original definition of the policy requires each user to memorize the entire history of feedback signals, in the actual implementation, each user only needs to know the current feedback signal y^t and memorize N normalized values $\{u'_j(t)\}_{j\in\mathcal{N}}$.

[3]For obedient users, this overhead can be avoided since there is no need to broadcast b_{ij}.

TABLE V
COMPARISON OF THE TOTAL OVERHEAD OF INITIAL INFORMATION EXCHANGE AND FEEDBACK.

	Overhead
[1]–[7]	Initial information exchange: N/A Feedback: Each user i feeds the interference temperature back in *each time slot* Total overhead: N *real numbers* in *each time slot*
Proposed (self-interested)	Initial information exchange: Each user i broadcasts to all the other users: $\rho(y=1\|\tilde{\mathbf{a}}^i)$, $U_i^{\min}$, and $\{b_{ji}\}_{j\neq i}$ Feedback: a binary feedback signal Total overhead: N^2+N real numbers initially, and a distress signal (possibly just a probe) when necessary
Proposed (obedient)	Initial information exchange: Each user i broadcasts to all the other users: $U_i^{\min}$ Feedback: a binary feedback signal Total overhead: N real numbers initially, and a distress signal (possibly just a probe) when necessary

TABLE VI
COMPUTATIONAL COMPLEXITY OF THE PROPOSED FRAMEWORK.

	Computation	Memory
Initial	Solve the convex program (2)	Needed for solving (2)
Run-time	N indices analytically	N indices

V. SIMULATION RESULTS

In this section, we apply our design framework to spectrum sharing scenarios, in which multiple users share the spectrum. Their actions are the transmit power levels, their payoffs are the throughput, and their costs are the energy consumption. We use the following system parameters. The noise powers at all the users' receivers are 0.05 W. For simplicity, we assume that direct channel gains have the same distribution $g_{ii} \sim \mathcal{CN}(0,1), \forall i$, and the cross channel gains have the same distribution $g_{ij} \sim \mathcal{CN}(0,\alpha), \forall i \neq j$, where α is defined as the *cross interference level*. Each user measures the interference temperature with a measurement error ε that is Gaussian distributed with zeros mean and variance 0.1. The energy efficiency criterion is the average energy consumption across the users. The discount factor is 0.95.

We compare the proposed policy against the optimal stationary policy in [1]–[7] and two (adapted) versions of the punish-forgive (PF) policies in [12]–[15]. Since the PF policies in [12]–[15] were originally proposed for the throughput maximization problem, we need to adapt them to solve the cost minimization problem in (1). We describe the state-of-the-art policies that we compare against as follows.

- The optimal stationary policy [1]–[7]: each user transmits at a constant power level that is just large enough to fulfill the throughput requirement under the interference from other users.
- The (adapted) stationary punish-forgive (SPF) policy [12]–[14]: the SPF policies are dynamic policies that have two phases. When the users have not received the distress signal (defined as the signal $y = 1$), they transmit at optimal *stationary* power levels. When they receive a distress signal that indicates deviation, they switch to the punishment phase, in which all the users

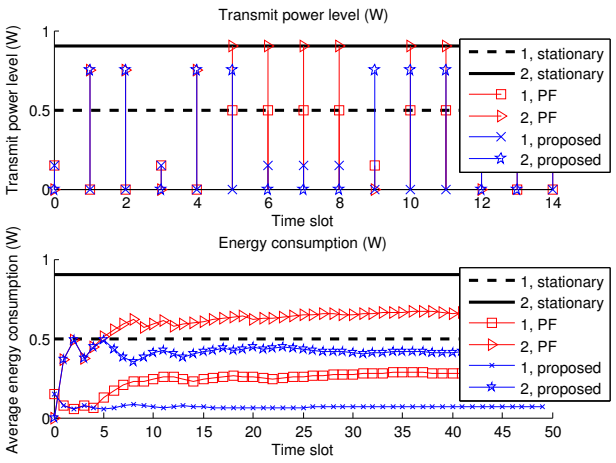

Fig. 4. Illustration of different policies. Upper plot: in the stationary policy, the users transmit at constant power levels (two flat black lines); in PF and proposed policies, only one user transmits in each time slot; in the PF policy, the users choose the same power levels as in the stationary policy after time slot 4 when $y = 1$ received. Lower plot: the two users' average energy consumptions under different policies.

transmit at the Nash equilibrium power levels. In the energy efficiency formulation, the optimal stationary power levels are the Nash equilibrium power levels. Hence, the adapted SPF policy is essentially the same as the optimal stationary policy.

- The adapted nonstationary punish-forgive (NPF) policy: the punish-forgive policy in [15] is different from those in [12]–[14], in that *nonstationary* power levels are used when the users have not received the distress signal. In the simulation, we adapt the NPF policy in [15] such that the users transmit in the same way as in the proposed policy when they have not received the distress signal. After receiving the distress signal, the NPF policy requires the users to transmit at the optimal stationary power levels.

Since the SPF policy is the same as the optimal stationary policy, in the rest of this section, we focus on the NPF policy, and simply refer to the NPF policy as the PF policy.

