An Indirect-Reciprocity Reputation Game for Cooperation in Dynamic Spectrum Access Networks

Biling Zhang, *Member, IEEE,* Yan Chen, *Member, IEEE,* and K. J. Ray Liu, *Fellow, IEEE*

Abstract—Cooperation is a promising approach to simultaneously achieve efficient utilization of spectrum resource and improve the quality of service of primary users in dynamic spectrum access networks. However, due to the selfish nature, secondary users may not act as cooperatively as primary users have expected. Therefore, how to stimulate the secondary users to play cooperatively is an important issue. In this paper, we propose a reputation-based spectrum access framework, where the cooperation stimulation problem is modeled as an indirect reciprocity game. In the proposed game, secondary users choose how to help primary users relay information and gain reputations, based on which they can access a certain amount of vacant licensed channels in the future. By formulating a secondary user's decision making as a Markov decision process, we obtain the optimal action rule, according to which the secondary user will use maximal power to help primary user relay data if the channel is not in an outage, and thus greatly improve the primary user's quality of service as well as the spectrum utilization efficiency. Moreover, we prove the uniqueness of stationary reputation distribution and theoretically derive the condition under which the optimal action rule is evolutionarily stable. Finally, simulation results are shown to verify the effectiveness of the proposed scheme.

Index Terms—Indirect reciprocity, dynamic spectrum access, game theory, Markov decision process, evolutionarily stable strategy.

I. INTRODUCTION

RECENTLY, it has been reported that a large portion of the radio spectrum assigned by Federal Communications Commission (FCC) to licensed users, also known as primary users (PUs), remains under-utilized [1]. Dynamic spectrum access (DSA) technique, where unlicensed users, also known as secondary users (SUs), are allowed to access and share the licensed spectrum temporarily, has gained many attentions since it can greatly improve the efficiency of spectrum utilization [2] [3] [4].

In the literature, many spectrum access approaches have been proposed. Spectrum sharing is one kind of these approaches, which studies the allocation of vacant spectrum where PUs are sensed as inactive [5] [6] [7] [8]. In [5], the authors introduced a cognitive multiple-access protocol where SUs sense and exploit idle channels to cooperate with the PUs for forwarding their packets to improve the PUs' quality of service (QoS), while the authors in [6] [7] [8] focused on designing spectrum sharing rules for SUs to efficiently and fairly share the vacant channels to forward their own packets. Another way for SUs to use the licensed spectrum is through spectrum auction [9] [10], spectrum trading [11] [42] or spectrum leasing [12] [13]. In [11], the trading mechanism is modeled as a market where PUs set the qualities and/or prices for their idle spectrums and SUs choose the spectrum with appropriate quality and/or cost. Instead of getting virtual currency, in spectrum leasing [12] [13], PUs ask for cooperation from SUs by leasing a portion of spectrum to them, due to which the PUs' QoS is improved while SUs gain the right to access licensed spectrum.

How to stimulate the cooperation from SUs in DSA networks is an important problem. Since PUs and SUs generally belong to different operators or service providers and pursue different objectives, SUs may not cooperate if cooperation cannot bring them benefits. Game theory, a well-developed mathematical tool that studies the interactions among rational users [14], has been used to analyze the cooperation behaviors among nodes in DSA networks [12] [13] [15] [16] [17] [18] [19] [41] [43]. However, most of the existing game theoretic frameworks for DSA networks are based on the direct reciprocity model [12] [13] and repeated game [18], where the underlying assumption is that the game among a group of nodes is played for infinite times. Nevertheless, this assumption is not true in reality since players need to change their partners frequently due to mobility or changes of environment. In such a case, the only optimal strategy for them is to always play non-cooperatively. Moreover, without punishment for the cheating behavior, players may cheat even after they have agreed to cooperate. For example, in the systems such as [13], an SU may use less power to relay than it had promised after it had gained the chance to use the spectrum since there is no supervision and evaluation for its action.

Another two categories of cooperation stimulation schemes are payment-based and reputation-based schemes. Payment-

based schemes, such as virtual currency [20] and auction [21] [22] [23], have been proposed to enforce cooperation in DSA networks. In [20], a payment mechanism where PUs and SUs can pay charges to each other, was proposed to motivate the cooperation between multiple PUs and multiple SUs. In [23], the authors proposed an auction-based protocol to encourage PUs to free a portion of their spectrum resources to SUs for the exchange of the help from SUs to relay their signals to the destinations. Although these schemes can achieve promising results, the requirement of tamper-proof hardware or central billing services greatly limits their potential applications. Reputation-based cooperation stimulation schemes are widely discussed in Ad-hoc networks [24] [25] and P2P networks [26] [27]. In [26], Satsiou and Tassiulas proposed a reputation-based system, where peers earned reputations according to their contributions and received resources in proportion to their reputations. Under such a reputation system, the authors showed that rational peers, who seek to maximize their utilities with the least possible contributions, have the incentive to cooperate. However, few reputation-based mechanisms have been proposed for DSA networks, and there is no theoretical analysis about using reputation to stimulate cooperation in DSA networks.

