Fiftieth Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
October 1 - 5, 2012

Repeated Resource Sharing Among Selfish Players With Imperfect Binary Feedback

Yuanzhang Xiao[1] and Mihaela van der Schaar[2]

***Abstract*— We develop a novel design framework for resource sharing among self-interested players, who adjust their resource usage levels to compete for a common resource. We model the interaction among the players as a *repeated resource sharing game with imperfect monitoring*, which captures four unique features of the considered interaction. First, the players inflict *negative externality* to each other due to the interference/congestion among them. Second, the players interact with each other *repeatedly* because of their long-term coexistence in the system. Third, since the players are decentralized, they are *selfish* and aim to maximize their own long-term payoffs from utilizing the resource rather than obeying any prescribed sharing rule. Finally, the players are informed of the interference/congestion level through a *binary* feedback signal, which is quantized from *imperfect* observation about the interference/congestion level.**

We first characterize the set of Pareto optimal operating points that can be achieved by deviation-proof resource sharing policies, which are policies that the selfish players find it in their self-interests to comply with. Next, for any given operating point in this set, we show how to construct a deviation-proof policy to achieve it. The constructed deviation-proof policy is amenable to distributed implementation, and allows players to use the resource in an alternating fashion. In the presence of strong negative externality, our policy outperforms existing resource sharing policies that dictate constant resource usage levels by the players. Moreover, our policy can achieve Pareto optimality even when the players 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.

I. Introduction

The design of resource sharing policies is essential for resource sharing systems such as power control, medium access control (MAC), and flow control. An important feature of a resource sharing system is the negative externality among the players: due to interference/congestion, an increase in one player's resource usage level will reduce the other players' payoffs. In addition, in a decentralized resource sharing system, due to the lack of coordination, each player can only maximize its own payoff. The lack of coordination, along with the negative externality, renders the equilibrium outcomes of many resource sharing systems inefficient [1].

To improve inefficient equilibrium outcomes, many resource sharing policies have been proposed. Most works modeled the resource sharing system as a one-shot game [2]– [11], and designed resource sharing policies based on pricing (e.g. [2] [4] [6] [7]) and auctions (e.g. [3]). Recently, a new class of resource sharing policies based on "intervention" [12] were proposed in[1] MAC games [14] and power control games [15].

However, the above resource sharing policies [2]– [15], designed under one-shot game models, neglect an important feature of many resource sharing systems: the repeated interaction among the players. Specifically, the (resource sharing) policies in [2]– [15] assume that the players interact only once, and require the players to consume resources at *constant* levels over the time horizon in which they interact[2]. We call such policies in [2]– [15] *static* policies. When the negative externality among the players is strong, we can significantly improve static policies by using *dynamic* policies, which allow players to choose time-varying resource usage levels. Since the players coexist in the system for a relatively long period of time, we can model their interaction as a repeated game, and design dynamic policies under repeated game models.

Several dynamic resource sharing policies have been proposed in a repeated game framework [16]– [19], under the assumption of *perfect* monitoring. Specifically, they assume that each player perfectly knows the individual resource usage levels of all the players. In the dynamic policies in [16]– [19], the assumption of perfect monitoring is vital for the accurate detection of deviation, following which a perpetual [16] or temporary [17]– [19] punishment phase will be triggered. In the punishment phase, all the players choose the maximum resource usage levels to create strong interference/congestion to each other, and hence experience low payoffs as a punishment. Due to the threat of the punishment, all the users will follow the policy in their self-interests. However, since the monitoring can never be perfect, the punishment phase, in which all the users receive low throughput, will be triggered by mistake even if no one actually deviates. Thus, the players' overall payoffs, averaged over time, cannot be Pareto optimal because of the low payoffs received in the punishment phases. Hence, the policies in [16]– [19] have large performance loss in practice due to the imperfect monitoring.

In this paper, we design dynamic resource sharing policies

[1]Y. Xiao is with the Electrical Engineering Department, UCLA, Los Angeles, CA 90095, USA. Email: yxiao@ee.ucla.edu

[2]M. van der Schaar is with the Electrical Engineering Department, UCLA, Los Angeles, CA 90095, USA. Email: mihaela@ee.ucla.edu

[1]An incentive scheme with a similar philosophy was proposed to flow control games in [13].

[2]Although some resource sharing policies 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.

978-1-4673-4539-2/12/$31.00 ©2012 IEEE

to achieve Pareto optimal equilibrium outcomes that are not achievable by existing static policies [2]– [15]. We provide a systematic design approach, which first characterizes the set of Pareto optimal equilibrium outcomes, and then achieves any outcome in this set by a deviation-proof policy, which is a policy followed by the players in their self-interests. The constructed policy can be easily implemented in a decentralized manner. Moreover, we prove that the proposed policy can achieve Pareto optimal equilibrium outcomes, even when the players have *imperfect and limited* monitoring on the interference/congestion level. More specifically, they only receive a *binary* feedback signal, which is quantized from the *erroneous* measurement on the interference/congestion level. The requirement of imperfect binary feedback is significantly relaxed compared to that of perfect monitoring in [16]– [19]. Note also that to achieve Pareto optimality, the state-of-the-art results in repeated game theory [20] require the cardinality of the (imperfect) feedback signal increases with the number of resource usage levels that each player can choose. By exploiting unique features in the resource sharing system, we can achieve Pareto optimality with binary feedback regardless of the number of resource usage levels.

We illustrate the performance gain of the proposed dynamic policies over the existing static policies and dynamic policies designed under the assumption of perfect monitoring. In Fig. 1, we show the best equilibrium outcomes achievable by different classes of policies in a two-player resource sharing system. Due to the strong negative externality, the best equilibrium outcomes achievable by static policies [2]–[15] (the dashed pink curve) are Pareto dominated by the best equilibrium outcomes achieved by dynamic policies (the black straight line). The proposed dynamic policy, restricted by the deviation-proof requirement, can achieve a portion of the Pareto optimal outcomes (the red thick line). Assuming imperfect binary feedback, the dynamic policies designed under the assumption of perfect monitoring [16]– [19] (the green solid curve) have large performance loss compared to the proposed policy.