A. Illustrations of Different Policies

Fig. 4 illustrates the differences among stationary, PF, and the proposed policies in a simple case of two users, whose minimum throughput requirements are 1 bits/s/Hz and 2 bits/s/Hz, respectively. In stationary policies, users transmit simultaneously with constant power levels (0.5 W and 0.9 W), which are higher than those (0.15 W and 0.75 W) in the proposed policy, because users need to overcome multi-user interference to achieve the minimum throughput requirements. In addition, users transmit all the time in stationary polices, which results in even higher average energy consumption.

The key difference between the proposed policy and the PF policy lies in time slot 5, after a distress signal is sent at $t = 4$. In the PF policy, users transmit together at the same high power levels as in the stationary policy at $t = 5$. In the proposed policy, user 2, the user who transmitted at $t = 4$, transmits again at $t = 5$. In summary, the punishment in the PF policy is the multi-user interference, which increases the energy consumptions of both users, while the punishment in the proposed policy is the delay in transmission, which keeps the energy consumptions low. This advantage of the proposed policy in terms of energy efficiency is also illustrated in Fig. 4.

Finally, we can see that in the steady state, the energy consumption of the proposed policy is much lower than those in the other policies.

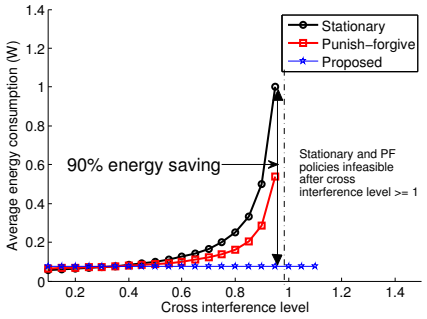

Fig. 5. Energy efficiency of the stationary, PF, and proposed policies under different cross interference levels.

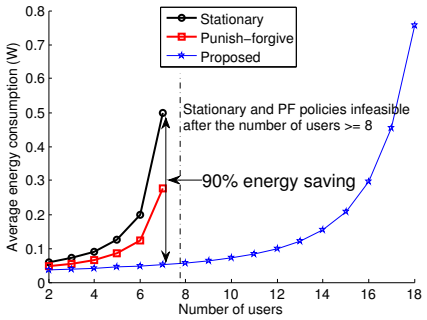

Fig. 6. Energy efficiency of the stationary, PF, and proposed policies under different number of users.

B. Performance Gains

We compare the energy efficiency of the optimal stationary policy, the optimal punish-forgive policy, and the proposed policy under different cross interference levels in Fig. 5. We consider a network of two users whose minimum throughput requirements are 1 bits/s/Hz. First, notice that the energy efficiency of the proposed policy remains constant under different cross interference levels, while the average transmit power increases with the cross interference level in the other two policies. The proposed policy outperforms the other two policies in medium to high cross interference levels (approximately when $\alpha \geq 0.3$). In the cases of high cross interference levels ($\alpha \geq 1$), there is no stationary policy that can fulfill the minimum throughput requirements. As a consequence, the punish-forgive policies cannot fulfill the throughput requirements when $\alpha \geq 1$, either.

In Fig. 6, we examine how the performance of these three policies scales with the number of users. The number of users

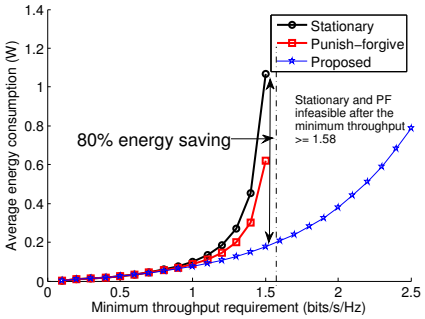

Fig. 7. Energy efficiency of the stationary, PF, and proposed policies under different minimum throughput requirements.

in the network increases, while the minimum throughput requirement for each user remains 1 bits/s/Hz. The cross interference level is $\alpha = 0.2$. We can see that the stationary and punish-forgive policies are infeasible when there are more than 6 users. In contrast, the proposed policy can accommodate 18 users in the network with each users transmitting at a power level less than 0.8 W.

Fig. 7 shows the joint spectrum and energy efficiency of the three policies. We can see that the optimal stationary and punish-forgive polices are infeasible when the minimum throughput requirement is larger than 1.6 bits/s/Hz. On the other hand, the proposed policy can achieve a much higher spectrum efficiency (2.5 bits/s/Hz) with a better energy efficiency (0.8 W transmit power). Under the same average transmit power, the proposed policy is always more energy efficient than the other two policies.

In summary, the proposed policy significantly improves the spectrum and energy efficiency of existing policies in most scenarios. In particular, the proposed policy achieves an energy saving of up to 90%, when the cross interference level is large or the number of users is large (e.g., when $\alpha = 0.9$ in Fig. 5 and when $N = 7$ in Fig. 6). These are exactly the deployment scenarios where improvements in spectrum and energy efficiency are much needed. In addition, the proposed policy can always remain feasible even when the other policies cannot maintain the minimum throughput requirements.