Indirect reciprocity has recently drawn a lot of attentions in the area of social science and evolutionary biology [28] [29]. In the context of indirect reciprocity, the interactions between any pair of donor and recipient are short-term. During the interaction, the donor's decision to help the recipient can be interpreted as an indirect expression of gratitude, that is, I help you not because you have helped me but because you have helped others. In [30], Chen and Liu have showed that indirect reciprocity game is a new paradigm to stimulate cooperation among cognitive nodes. However, the framework proposed in [30] is too general and cannot be directly applied to DSA networks due to the following two reasons. Firstly, unlike in other cognitive networks where players are homogeneous, the cooperation in DSA networks takes place more frequently between a primary system and a secondary system, where in most cases a PU acts as a leader and an SU is a follower. Secondly, cooperation between a primary system and a secondary system is asymmetric, e.g., PUs offer spectrum while SUs offer power. Therefore a careful design for the resource allocation in DSA networks by taking into account the overall system performance is needed.

In this paper, we propose a spectrum access framework to address the efficient allocation of channels by incorporating indirect reciprocity based incentive mechanism to stimulate the cooperation of secondary users and thus improve the overall system performance. Our major contributions are summarized as follows.

- Different from previous works which focused on direct reciprocity models, we model the cooperation stimulation problem in DSA networks as an indirect reciprocity game. In the game, PUs are source nodes and SUs are selected as relays while the base station (BS) is the destination as well as the observer who assigns SUs' reputations based on their actions. Indirect reciprocity game here is used not only to stimulate the cooperation of SUs, but also to overcome the cheating behaviors. On
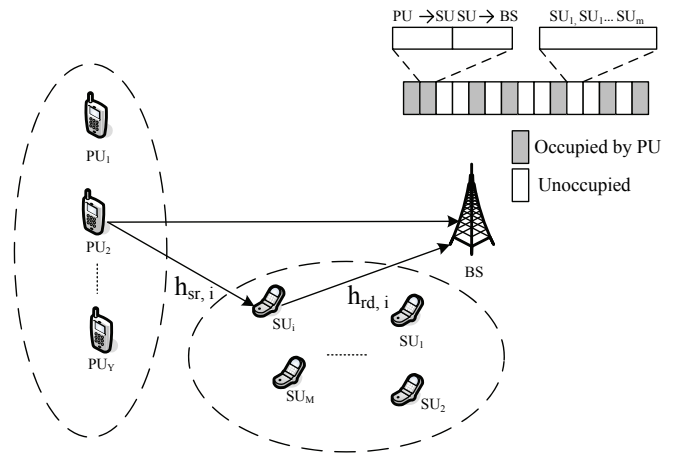

Fig. 1. The system model.

one hand, we incorporate the energy detection mechanism in the BS, which can help the BS detect the cheating behavior when the SU uses less power for relaying data than it has promised to. On the other hand, if the SU uses less power for relaying data, it will get less reputation, due to which the SU has no incentive to cheat.

- While previous works focused on the reputation propagation, our proposed model does not have reputation propagation error since the reputation is assigned by and stored in the BS. However, the noise in channels between SUs and the BS may distort the SUs' actions and affect the detection accuracy of the BS. To overcome such a problem, we take into account the impact of channel noise and develop a reputation updating policy to update SUs' reputations. In our model, we adopt the concept of reputation distribution to capture the likelihoods of the SU's reputation and prove theoretically the uniqueness of stationary reputation distribution of the whole population.

- In our proposed scheme, we formulate the SUs' decision making as a Markov Decision Process (MDP) and use a modified value iteration algorithm to find the optimal action rule. We show that within an appropriate cost-to-gain ration, the optimal strategy for SUs is to use the maximum power to relay if the source-relay-destination channel does not encounter an outage, which greatly improves the PUs' QoS. We also find that this optimal action rule will lead to a "good" society where most of the SUs have the highest reputation. Finally, we derive theoretically the stable condition of the optimal action rule.

The rest of the paper is organized as follows. Section II describes in details our system model for the DSA networks. In Section III, we introduce the reputation updating policy, characterize the equilibrium of the indirect reciprocity game, and analyze the stability of optimal action rule. Finally, we show the simulation results in Section IV and draw conclusions in Section V.

II. SYSTEM MODEL

As shown in Fig. 1, we consider the system where PUs far away from the BS are experiencing low achievable rates

while some SUs between the PUs and the BS are searching opportunities to use the licensed spectrum. It is well known that cooperative communications can greatly improve the communication performance [31]. To improve the PU's QoS, e.g., achievable rate or throughput, if there are SUs having more favorable channel conditions than the PU, then one of them can be selected as a relay by the BS based on the relay selection strategy in [32][1]. In this paper, we address the uplink transmission case, where PUs work as the source nodes and SUs and the BS are the relays and destination, respectively. In the rest of this paper, the channel between a PU and an SU, an SU and the BS, and a PU and the BS are referred as S-R channel, R-D channel and S-D channel, respectively, for short.