The rest of this paper is organized as follows. We discuss related works on dynamic policies in Section II. Section III describes the model of repeated resource sharing games with imperfect binary feedback, and illustrates how the general model applies to several different communication systems. In Section IV, we formulate the policy design problem, which is solved in Section V. Simulation results are presented in Section VII. Finally, Section VIII concludes the paper.

II. Related Works On Dynamic Policies

As discussed above, the dynamic policies outperform the static policies in resource sharing games. Hence, we focus on comparing the proposed policy with existing dynamic policies.

A. Dynamic Policies Based on Repeated Games

As discussed above, the major limitation of the works based on repeated games [17]– [19] is the assumption of perfect monitoring, which requires an unlimited amount of feedback. Limited feedback is assumed in [20]. However, [20] requires that the amount of feedback increases with the number of resource usage levels that the players can choose. In contrast, we only require binary feedback regardless of the number of resource usage levels.

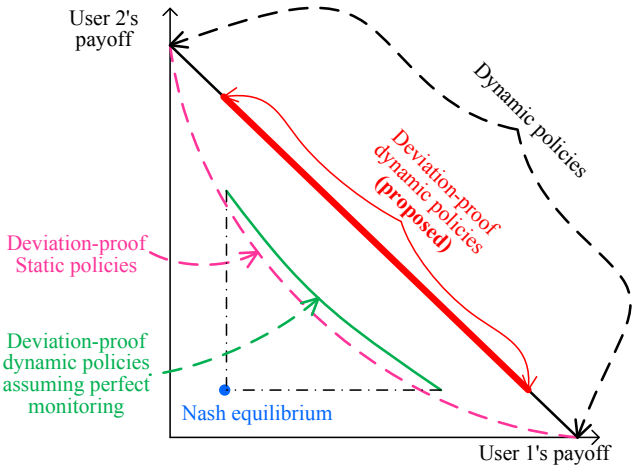

Fig. 1. An illustration of the best equilibrium outcomes achievable by static policies, dynamic policies assuming perfect monitoring, and the proposed policies in a two-player resource sharing system.

TABLE I
Related Works on Dynamic Policies.

Related Works	Limitations
Repeated Games [17]– [20]	Large feedback overhead
MDP [21]	Not applicable to the case of multiple users
MAB [22] [23]	Feedback schemes given, homogeneous and sufficiently patient players

B. Dynamic Policies Based on MDP

Many works developed optimal dynamic policies for resource sharing problems based on Markov decision processes (MDP). However, almost all the approaches based on MDP focus on optimal dynamic policies for a *single* user, and cannot be easily applied to the case of multiple users even when the users are obedient [21]. When multiple users compete for a single resource, they are optimizing their own payoffs in a competitive multi-user MDP, which cannot be solved by existing MDP theory.

C. Dynamic Policies Based on Multi-arm Bandit

Optimal policies based on multi-arm bandit (MAB) have been proposed in [22] [23]. However, [22] assumed that there is no observation error, although the feedback is limited. In addition, the feedback scheme is assumed to be given and is not one of the design parameters in [22] [23], while it is an important design parameter in our work. Moreover, [22] [23] assumed that the players are homogeneous and sufficiently patient (time-average payoffs are used). In our work, we consider heterogeneous and impatient players (discounted average payoffs are used).

We summarize the major limitations of the existing dynamic policies compared to our proposed policy in Table I.

III. System Model

A. Resource Sharing Games

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: The 3-tuple $\mathcal{G}$ represents a resource sharing game, if the action sets and the utility functions satisfy the following properties:

- Each player i's action set A_i is compact. In addition, A_i includes 0 as an element, i.e. $0 \in A_i$.
- Each player i's utility function u_i is decreasing in another player's action a_j, $\forall j \neq i$. When $a_i = 0$, we have $u_i(\mathbf{a}) = 0$.

The above definition captures the main characteristics of a resource sharing system. The first property on the action sets indicates that each player's action set is closed and bounded, which is consistent with the players' inability to consume unlimited amount of resources. In addition, each player can choose not to use any resource by taking action 0. The second property reflects the negative externality among the players in a resource sharing system: the increased resource usage by one player results in a decrease in the other players' payoffs. Moreover, when a player does not use the resource by choosing action 0, it will receive zero payoff.

Among all the resource sharing games, we are particularly interested in those with strong negative externality. The strong negative externality results from strong interference/congestion among the players. It is important to study the games with strong negative externality, because efficient resource sharing policies are essential to mitigate the strong interference/congestion among players. On the contrary, if the interference/congestion among players is weak, efficient resource sharing policies are less important, because the players can just choose their optimal resource usage levels individually without influencing the other players. We formally define the games with strong negative externality as follows. Before the definition, we define each player i's *preferred action profile* as $\tilde{\mathbf{a}}^i = \arg\max_{\mathbf{a}\in\mathcal{A}} u_i(\mathbf{a})$. Since each player's payoff is decreasing in the other players' actions, we know that $\tilde{a}^i_j = 0$, $\forall j \neq i$.

Definition 2: The 3-tuple $\mathcal{G}$ represents a resource sharing game with strong negative externality, if it is a resource sharing game, and if the following conditions are satisfied:

- The set of feasible payoffs $\mathcal{V} = \text{conv}\{\mathbf{u}(\mathbf{a}) = (u_1(\mathbf{a}), \ldots, u_N(\mathbf{a})) : \mathbf{a} \in \mathcal{A}\}$, where $\text{conv}(X)$ is the convex hull of X, has $N + 1$ extremal points[3]: $(0, \ldots, 0) \in \mathbb{R}^N$, $\mathbf{u}(\tilde{\mathbf{a}}^1), \ldots, \mathbf{u}(\tilde{\mathbf{a}}^N)$.