VI. Conclusion

In this paper, we proposed a design framework for nonstationary policies to achieve cost minimization in decentralized resource sharing problems. The proposed nonstationary resource sharing policies greatly outperform existing stationary policies. In one of many applications of our design framework, namely power control, we demonstrate the significant performance improvement (up to 90% energy saving) over existing policies. We also proposed a low-complexity online algorithm, which can be run by each user in a decentralized manner, to construct the policy. Furthermore, the proposed policy only requires the users to have imperfect and binary feedback on the resource usage status.

References

[1] R. D. Yates, "A framework for uplink power control in cellular radio systems," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 7, pp. 1341–1347, Sep. 1995.

[2] T. Alpcan, T. Basar, R. Srikant, and E. Altman, "CDMA uplink power control as a noncooperative game," *Wireless Networks*, vol. 8, pp. 659–670, 2002.

[3] M. Xiao, N. B. Shroff, and E. K. P. Chong, "A utility-based power control scheme in wireless cellular systems," *IEEE/ACM Trans. Netw.*, vol. 11, no. 2, pp. 210–221, Apr. 2003.

[4] E. Altman and Z. Altman, "S-modular games and power control in wireless networks," *IEEE Trans. Autom. Control*, vol. 48, no. 5, pp. 839–842, May 2003.

[5] J.-S. Pang, G. Scutari, F. Facchinei, and C. Wang, "Distributed power allocation with rate constraints in Gaussian parallel interference channels," *IEEE. Trans. Inform. Theory*, vol. 54, no. 8, Aug. 2008.

[6] P. Hande, S. Rangan, M. Chiang, and X. Wu, "Distributed uplink power control for optimal SIR assignment in cellular data networks," *IEEE/ACM Trans. Netw.*, vol. 16, no. 6, pp. 1420–1433, Dec. 2008.

[7] S. Sorooshyari, C. W. Tan, M. Chiang, "Power control for cognitive radio networks: Axioms, algorithms, and analysis," *IEEE/ACM Trans. Netw.*, vol. 20, no. 3, pp. 878–891, Jun. 2012.

[8] L. Yang, H. Kim, J. Zhang, M. Chiang, and C. W. Tan, "Pricing-based spectrum access control in cognitive radio networks with random access," in *Proc. IEEE INFOCOM 2011*, pp. 2228–2236, 2011.

[9] K. Bharath-Kumar and J. M. Jaffe, "A new approach to performanceoriented flow control," *IEEE Trans. on Commun.*, vol. 29, pp. 427-435, 1981.

[10] Y. Su and M. van der Schaar, "Linearly coupled communication games," *IEEE Trans. Commun.*, vol. 59, no. 9, pp. 2543–2553, Sep. 2011.

[11] H. Shen and T. Basar, "Optimal nonlinear pricing for a monopolistic network service provider with complete and incomplete information," *IEEE J. Select. Areas Commun.*, vol. 25, pp. 1216–1223, Aug. 2007.

[12] R. Etkin, A. Parekh, and D. Tse, "Spectrum sharing for unlicensed bands," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, pp. 517–528, Apr. 2007.

[13] Y. Wu, B. Wang, K. J. R. Liu, and T. C. Clancy, "Repeated open spectrum sharing game with cheat-proof strategies," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1922–1933, 2009.

[14] M. Le Treust and S. Lasaulce, "A repeated game formulation of energy-efficient decentralized power control," *IEEE Trans. on Wireless Commun.*, vol. 9, no. 9, pp. 2860–2869, Sep. 2010.

[15] Y. Xiao, J. Park, and M. van der Schaar, "Repeated games with intervention: Theory and applications in communications," *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 3123–3132, Oct. 2012.

[16] D. Fudenberg, D. K. Levine, and E. Maskin, "The folk theorem with imperfect public information," *Econometrica*, vol. 62, no. 5, pp. 997–1039, Sep. 1994.

[17] E. Altman, "Constrained Markov decision processes," *Chapman and Hall/CRC*, 1999.

[18] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, pp. 589–600, Apr. 2007.

[19] K. Liu and Q. Zhao, "Distributed learning in multi-armed bandit with multiple players," *IEEE Trans. Signal Proc.*, vol. 58, no. 11, pp. 5667-5681, Nov., 2010.

[20] K. Liu and Q. Zhao, "Cooperative game in dynamic spectrum access with unknown model and imperfect sensing," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, Apr. 2012.

[21] Y. Xiao and M. van der Schaar, "Repeated resource sharing among selfish players with imperfect binary feedback," *Prof. Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 452–459, Oct. 2012.

[22] Y. Xiao and M. van der Schaar, "Dynamic spectrum sharing among repeatedly interacting selfish users with imperfect monitoring," *IEEE J. Sel. Areas Commun., Special issue on Cognitive Radio Series*, vol. 30, no. 10, pp. 1890–1899, Nov. 2012.

[23] D. P. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, Aug. 2006.

[24] Y. Xiao and M. van der Schaar, "Appendix," Available: www.seas.ucla.edu/~yxiao/Allerton2013.pdf