An SU who is selected as the relay needs to make decision on whether to help the PU relay the information, and if help, how much power to use for relaying. Although our scheme is applicable to any cooperation protocol, here, we assume SUs use Decode-and-Forward (DF) protocol to relay PUs' signals. We assume that a time frame is divided into two time slots, where the PU transmits in the first time slot and the selected SU relays PU's information in the second time slot. After the transmission, the BS will update the selected SU's reputation. Later, when the SU applies for the usage of vacant channels, it will be given the right for accessing a certain amount of vacant channels by the BS according to its reputation.

From the discussion above, we can see that the concept of indirect reciprocity adopted in our model is that the BS helps certain SUs because they have helped PUs.

A. Action

In our model, action $a \in \mathcal{A}$ stands for the power level used by an SU when relaying the data of a PU, where $\mathcal{A} = \{1, ..., N\}$ is the action set that an SU can choose its action from when it is selected as a relay. In the action set $\mathcal{A}$, "1" represents that an SU uses zero power to relay, i.e., the SU denies cooperation, while "N" denotes that an SU relays a PU's data with its maximum power P_{max}, i.e., SU makes its best effort to cooperate.

While an SU makes its decision, it needs to consider how many reputations it has gained as well as what the source-relay-destination channel condition is. Then $a_{i,j} \in \mathcal{A}$ is used to describe the action performed by an SU with reputation i, $i \in \{1, ..., N\}$, under source-relay-destination channel condition j, $j \in \{1, ..., N\}$. Here, without loss of generality, we assume that both the SU's reputation and source-relay-destination channel quality are quantized to N levels and we will discuss how to acquire them in subsection II-C and II-D, respectively.

B. Social Norm: How to Assign Reputation

A social norm $\mathbf{\Omega}$ is a matrix used for assigning the immediate reputation of players. In a social norm $\mathbf{\Omega}$, the element $\Omega_{i,j}$ represents the reputation assigned to an SU immediately after it takes an action i based on the source-relay-destination

[1]Since whether an SU will be selected only depends on its channel condition, there is no competition among SUs.

channel quality j. In our model, we assume that all SUs in the system share the same norm.

With the DF protocol, the SNR at the BS can be written as [31]

$$SNR_{DF} = \frac{P_r|h_{rd}|^2}{\sigma_n^2}, \tag{1}$$

where P_r is the relay power, h_{rd} is the channel coefficient of R-D channel, and σ_n^2 is the variance of additive Gaussian white noise in R-D channel.

From (1), we can see that the more power an SU uses, the higher SNR the BS achieves. Consequently, in our model, $\Omega_{i,j}$ is defined as

$$\Omega_{i,j} = \begin{cases} 1, & j = 1, \\ i, & j \neq 1, \end{cases} \tag{2}$$

and the corresponding social norm is

$$\mathbf{\Omega} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 2 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & N & \cdots & N \end{pmatrix}. \tag{3}$$

The social norm (3) is designed to encourage an SU to relay the PU's signal using a higher power by assigning the SU a larger reputation as long as the channel is not in an outage.

C. Power Level and Relay Power

Based on the social norm, the BS will assign a reputation to an SU according to its action, i.e., the relay power level it used. In the following, we will discuss how to quantize the relay power to N levels based on the outage probability. Similar analysis can be done for other QoS measurements such as bit error rate (BER).

Assuming that all S-R, R-D and S-D channels are Rayleigh fading channels where the channel coefficient is complex Gaussian, i.e., $h_{sr} \sim \mathcal{CN}\left(0, \sigma_{sr}^2\right)$, $h_{rd} \sim \mathcal{CN}\left(0, \sigma_{rd}^2\right)$ and $h_{sd} \sim \mathcal{CN}\left(0, \sigma_{sd}^2\right)$, and σ_n^2 is the same for all S-R, R-D and S-D channels. Then, given the transmission rate threshold R, the outage probability at the BS with DF relay protocol [39] can be derived as shown in (4), where $\lambda_u = P_s\sigma_{sd}^2/\sigma_n^2$, $\lambda_v = P_r\sigma_{rd}^2/\sigma_n^2$, $\omega = 2^{2R}-1$, and P_s is the PU's transmission power.

Since p^{out} is a monotonic function of P_r, if we quantize the required outage probability interval $[p_{min}^{out}, p_{max}^{out}]$ to $N-2$ levels $\mathcal{P}^{out} = \{p_1^{out}, p_2^{out}, ..., p_{N-2}^{out}\}$ with $p_1^{out} = p_{max}^{out}$ and $p_{N-2}^{out} = p_{min}^{out}$, then $\forall p_i^{out} \in \mathcal{P}^{out}$, the actual power an SU should use can be obtained as

$$P_{r_{i+1}} = F^{-1}(p_i^{out}), \tag{5}$$

where F^{-1} is the inverse function of p^{out} in (4).