This definition characterizes the strong negative externality among the players: the increase of one player's payoff comes at such an expense of the other players' payoffs that the set of feasible payoffs without time sharing is nonconvex. An illustration of the feasible payoff region in a resource sharing game with strong negative externality is shown in Fig. 1. In the two-player system shown, the feasible payoff region is the convex hull of three points (the triangle bounded by the two axis and the solid black line): the origin $(0, 0)$, the payoff at player 1's preferred action profile, and the payoff at player 2's preferred action profile. Any payoff on the Pareto boundary can be achieved only by alternating between player 1's and player 2's preferred action profiles, but not by any other pure action profile.

In this paper, since we focus on resource sharing games with strong negative externality, we will sometimes referred to them simply as resource sharing games.

B. Repeated Resource Sharing Games With Imperfect Binary Feedback

Since the players interact with each other repeatedly, we model their interaction as a repeated game with the stage game being the resource sharing game. In the repeated game, the stage game is played in every period $t = 0, 1, 2, \ldots$. At the beginning of each period t, the players choose the action profile $\mathbf{a}^t$, which will lead to a interference/congestion level $c(\mathbf{a}^t)$ in this period. We assume that $c(\mathbf{a})$ is increasing in a_i, $\forall i \in \mathcal{N}$. The interference/congestion level is observed[4] with errors. We denote the noisy observation at period t by z^t, which is characterized by the conditional probability density function $\eta(z^t|c(\mathbf{a}^t))$. The noisy observation is then quantized and fedback to the players. 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 and will be proved to be sufficient to achieve Pareto optimality. 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

$$\begin{aligned} \rho_{z_0}(y^t = 0|\mathbf{a}^t) &= \int_{z^t<z_0} \eta(z^t|c(\mathbf{a}^t))dz^t, \\ \rho_{z_0}(y^t = 1|\mathbf{a}^t) &= 1 - \rho(y^t = 0|\mathbf{a}^t). \end{aligned} \quad (1)$$

The conditional probability distribution ρ_{z_0} completely characterizes the binary feedback scheme, which will be part of our design. In particular, the quantization threshold z_0 will be a key design parameter to optimize.

In summary, the repeated resource sharing game with imperfect binary feedback is characterized by the stage game (i.e. the resource sharing game) $\mathcal{G}$ and the feedback scheme ρ_{z_0}.

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

[4]In the next subsection, we will make it clear who observes the interference/congestion level and sends the feedback signal in concrete examples.

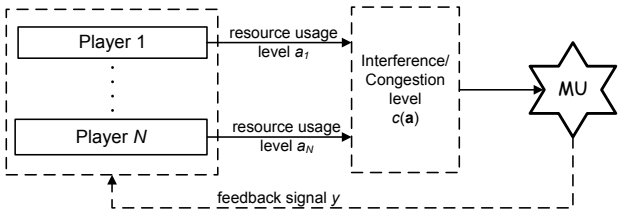

Fig. 2. An illustration of the role of the monitoring unit (MU) in the resource sharing system.

For better exposition, we define a logic unit, the *monitoring unit* (MU), as an entity that measures the interference/congestion level and sends the feedback signal. The monitoring unit can be controlled by the designer in determining the quantization threshold, as well as in designing resource sharing policies, which will be explained later. The role of the monitoring unit in the resource sharing system is illustrated in Fig. 2. In the following subsection, it will be clear what are the monitoring units in different scenarios.

C. Instantiation of The Model in Different Scenarios

In the following, we will show the instantiation of the general model in different scenarios of communication networks. Note that the applications are not limited to communication networks.

1) Power Control: Consider the power control problem in a one-hop wireless ad hoc network, where all the players transmit in the same channel [2]– [5] [15] [17] [18]. Each user i chooses its transmit power level (i.e. its action) $a_i \in [0, \bar{a}_i]$, and receives a payoff that can be defined as the Shannon rate:

$$u_i(\mathbf{a}) = \log_2 \left(1 + \frac{g_{ii} a_i}{\sum_{j \neq i} g_{ij} a_j + \sigma_i} \right),$$

where g_{ij} is the channel gain from player j's transmitter to player i's receiver, and σ_i is the noise power at player i's receiver. There is strong negative externality among the players when the cross channel gains are large [24] [25].

The interference level is the interference temperature measured at a measure center [3] [5], namely

$$c(\mathbf{a}) = \sum_{i \in \mathcal{N}} g_{0i} a_i,$$

where g_{0i} is the channel gain from player i's transmitter to the measure center's receiver. In this case, the MU is the measure center, who measures the interference temperature with some additive error e, namely $z = c(\mathbf{a}) + e$, and sends the binary feedback signal y to the players.

2) MAC: Consider a MAC problem in which multiple players transmit to an access point (AP) using slotted Aloha protocol [6] [7]. Each user i chooses whether to transmit or not. Hence, user i's action is $a_i \in \{0, 1\}$, with 0 being "transmit" and 1 being "not transmit". User i's payoff can be the throughput (assuming that simultaneous transmission leads to packet loss):

$$u_i(\mathbf{a}) = a_i \cdot \prod_{j \neq i} (1 - a_j).$$

The MAC system exhibits the extreme of strong interference among players: simultaneous transmissions from different players result in packet loss. Hence, MAC games are resource sharing games with strong negative externality.

The interference level is also the interference temperature. Since we assume homogeneous users in MAC, the interference temperature is proportional to the total number of transmissions. Normalized by the transmit power level, the interference level is thus defined as

$$c(\mathbf{a}) = \sum_{i \in \mathcal{N}} a_i.$$

In this case, the MU is the AP, who measures the interference temperature with some additive error e, namely $z = c(\mathbf{a}) + e$, and sends the binary feedback signal y to the players.

3) Flow Control: Consider a flow control problem in which multiple players transmit packets to a server with a service rate s [8]– [11]. Each user i chooses its transmission rate (i.e. its action) $a_i \in [0, s]$, and receives its payoff that can be defined as follows:

$$u_i(\mathbf{a}) = a_i^{\beta_i} \cdot \max \left\{ 0, s - \sum_{i \in \mathcal{N}} a_i \right\},$$

where β_i is a parameter representing the trade-off between the transmission rate a_i and the delay $1/\left(s - \sum_{i \in \mathcal{N}} a_i\right)$. A larger β_i means that the transmission rate is more important. The congestion level is the total transmission rate, namely

$$c(\mathbf{a}) = \sum_{i \in \mathcal{N}} a_i.$$

In this case, the MU is the server, who measures the total transmission rate with errors, and sends the binary feedback signal to the players.

IV. Formulation of The Policy Design Problem

In this section, we first define the deviation-proof resource sharing policy. Then we formulate the policy design problem.

A. Deviation-proof Resource Sharing Policies

A resource sharing policy specifies the action profile $\mathbf{a}^t$ the players choose at each period t based on the past history. Formally, the history up to period t is $h^t = \{y^0, \ldots, y^{t-1}\}$ for $t \geq 1$ and $h^0 = \varnothing$. Then a resource sharing policy π is a mapping from the set of possible histories $\mathcal{H} \triangleq \sqcup_{t=0}^{\infty} \{0, 1\}^t$ to the joint action set $\mathcal{A}$. We denote player i's policy by $\pi_i : \mathcal{H} \to A_i$.

Each player i's payoff is defined as the expected discounted average payoff per time slot. Assuming, as in [16]– [20], the same discount factor $\delta \in [0, 1)$ for all the players, player i's payoff can be written as

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

Note that the distribution of the history h^t is determined by the policy π and the distribution of the feedback signal ρ_{z_0}. The discount factor represents the "patience" of the users; a larger discount factor indicates that a user is more patient.

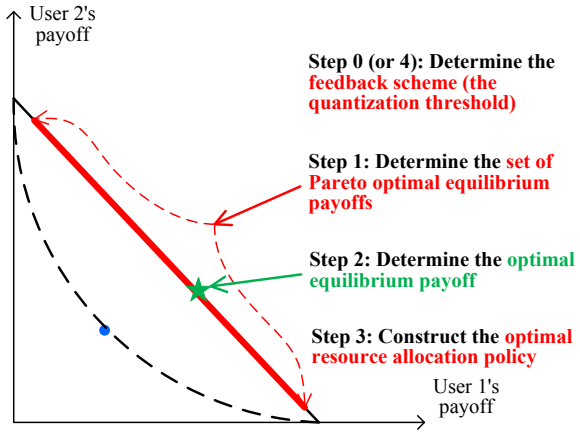

Fig. 3. Illustration of the design framework.

The discount factor is determined by the delay sensitivity of the user's applications.

We would like to implement deviation-proof policies, which are policies that the players find it in their self-interest to comply with. The deviation-proof policies are formally defined as follows.

Definition 3 (Deviation-proof Policies): In a repeated resource sharing game, the policy π is deviation-proof if for all $i \in \mathcal{N}$,

$$U_i(\pi) \geq U_i(\pi_i', \pi_{-i}), \ \forall \pi_i'. \tag{2}$$

We define the *equilibrium payoff* as the payoff $(U_1(\pi), \ldots, U_N(\pi))$ that can be achieved by a deviation-proof policy π.

B. The Policy Design Problem

The designer aims to maximize an objective function defined on the players' payoffs, $W(U_1(\pi), \ldots, U_N(\pi))$. This definition of the objective function is general enough to include the objective functions deployed in many existing works, such as [2]– [3] [16] [17]. An example of the objective function is the weighted sum payoff with minimum payoff guarantees, defined as

$$W(U_1, \ldots, U_N) = \begin{cases} \sum_{i=1}^N w_i U_i & \text{if } U_i \geq \gamma_i, \ \forall i \in \mathcal{N} \\ -\infty & \text{otherwise} \end{cases}, \tag{3}$$

where $\{w_i\}_{i=1}^N$ are the weights satisfying $w_i \in [0,1], \forall i$ and $\sum_{i=1}^N w_i = 1$, and $\gamma_i \geq 0$ is the minimum payoff guarantee for each player i. The minimum payoff guarantees are imposed to prevent each player from having extremely low payoffs. To sum up, we can formally define the policy design problem as

$$\begin{aligned} \max_{\pi} \quad & W(U_1(\pi), \ldots, U_N(\pi)) \\ s.t. \quad & \pi \text{ is deviation} - \text{proof}. \end{aligned} \tag{4}$$

V. The Design Framework

We outline the proposed design framework in Fig. 3. Given the feedback scheme, we first quantify the set of Pareto optimal equilibrium payoffs, namely the Pareto optimal payoffs that can be achieved by deviation-proof policies. Then, we determine the optimal equilibrium payoff based on the welfare function. Finally, we construct the deviation-proof policy to achieve the optimal equilibrium payoff. Since we can characterize the set of Pareto optimal equilibrium payoffs given the feedback scheme, we can design the optimal feedback scheme, namely the optimal quantization threshold.