For $p^{out} > p_{max}^{out}$, we quantize it to the 1th level and set $P_{r_1} = 0$, and for $p^{out} < p_{min}^{out}$, we quantize it to the Nth level and set $P_{r_N} = P_{max}$. In such a way, we successfully obtain the N power levels and their corresponding transmission power.

$$p^{out} = \begin{cases} 1 - [(\frac{\lambda_v}{\lambda_v - \lambda_u})\exp(-\lambda_u \omega) + (\frac{\lambda_u}{\lambda_u - \lambda_v})\exp(-\lambda_v \omega)], & \lambda_u \neq \lambda_v, \\ 1 - (1 + \lambda \omega)\exp(-\lambda \omega), & \lambda_u = \lambda_v = \lambda, \end{cases} \qquad (4)$$

D. Channel Quality Distribution

The quality of source-relay-destination channel is also quantized to N levels, i.e., 1, 2, ..., N. Here the level "1" means the channel has encountered an outage while the level "2" to the level "N" stand for different channel qualities, i.e., level "N" being the best channel quality. Next, we will discuss how to acquire such N levels of the channel quality and the channel quality distribution $\mathbf{q} = [q_1, q_2, ..., q_N]$ with the kth element q_k, $k \in \{1, 2, ..., N\}$, being the probability of channel quality at the level k.

The channel quality distribution depends on the relay selection strategy. In our work, an SU which has the "best" source-relay-destination channel will be selected according to the relay selection strategy in [32]. In other words, denote the relay metric of SU$_i$, h_i, as

$$h_i = \frac{2\beta_1 \beta_2 |h_{sr,i}|^2 |h_{rd,i}|^2}{\beta_1 |h_{rd,i}|^2 + \beta_2 |h_{sr,i}|^2}, \qquad (6)$$

where β_1 and β_2 are two parameters defined in (5) in [32], then the SU with the maximum metric $h^* = \max_i h_i$ is the one with the "best" source-relay-destination channel in the secondary system.

Suppose there are M SUs in the secondary system and the S-R channel and the R-D channel for all SUs are independent identical distribution Rayleigh fading channels with complex Gaussian channel coefficient, i.e., $\forall i \in \{1, 2, ..., M\}$, $h_{sr,i} \sim \mathcal{CN}(0, \sigma_{sr}^2)$ and $h_{rd,i} \sim \mathcal{CN}(0, \sigma_{rd}^2)$. In such a case, the probability distribution function (PDF) of h^* can be derived according to [32]. Finally, by dividing the domain of h^* into N intervals we get N levels of channel quality, and by integrating the PDF of h^* over each interval, we obtain the channel quality distribution $\mathbf{q}$.

III. PROPOSED ALGORITHM AND ANALYSIS

In this section, we first propose a reputation updating policy by incorporating the impact of channel noise on SUs' actions. Then, we analyze the uniqueness of stationary reputation distribution of the whole population. Finally, we formulate the cooperation problem as an indirect reciprocity game, characterize the equilibrium of the proposed game, and derive the stable condition for the optimal action rule.

A. Reputation Updating Policy

As we have discussed in Section II, after an SU finishes relaying, it will be assigned a reputation, based on which it can apply for the usage of vacant channels in the future. In order to establish the reputation of an SU with the social norm, we develop a reputation updating policy as shown in Fig. 2. Different from the reputation updating policy in [30] which considered the reputation propagation, our proposed model does not have reputation propagation error since the reputation is assigned by and stored in the BS. However,
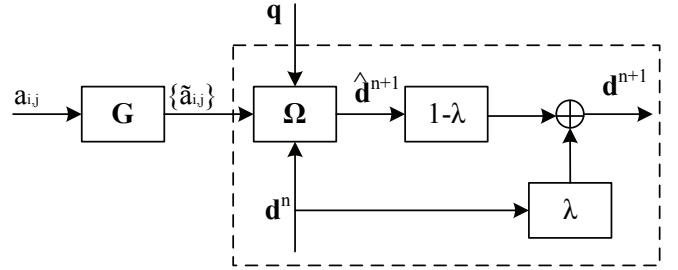

Fig. 2. Reputation updating policy.

the noise in channels between SUs and the BS may distort the SUs' actions, affect the detection accuracy of the BS and hence influence the SUs' reputation updating. Therefore, in our reputation updating policy, we define $\mathbf{G}$, which will be discussed in the next subsection, as the power detection transition matrix to describe the impact of R-D channels on SUs' actions.