A. Characterizing The Set of Pareto Optimal Equilibrium Payoffs

The first step in solving the design problem (4) is to characterize the set of Pareto optimal equilibrium payoffs for the repeated resource sharing game. First, we know from Definition 2 that the set of Pareto optimal payoffs can be written as $\mathcal{P} = \{\mathbf{v} : \sum_{i=1}^N v_i/\bar{v}_i = 1, \ v_i \geq 0, \forall i\}$, where $\bar{v}_i \triangleq u_i(\tilde{\mathbf{a}}^i)$ is the maximum payoff that player i can achieve. The set of Pareto optimal equilibrium payoffs is a subset of $\mathcal{P}$. Given the system parameters, in particular the feedback scheme, we define the set of Pareto optimal equilibrium payoffs as $\mathcal{P}^*(\rho_{z_0})$. Before determining $\mathcal{P}^*(\rho_{z_0})$, we define the *benefit from deviation* as follows.

Definition 4 (Benefit From Deviation): We define player j's benefit from deviation from player i's preferred action profile $\tilde{\mathbf{a}}^i$ as

$$b_{ij} = \sup_{a_j \in A_j, a_j \neq \tilde{a}_j^i} \frac{\rho_{z_0}(1|\tilde{\mathbf{a}}^i) - \rho_{z_0}(1|a_j, \tilde{\mathbf{a}}_{-j}^i)}{u_j(a_j, \tilde{\mathbf{a}}_{-j}^i)/\bar{v}_j}.$$

Now we state Theorem 1, which characterizes the set of Pareto optimal equilibrium payoffs $\mathcal{P}^*(\rho_{z_0})$. (Since $\tilde{a}_j^i = 0, \ \forall j \neq i$, at the action profile $\tilde{\mathbf{a}}^i$, we call player i the *active* player and player j the *inactive* player for all $j \neq i$.)

Theorem 1: Pareto efficient equilibrium payoffs can be achieved if and only if the following three conditions are satisfied:

- Condition 1: the inactive player has no benefit from deviation from the active player's preferred action profile, namely $b_{ij} < 0$ for all i and $j \neq i$;
- Condition 2: the active player has no incentive to deviate, namely for all $i \in \mathcal{N}$ and $a_i \in A_i$,

$$1 - \frac{u_i(a_i, \tilde{\mathbf{a}}_{-i}^i)}{\bar{v}_i} \geq \sum_{j \neq i} \frac{\rho_{z_0}(1|a_i, \tilde{\mathbf{a}}_{-i}^i) - \rho_{z_0}(1|\tilde{\mathbf{a}}^i)}{-b_{ij}};$$

- Condition 3: the discount factor δ is no smaller than the threshold $\underline{\delta}$, namely

$$\delta \geq \underline{\delta} \triangleq \frac{1}{1 + \frac{1 - \sum_{i \in \mathcal{N}} \underline{\mu}_i}{N - 1 + \sum_{i \in \mathcal{N}} \sum_{j \neq i} (-\rho_{z_0}(1|\tilde{\mathbf{a}}^i)/b_{ij})}}, \tag{5}$$

where

$$\underline{\mu}_i \triangleq \max_{j \neq i} \frac{1 - \rho_{z_0}(1|\tilde{\mathbf{a}}^i)}{-b_{ij}}. \tag{6}$$

If Pareto efficient equilibrium payoffs can be achieved, the set $\mathcal{P}^*(\rho_{z_0})$ is

$$\mathcal{P}^*(\rho_{z_0}) = \left\{ \mathbf{v} : \sum_{i=1}^N \frac{v_i}{\bar{v}_i} = 1, \ \frac{v_i}{\bar{v}_i} \geq \underline{\mu}_i, \forall i \in \mathcal{N} \right\}, \tag{7}$$

TABLE II

THE ALGORITHM RUN BY PLAYER i.

Require: The optimal equilibrium payoff $\mathbf{v}^\star$ obtained from the MU

Initialization: Sets $t=0$, $v_j(0)=v_j^\star$ for all $j\in\mathcal{N}$.

repeat

Calculates the index $\alpha_j(t)=\frac{v_j(t)/\bar{v}_j-\underline{\mu}_j}{1-v_j(t)/\bar{v}_j+\sum_{k\neq j}(-\rho_{z_0}(1|\tilde{\mathbf{a}}^j)/b_{jk})},\forall j$

Finds the largest index $i^*\triangleq\arg\max_{j\in\mathcal{N}}\alpha_j(t)$

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

Chooses the resource usage level $\tilde{a}_i^i$

end if

Updates $v_j(t+1)$ for all $j\in\mathcal{N}$

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

$$v_{i^*}(t+1)=\tfrac{1}{\delta}\cdot v_{i^*}(t)-(\tfrac{1}{\delta}-1)\cdot(1+\sum_{j\neq i^*}\frac{\rho_{z_0}(1|\tilde{\mathbf{a}}^{i^*})}{-b_{i^*j}})\cdot\bar{v}_{i^*}$$

$$v_j(t+1)=\tfrac{1}{\delta}\cdot v_j(t)+(\tfrac{1}{\delta}-1)\cdot\frac{\rho_{z_0}(1|\tilde{\mathbf{a}}^{i^*})}{-b_{i^*j}}\cdot\bar{v}_j,\ \forall j\in\mathcal{N},j\neq i^*$$

else

$$v_{i^*}(t+1)=\tfrac{1}{\delta}\cdot v_{i^*}(t)-(\tfrac{1}{\delta}-1)\cdot(1-\sum_{j\neq i^*}\frac{\rho_{z_0}(0|\tilde{\mathbf{a}}^{i^*})}{-b_{i^*j}})\cdot\bar{v}_{i^*}$$

$$v_j(t+1)=\tfrac{1}{\delta}\cdot v_j(t)-(\tfrac{1}{\delta}-1)\cdot\frac{\rho_{z_0}(0|\tilde{\mathbf{a}}^{i^*})}{-b_{i^*j}}\cdot\bar{v}_j,\ \forall j\in\mathcal{N},j\neq i^*$$

end if

$t\leftarrow t+1$

until $\varnothing$

which is nonempty if and only if $\sum_{i\in\mathcal{N}}\underline{\mu}_i\leq 1$.