To capture all the likelihoods of the SU's reputation after it takes action $a_{i,j}$, we assign for the SU a reputation distribution $\mathbf{d}(a_{i,j}) = [d_1(a_{i,j}), d_2(a_{i,j}), ..., d_N(a_{i,j})]^T$ where $d_k(a_{i,j})$, $k \in \{1, 2, ..., N\}$, denotes the probability of the SU's being assigned with reputation k. Due to the channel noise, it is impossible for the BS to perfectly recognize the power level an SU used. In such a case, $a_{i,j}$ may be falsely detected as $\tilde{a}_{i,j}$. Consequently, for an SU which performs action $a_{i,j}$ at time index $n + 1$, it will be assigned an immediate reputation $\hat{\mathbf{d}}^{n+1} = \mathbf{d}(\tilde{a}_{i,j})$ according to the social norm. Finally, the BS updates the SU's reputation distribution at time index $n + 1$, $\mathbf{d}^{n+1}$, using a linear combination of the SU's original reputation distribution $\mathbf{d}^n$ and the immediate reputation distribution $\hat{\mathbf{d}}^{n+1}$ with a weight λ. Here, the weight λ can be treated as a discount factor of the past reputation.

In a simple case, an SU's original reputation distribution is $\mathbf{e}_i$, where $\mathbf{e}_i$ is the standard basis vector. According to our reputation updating policy in Fig. 2, after the SU takes action $a_{i,j}$, its reputation distribution $\mathbf{d}(a_{i,j})$ can be updated as

$$\mathbf{d}(a_{i,j}) = (1 - \lambda) \sum_{l=1}^{N} \left[\mathbf{e}_{\Omega_{l,j}} \cdot G_{a_{i,j},l} \right] + \lambda \mathbf{e}_i, \qquad (7)$$

where $G_{a_{i,j},l}$ is the probability of the SU's action $a_{i,j}$ being detected as l, and $\Omega_{l,j}$ is the immediate reputation assigned to the SU when the BS considers that the SU takes an action l with channel quality j.

B. Power Detection and Power Detection Transition Matrix

In this subsection, we will discuss in details how the BS estimates the power level an SU uses through energy detection. Let x_i, $i = 1, 2, \ldots L$, be the signals modulated with ASK, PSK or FSK and broadcasted by a PU in L frames. After relayed by an SU with the DF protocol, the signals received

by the BS are

$$y_i = \sqrt{P_r} h_{rd} \hat{x}_i + n_i, \ i = 1, 2, \ldots, L, \qquad (8)$$

where $\hat{x}_i$ is the signal decoded by SU, $E\left(\hat{x}_i^2\right) = \sigma_x^2$ and n_i is channel noise. After coherent demodulation and being matched with the constellation map, the value of $\hat{x}_i$ can be obtained, based on which the BS can further estimate the relay power the SU used.

Suppose that both the channel state information of the R-D channel and the relay power the SU used remain unchanged in L frames. To improve the accuracy of estimation, the BS combines L received signals y_i, $i = 1, 2, \ldots, L$, using Maximal Ratio Combining (MRC) and gets

$$z = \sqrt{P_r} |h_{rd}|^2 ||\mathbf{x}||^2 + h_{rd}^* \mathbf{x}^H \mathbf{n}, \qquad (9)$$

where h_{rd}^* is the complex conjugate of h_{rd}, $\mathbf{x} = [x_1, x_2, ..., x_L]$ and $\mathbf{n} = [n_1, n_2, ..., n_L]$.

From (9), we can see that z follows a Gaussian distribution with the mean $E[z] = L\sqrt{P_r}\sigma_{rd}^2\sigma_x^2$ and variance $Var[z] = L\sigma_{rd}^2\sigma_x^2\sigma_n^2$. Therefore, for each of the N possible power levels, i.e., $P_{r_i} \in \mathcal{P}_r = \{P_{r_i}, i = 1, 2, ...N\}$, let $B_i = \sqrt{P_{r_i}}$, $B_i \in \mathcal{B} = \{B_i, i = 1, 2, ...N\}$, we have

$$z_i \sim \mathcal{N}\left(L\sigma_{rd}^2\sigma_x^2 B_i, \ L\sigma_{rd}^2\sigma_x^2\sigma_n^2\right) = \mathcal{N}\left(\mu_i, \sigma^2\right), \qquad (10)$$

where $\mu_i = L\sigma_{rd}^2\sigma_x^2 B_i$ and $\sigma^2 = L\sigma_{rd}^2\sigma_x^2\sigma_n^2$.

Since all z_is are Gaussian, the optimal detection threshold would be $v_i = \frac{\mu_i + \mu_{i+1}}{2}$, $i = 1, 2, ...N - 1$. With the optimal thresholds, we construct the power detection transition matrix $\mathbf{G} = (G_{i,j} | \forall i, j)$ with $G_{i,j}$, the probability of an SU using power level i to relay while being detected as power level j, being defined as follows

$$G_{i,j} = \begin{cases} 1 - Q(\frac{\mu_1 + \mu_2 - 2\mu_i}{2\sigma}), & j = 1, \\ Q(\frac{\mu_{j-1} + \mu_j - 2\mu_i}{2\sigma}) - Q(\frac{\mu_j + \mu_{j+1} - 2\mu_i}{2\sigma}), & 1 < j < N, \\ Q(\frac{\mu_{N-1} + \mu_N - 2\mu_i}{2\sigma}), & j = N, \end{cases} \qquad (11)$$

where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp(-\frac{u^2}{2}) du$ is the Q-function.

The power detection transition matrix $\mathbf{G}$ in (11) has two nice properties as described below in Lemma 1 and Lemma 2, and these two properties will be used in later analysis.

Lemma 1: $G_{N,j} < G_{i,j}$, $\forall 1 \le i < N$ and $1 \le j < i$.

Proof: See proof in the appendix. ∎

Lemma 2: $G_{j,j} > G_{i,j}$, $\forall 1 \le i, j \le N$ and $i \ne j$.

Proof: See proof in the appendix. ∎

C. Stationary Reputation Distribution

Let $\boldsymbol{\eta} = [\eta_1, \eta_2, ..., \eta_N]$ be the reputation distribution of the entire population, where η_i, $i \in \{1, 2, ..., N\}$, is the portion of the population that has reputation i. After the transmission, the reputation of the SU is updated according to the reputation updating policy and the new reputation distribution of the entire population $\boldsymbol{\eta}^{new}$ can be computed by

$$\boldsymbol{\eta}^{new} = (1 - \lambda)\boldsymbol{\eta}\mathbf{P}_t + \lambda\boldsymbol{\eta}, \qquad (12)$$

where $\mathbf{P}_t$ is the reputation transition matrix. Each element $P_{t_{i,j}}$ in $\mathbf{P}_t$ stands for the probability of reputation i turning into j after the SU has taken an action. Recalling that q_k,

$k \in \{1, 2, ..., N\}$, is the probability of channel quality at the level k, according to Fig. 2, $P_{t_{i,j}}$ can be calculated by

$$P_{t_{i,j}} = \sum_{k,l:\Omega_{l,k}=j} G_{a_{i,k}^*,l} q_k, \qquad (13)$$

where $a_{i,k}^*$ is the optimal action an SU with reputation i takes under channel condition k.

In the steady state, the new reputation distribution $\boldsymbol{\eta}^{new}$ should be equal to $\boldsymbol{\eta}$, i.e., $\boldsymbol{\eta}^{new} = \boldsymbol{\eta}$. Therefore, if there exists a stationary reputation distribution $\boldsymbol{\eta}^*$, it must be the solution to the following equation

$$(1 - \lambda)\boldsymbol{\eta}^*\mathbf{P}_t + \lambda\boldsymbol{\eta}^* = \boldsymbol{\eta}^*, \qquad (14)$$

which is equivalent to

$$\boldsymbol{\eta}^*\mathbf{P}_t = \boldsymbol{\eta}^*. \qquad (15)$$

In the following Theorem 1, we will show that there is a unique stationary reputation distribution.

Theorem 1: There exists a unique stationary reputation distribution of the whole population for any given optimal action rule, and the stationary reputation distribution is $\mathbf{P}_t$'s eigenvector with the corresponding eigenvalue one.

Proof: For any given optimal action rule $a_{i,j}^*$, we have

$$P_{t_{i,j}} = \sum_{k,l:\Omega_{l,k}=j} G_{a_{i,k}^*,l} q_k > 0, \qquad \forall i, j. \qquad (16)$$

Moreover, we have $\sum_j P_{t_{i,j}} = 1$. Therefore, $\mathbf{P}_t$ is a stochastic matrix with strictly positive entries. According to the Perron−Frobenius theorem [33] [34], there exists a unique eigenvector of $\mathbf{P}_t$ corresponding to eigenvalue 1. Such an eigenvector is the stationary reputation distribution. ∎

D. Payoff Function and Equilibrium of Indirect Reciprocity Game

In this subsection, we will discuss how an SU chooses the optimal action. Intuitively, it will cooperate only if cooperation can bring it benefits. By helping relay the PU's data, the SU will incur a cost that is determined by power consumption, i.e., the power level $a_{i,j}$. Here, we assume a linear cost function as follows

$$C(a_{i,j}) = c(a_{i,j} - 1), \qquad (17)$$

where c is the price per unit power level.