Proof: The proof heavily replies on the concept of self-generating sets [26]. Due to space limit, we refer interesting readers to [27, Appendix A] for the complete proof. ■

Theorem 1 first provides the sufficient conditions for the existence of Pareto optimal equilibrium payoffs. Condition 1 (respectively, Condition 2) ensures that at action profile $\tilde{\mathbf{a}}^i$, the inactive players (respectively, the active player) has no incentive to deviate. Theorem 1 also gives us the minimum discount factor under which any payoff in $\mathcal{P}^*(\rho_{z_0})$ is achievable. Theorem 1 analytically quantifies the set of Pareto optimal equilibrium payoffs $\mathcal{P}^*(\rho_{z_0})$.

B. Determining The Optimal Equilibrium Payoff

Since we have identified the set of Pareto optimal equilibrium payoffs $\mathcal{P}^*(\rho_{z_0})$, the problem of find the optimal equilibrium payoff that solves the policy design problem can be rewritten as

$$\begin{aligned}\max_{\mathbf{v}}\quad & W(v_1,\ldots,v_N)\\ s.t.\quad & (v_1/\bar{v}_1,\ldots,v_N/\bar{v}_N)\in\mathcal{P}^*(\rho_{z_0}).\end{aligned}\tag{8}$$

The constraint in the above problem can be further simplified as $v_i\geq\underline{\mu}_i\cdot\bar{v}_i,\ \forall i\in\mathcal{N}$.

The optimization problem (8) is easy to solve when W is concave in $(v_1,\ldots,v_N)$. For example, if the welfare function is the weighted sum payoffs with minimum payoff guarantees defined in (3), the solution can be obtained analytically as $v_{i^*}^\star=(1-\sum_{j\neq i}\max\{\underline{\mu}_j,\gamma_j/\bar{v}_j\})\cdot\bar{v}_i$ for $i^*=\arg\max_{j\in\mathcal{N}}w_j\bar{v}_j$, and $v_i^\star=\max\{\underline{\mu}_i,\gamma_i/\bar{v}_i\}\cdot\bar{v}_i$ for all $i\neq i^*$.

C. Constructing The Deviation-Proof Policy

Given the optimal equilibrium payoff $\mathbf{v}^\star\in\mathcal{P}^*(\rho_{z_0})$, we can construct the deviation-proof policy that achieves the payoff $\mathbf{v}^\star$. According to Definition 2, any payoff $\mathbf{v}^\star\in\mathcal{P}^*(\rho_{z_0})$ should be achieved by alternating among the players' preferred action profiles. Hence, the optimal deviation-proof policy π satisfies $\pi(h^t)\in\{\tilde{\mathbf{a}}^1,\ldots,\tilde{\mathbf{a}}^N\}$ for any $t\geq 0$ and for any public history $h^t\in\{0,1\}^t$. Since only player i consumes the resources at the action profile $\tilde{\mathbf{a}}^i$, the deviation-proof policy can also be regarded as a scheduling in a TDMA (time-division multiple access) fashion.

Theorem 2 ensures that if all the players run the policy in Table II distributively, they will achieve the optimal equilibrium payoff $\mathbf{v}^\star$, and will have no incentive to deviate.

Theorem 2: For any payoff $\mathbf{v}^\star\in\mathcal{P}^*(\rho_{z_0})$, and any discount factor $\delta\geq\underline{\delta}$, the policy generated by each player running the algorithm in Table II is deviation-proof and achieves $\mathbf{v}^\star$.

Proof: See [27, Appendix B]. ■

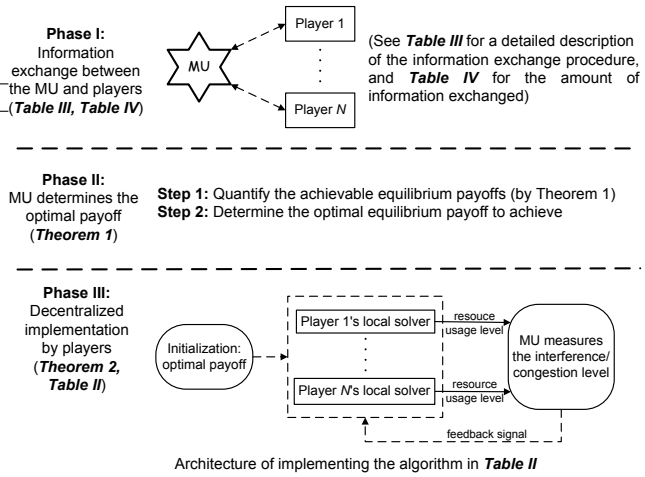

Fig. 4. Illustration of the implementation.

TABLE III

THE INFORMATION EXCHANGE PHASE.

Events	Information obtained
Players choose $\{\tilde{\mathbf{a}}^i\}_{i\in\mathcal{N}}$	MU: $\{\rho_{z_0}(1\|\tilde{\mathbf{a}}^i)\}_{i\in\mathcal{N}}$
Players choose $(a_j,\tilde{\mathbf{a}}^i_{-j}),\forall a_j$	MU: $\rho_{z_0}(1\|a_j,\tilde{\mathbf{a}}^i_{-j}),\forall a_j$
LSS broadcasts	Player i: $\rho_{z_0}(1\|\tilde{\mathbf{a}}^i),\rho_{z_0}(1\|a_j,\tilde{\mathbf{a}}^i_{-j})$
Players broadcast	MU, Players: $b_{ij},\forall i,j\neq i$
Players send to MU	MU: $\{\bar{v}_i\}_{i\in\mathcal{N}}$

D. Implementation Issues

We discuss the implementation issues of our proposed design framework. The proposed design framework can be implemented in three phases as illustrated in Fig. 4. In Phase I, the MU exchanges some information with the players following the procedure described in Table III. In Phase II, using the information obtained in Phase I, the MU quantifies the set of Pareto optimal equilibrium payoffs, and solves the policy design problem for the optimal equilibrium payoff. Finally in Phase III, the MU sends the optimal equilibrium payoff to the players, as an input to each player's decentralized algorithm of constructing the optimal deviation-proof

TABLE IV

COMPARISON OF THE AMOUNT OF INFORMATION EXCHANGED.

	Amount of information exchanged
[2]– [7]	$O(N)$ or $O(N^2)$ per iteration $\cdot$ # of iterations
Proposed	$\sum_i \sum_{j \neq i} \lvert A_j \rvert + N^2 + N$

policy. In our design framework, the designer programs the MU according to the three-phase procedure shown in Fig. 4, and then *leaves* the system. The players will autonomously achieve the optimal equilibrium payoff under the guidance of the MU.

1) Overhead of information exchange: We briefly comment on the overhead of the information exchange in the proposed framework. First, the information exchange in Phase I is necessary for the MU to determine and for the players to achieve the optimal equilibrium payoff. A similar information exchange phase is proposed in [16] [17] [28]. The information exchange phase can be considered as a substitute for the convergence process needed by the algorithms in [2]– [7]. In the proposed policy, since the players implement the policy without any information exchange in Phase III, the only information exchange happen in Phase I and at the end of Phase II (when the MU broadcasts the optimal equilibrium payoff). The information exchange method in our framework is advantageous in that its duration and the amount of information to exchange are fixed and predetermined. On the other hand, the amount of information to exchange in [2]– [7] is proportional to the convergence time of their algorithms, which are generally unbounded. We summarize the overhead of information exchange (measured by the number of real numbers or pilot signals transmitted) in Table IV.

2) Computational complexity: As we can see from Table II, the computational complexity of each player in constructing the optimal policy is very small. At each period t, each player needs to compute N indices $\{\alpha_j(t)\}_{j\in\mathcal{N}}$, and N future payoffs $\{v_j(t)\}_{j\in\mathcal{N}}$, all of which can be calculated analytically. In addition, although the original definition of the resource sharing policy requires each player to memorize the entire history of feedback signals, in the actual implementation, each player only needs to memorize N future payoffs $\{v_j(t)\}_{j\in\mathcal{N}}$.

VI. SPECIAL CASES

The results in Section V were derived for general resource sharing games in Definition 2. Now we discuss how the results simplify in some special cases. The discussion is summarized in Table V.

A. Unselfish Users

If the users are not selfish, the conditions to achieve Pareto optimal equilibrium payoffs reduce to Condition 3 with $\underline{\delta} = \frac{N-1}{N}$. In addition, all the Pareto optimal equilibrium payoffs can be achieved (i.e. $\underline{\mu}_i = 0, \forall i$). The algorithm can also be simplified by setting $b_{ij} = -\infty, \forall i, j$. Then we can see that the update of $v_j(t+1)$ is the same under different feedback signals. In other words, no feedback signal is needed. Hence, the amount of information exchange reduces to $2N$.

TABLE V

RESULTS FOR SPECIAL CASES.

	Conditions	$\mathcal{P}^*(\rho_{z_0})$	Algorithm	Info. Exchange
Unselfish	Condition 3 ($\underline{\delta} = \frac{N-1}{N}$)	$\underline{\mu}_i = 0, \forall i$	$b_{ij} = -\infty$ $\forall i, j$	$2N$
Power Control	Condition 1,3	(6)	same	same
MAC	Condition 3 ($\underline{\delta} = \frac{N-1}{N}$)	$\underline{\mu}_i = 0, \forall i$	$b_{ij} = -\infty$ $\forall i, j$	$2N$
Flow Control	Condition 1,2,3	(6)	same	same

B. Power Control

In power control, each player's payoff is increasing in its own action. Hence, Condition 2 is always satisfied. The conditions to achieve Pareto optimal equilibrium payoffs reduce to Condition 1 and 3. The other results remain the same.

C. MAC

In MAC, each player's payoff is also increasing in its own action. In addition, a player cannot gain by deviating, because it receives zero payoff when if it transmits at another player's slot. Hence, both Condition 1 and 2 are satisfied automatically. In summary, the results for MAC are the same as the results for unselfish users.

D. Flow Control

In flow control, each player's payoff is not monotone in its own action. Hence, Condition 2 is needed. Actually, the results for the general resource sharing games apply to flow control without simplification.

VII. SIMULATION RESULTS

In this section, we demonstrate the performance gain of our proposed resource sharing policy over existing policies. We conduct the comparison in a power control system. We use the following system parameters. The noise powers at all the players' receivers are normalized as 0 dB. The maximum transmit powers of all the players are $\bar{a}_i = 10$ dB, $\forall i$. For simplicity, we assume that the 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 $g_{ij} \sim \mathcal{CN}(0,\beta), \forall i \neq j$, where β is defined as the *cross interference level*. The channel gain from each player to the MU also satisfies $g_{0i} \sim \mathcal{CN}(0,1), \forall i$. The additive measurement error e is Gaussian distributed with zero mean and variance 0.1. The quantization threshold is $z_0 = 10$ dB. The welfare function is the spectrum efficiency, i.e. the welfare function defined in (3) with equal weights. The minimum payoff guarantee is 10% of the maximum achievable payoff, i.e. $\gamma_i = 0.1 \cdot \bar{v}_i, \forall i$.

We compare the performance of the proposed policy with those of the optimal static policy and of the existing dynamic policies designed under the assumption of perfect monitoring

TABLE VI

COMPARISON OF DIFFERENT POLICIES IN TERMS OF SPECTRUM AND ENERGY EFFICIENCY.