On the other hand, the SU will benefit from the allocation of certain vacant channels in the future based on its newest reputation after relaying. Suppose there are Y channels and let p be the probability of a channel being unoccupied, the expected amount of vacant channels is $N_v = pY$. In our model, the BS allocates the vacant channels to all M SUs in a proportionally fair manner. In such a case, the number of vacant channels $t(a_{i,j})$ that an SU will be allocated after it takes action $a_{i,j}$ is

$$t(a_{i,j}) = N_t \frac{\sum_{n=1}^N n \cdot d_n(a_{i,j}) - 1}{\sum_{n=1}^N n \cdot \eta_n^* - 1}, \qquad (18)$$

where $N_t = N_c N_v / M$ and N_c is the number of SUs one channel is allowed to accommodate, $\sum_{n=1}^N n \cdot d_n(a_{i,j})$ is the

Algorithm 1 : Find the Optimal Action Using Value Iteration

1. Given the tolerance $\varepsilon = 0.01$ and set $\varepsilon_1 = 1$, and initialize a^* with a^0.
2. While $\varepsilon_1 > \varepsilon$
   - Set $\varepsilon_2 = 1$
   - Initialize $W_{i,j} = 0, \forall i, \forall j$.
   - While $\varepsilon_2 > \varepsilon$
     – Compute reputation distribution $\mathbf{d}(a_{i,j})$ using (7).
     – Find the stationary reputation distribution $\hat{\boldsymbol{\eta}}^*$ using (15) and Theorem 1.
     – Calculate the number of vacant channels $t(a_{i,j})$ using (18).
     – Obtain $\hat{W}_{i,j}$ based on the stationary reputation distribution $\hat{\boldsymbol{\eta}}^*$ using (20).
     – Find the optimal action $\hat{a}^*_{i,j} = \arg \max_{a_{i,j}} \hat{W}_{i,j}$.
     – Update the parameter ε_2 by $\varepsilon_2 = ||\hat{W}_{i,j} - W_{i,j}||^2$.
     – Update $W_{i,j}$ with $W_{i,j} = \hat{W}_{i,j}$.
     – End
   - Update the parameter ε_1 by $\varepsilon_1 = ||\hat{a}^* - a^*||^2$.
   - Update a^* with $a^* = \hat{a}^*$.

End

SU's expected reputation, and $\sum_{n=1}^{N} n \cdot \eta_n^*$ is the expected reputation of the whole system. Note that with $\boldsymbol{\eta}^*$, the stationary reputation distribution of the entire population, the impact of other users' reputations have been incorporated into $t(a_{i,j})$. Suppose the gain from accessing one channel is a constant g, then the total gain of an SU taking action $a_{i,j}$ is

$$T(a_{i,j}) = P_v g t(a_{i,j}), \tag{19}$$

where $P_v = 1 - (1-p)^Y$ is the probability of BS having vacant channels.

Let $\mathbf{W} = (W_{i,j} | i = 1, 2, ..., N, \ j = 1, 2, ..., N)$ with $W_{i,j}$ denoting the maximum payoff an SU, currently with reputation i and taking action $a_{i,j}$ under channel condition j, can obtain from this iteration to future. Then, we have

$$W_{i,j} = \max_{a_{i,j} \in \mathcal{A}} - c(a_{i,j} - 1) + P_v g t(a_{i,j})$$
$$+ \gamma \sum_k \sum_l d_k(a_{i,j}) q_l W_{k,l}. \tag{20}$$

The first term in (20) is the immediate cost an SU incurs by taking action $a_{i,j}$ as denoted in (17), the second term is the immediate gain it may obtain with a certain probability as denoted in (19), and the last term is the long-term benefit it receives in the future with a discount factor γ. From the definition of $W_{i,j}$ we can see that an SU's maximum payoff is not only related to its own reputation and channel condition, but also related to other players' actions due to $t(a_{i,j})$ defined in (18).

Note that (20) is a Bellman equation of $W_{i,j}$ [35] and thus finding the optimal action rule $a^*_{i,j}$ is an MDP. In the following, we adopt a modified value iteration algorithm to solve the MDP and the details of the algorithm are listed in Algorithm 1. From Algorithm 1, we can see that when finding

$a^*_{i,j}$ for a specific SU, all other SUs have used the optimal actions. In such a case, $a^*_{i,j}$ is the Nash equilibrium according to the definition in [14].

By adjusting the initial conditions in Algorithm 1, we find two possible optimal actions as follows.

$$a_1^* = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}$$

$$\text{and} \quad a_2^* = \begin{pmatrix} 1 & N & \cdots & N \\ 1 & N & \cdots & N \\ \vdots & \vdots & \ddots & \vdots \\ 1 & N & \cdots & N \end{pmatrix}. \tag{21}$$

Recalling that "1" denotes that an SU uses zero power to relay, the first optimal action a_1^* is a non-cooperative equilibrium, where SUs will not act as relays regardless their own reputations and the channel qualities. Obviously, a_1^* is a bad strategy since, with such a strategy, SUs always deny cooperation and thus their payoffs are zero. The second optimal action a_2^* is the desired equilibrium where SUs always use their maximum power to relay PUs' data unless the source-relay-destination channels encounter an outage.

Given the optimal action rule a_2^* in (21), the payoff function $W_{i,j}$ in (20) has an important property as described below in Lemma 3 and will be used in later analysis.