	6 users, $\beta = 0.2$	18 users, $\beta = 0.2$
Static	1.6 bits/s/Hz, 0.20 W	infeasible, infeasible
Trigger	1.8 bits/s/Hz, 0.12 W	infeasible, infeasible
Proposed	3.0 bits/s/Hz, 0.04 W	2.5 bits/s/Hz, 0.8 W

[17]– [19]. We call the dynamic policies in [17]– [19] *trigger* policies, which requires users to switch to the punishment phase once a deviation is detected. In the punishment phase, all the users transmit at the maximum power levels to create high interference to the deviator.

We list the spectrum and energy efficiency of the three different policies for some typical scenarios in Table VI. The energy efficiency is the average transmit power of each player to achieve the optimal spectrum efficiency. In both scenarios, we can see that the proposed policy outperforms the other two in both criteria. The performance gain is substantial when the number of users is large. For example, when the number of users is as large as 18, the other two policies are infeasible, because the minimum throughput guarantee cannot be satisfied, while the proposed policy can still achieve good spectrum and energy efficiency.

VIII. CONCLUSION

In this paper, we defined repeated resource sharing games with strong negative externality, and proposed a design framework for selfish players to achieve Pareto optimal equilibrium payoffs in a decentralized fashion. The proposed policy can achieve Pareto optimal equilibrium payoffs, which are not achievable under optimal static policies. In addition, the proposed policy can achieve Pareto optimality even with imperfect binary feedback, which significantly reduces the feedback overhead required in existing dynamic policies designed under the assumption of perfect monitoring.

REFERENCES

[1] A. MacKenzie and S. Wicker, "Game theory and the design of self-configuring, adaptive wireless networks," *IEEE Communication Magazine*, vol. 39, no. 11, pp. 126–131, Nov. 2001.

[2] J. Huang, R. A. Berry, and M. L. Honig, "Distributed interference compensation for wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 5, pp. 1074–1084, May 2006.

[3] J. Huang, R. A. Berry, and M. L. Honig, "Auction-based spectrum sharing," *Mobile Networks and Applications*, vol. 11, pp. 405–418, 2006.

[4] C. U. Saraydar, N. B. Mandayam, and D. J. Goodman, "Efficient power control via pricing in wireless data networks," *IEEE Transactions on Communications*, vol. 50, no. 2, pp. 291–303, Feb. 2002.

[5] S. Sharma and D. Teneketzis, "An externalities-based decentralized optimal power allocation algorithm for wireless networks," *IEEE/ACM Trans. Netw.*, vol. 17, no. 6, pp. 1819–1831, Dec. 2009.

[6] D. Wang, C. Comaniciu, and U. Tureli, "Cooperation and fairness for slotted Aloha," *Wireless Personal Communications*, vol. 43, no. 1, pp. 13–27, 2007.

[7] 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.

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

[9] C. Douligeris and R. Mazumdar, "A game theoretic perspective to flow control in telecommunication networks," *Journal of the Franklin Institute*, vol. 329, no.2, pp. 383-402, 1992.

[10] Z. Zhang and C. Douligeris, "Convergence of synchronous and asynchronous greedy algorithm in a multiclass telecommunications environment," *IEEE Trans. on Commun.*, vol. 40, no. 8, pp. 1277-1281, Aug. 1992.

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

[12] J. Park and M. van der Schaar, "The theory of intervention games for resource sharing in wireless communications," *IEEE J. Select. Areas Commun.*, vol. 30, no. 1, pp. 165–175, Jan. 2012.

[13] Yi Gai, Hua Liu, and Bhaskar Krishnamachari, "A packet dropping-based incentive mechanism for M/M/1 queues with selfish users," in *Proceedings of the 30th IEEE International Conference on Computer Communications (INFOCOM 2011)*, Shanghai, China, April, 2011.

[14] J. Park and M. van der Schaar, "Stackelberg contention games in multiuser networks," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, pp. 1–15, 2009.

[15] Y. Xiao, J. Park, and M. van der Schaar, "Intervention in power control games with selfish users," *IEEE J. Sel. Topics Signal Process., Special issue on Game Theory in Signal Processing*, vol. 6, no. 2, pp. 165–179, Apr. 2012.

[16] 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.

[17] 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.

[18] 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, september 2010.

[19] Y. Xiao, J. Park, and M. van der Schaar, "Repeated games with intervention: Theory and applications in communications," To appear in *IEEE Trans. Commun.*. Available: "http://arxiv.org/abs/1111.2456".

[20] 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.

[21] F. Fu and M. van der Schaar, "A systematic framework for dynamically optimizing multi-user video transmission," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 3, pp. 308-320, Apr. 2010. (Also featured in the IEEE MMTC R-Letter, Apr. 2011.)

[22] 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.

[23] 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.

[24] S. Stańczak and H. Boche, "On the convexity of feasible QoS regions," *IEEE Transactions on Information Theory*, vol. 53, no. 2, Feb. 2007.

[25] E. G. Larsson, E. A. Jorswieck, J. Lindblom, and R. Mochaourab, "Game theory and the flat-fading gaussian interference channel," *IEEE Signal Processing Magazine*, vol. 26, no. 5, pp. 18-27, Sep. 2009.

[26] D. Abreu, D. Pearce, and E. Stacchetti, "Toward a theory of discounted repeated games with imperfect monitoring," *Econometrica*, vol. 58, no. 5, pp. 1041–1063, 1990.

[27] Y. Xiao and M. van der Schaar, "Dynamic Spectrum Sharing Among Repeatedly Interacting Selfish Users With Imperfect Monitoring," Available at: http://arxiv.org/abs/1201.3328.

[28] A. De Domenico, E. C. Strinati, and M. G. Di Benedetto, "A survey on MAC strategies for cognitive radio networks," *IEEE Commun. Surveys Tutorials*, vol. 14, no. 1, pp. 21–44, 2012.