Lemma 3: Given the optimal action rule a_2^* and the corresponding stationary reputation distribution $\boldsymbol{\eta}^* = [\eta_1^*, \ \eta_2^*, \ ..., \eta_N^*]$, we have

$$[W_{i,:} - W_{k,:}]\mathbf{q}^T = W_{i,j} - W_{k,j} = \frac{P_v g \lambda N_t}{\bar{\eta}(1 - \gamma\lambda)}(i - k), \tag{22}$$

where $\bar{\eta} = \sum_{n=1}^{N} n \cdot \eta_n^* - 1$, $\mathbf{q} = [q_1, \ q_2, \ ..., q_N]$ and $W_{k,:} = [W_{k,1}, \ W_{k,2}, \ ..., W_{k,N}]$.

Proof: For any $a^*_{i,j}$ and the corresponding stationary reputation distribution $\boldsymbol{\eta}^* = [\eta_1^*, \ \eta_2^*, \ ..., \eta_N^*]$, we have

$$W_{i,j} = -c(a_{i,j}^* - 1) + P_v g t(a_{i,j}^*)$$
$$+ \gamma \sum_k \sum_l d_k(a_{i,j}^*) q_l W_{k,l}. \tag{23}$$

- If $j = 1$, $a^*_{i,j} = 1, \forall i$, we have

$$\mathbf{d}(a_{i,j}^*) = (1 - \lambda)\mathbf{e}_1 + \lambda \mathbf{e}_i, \tag{24}$$

and

$$t(a_{i,j}^*) = \frac{(i-1)\lambda N_t}{\bar{\eta}}. \tag{25}$$

Then, by substituting (24) and (25) into (23), we have

$$W_{i,j} = P_v g \frac{(i-1)\lambda N_t}{\bar{\eta}}$$
$$+ \gamma(1-\lambda)W_{1,:}\mathbf{q}^T + \gamma\lambda W_{i,:}\mathbf{q}^T. \tag{26}$$

According to (26), we have

$$W_{i,j} - W_{k,j} = P_v g \frac{\lambda N_t}{\bar{\eta}}(i - k) + \gamma\lambda[W_{i,:} - W_{k,:}]\mathbf{q}^T. \tag{27}$$

- If $j \neq 1$, $a_{i,j}^* = N, \forall i, j$, we have

$$\mathbf{d}(a_{i,j}^*) = (1 - \lambda) G_{N,:} + \lambda \mathbf{e}_i, \quad (28)$$

where $G_{N,:} = [G_{N,1}, \ G_{N,2}, \ ..., G_{N,N}]$, and

$$t(a_{i,j}^*) = N_t \frac{(1 - \lambda) \left(\sum_{n=1}^{N} n \cdot G_{N,n} - 1\right) + \lambda(i - 1)}{\bar{\eta}}. \quad (29)$$

Then, by substituting (28) and (29) into (23), we have

$$W_{i,j} = -c(N - 1) + P_v g \frac{(i - 1)\lambda N_t}{\bar{\eta}}$$
$$+ \frac{P_v g N_t (1 - \lambda)}{\bar{\eta}} (\sum_{n=1}^{N} n G_{N,n} - 1) + \gamma (1 - \lambda)$$
$$\sum_{k=1}^{N} G_{N,k} W_{k,:} \mathbf{q}^T + \gamma \lambda W_{i,:} \mathbf{q}^T. \quad (30)$$

According to (30), we have

$$W_{i,j} - W_{k,j} = P_v g \frac{\lambda N_t}{\bar{\eta}} (i - k) + \gamma \lambda [W_{i,:} - W_{k,:}] \mathbf{q}^T. \quad (31)$$

From (27) and (31), we can see that $W_{i,j} - W_{k,j}$ does not depend on j, which means that entries in the vector $W_{i,:} - W_{k,:}$ are the same. In such a case, let $W_{i,j} - W_{k,j} = x$, then we have

$$x = P_v g \frac{\lambda N_t}{\bar{\eta}} (i - k) + \gamma \lambda x. \quad (32)$$

Therefore, we can obtain the conclusion shown in (22). ∎

E. Stability of Optimal Action Rule

In this subsection, we will discuss the stable condition for a_2^* and we choose the evolutionarily stable strategy (ESS). The concept of ESS [36] can be explained as "a strategy such that, if all members of the population adopt it, then no mutant strategies could invade the population under the influence of natural selection". According to [37], an optimal strategy a^* is an ESS if and only if, $\forall a \neq a^*$, a^* satisfies

- equilibrium condition: $U_i(a, a^*) \leq U_i(a^*, a^*)$, and
- stability condition: if $U_i(a, a^*) = U_i(a^*, a^*)$, $U_i(a, a) < U_i(a^*, a)$,

where U_i denotes the payoff function of player i in the game.

To obtain the stable condition for the optimal action a_2^*, we first denote $\boldsymbol{\delta}(a_{i,j}) = (G_{N,:} - G_{a_{i,j},:}) - (\mathbf{e}_N - \mathbf{e}_{a_{i,j}}) = [\delta_1(a_{i,j}), \delta_2(a_{i,j}), ..., \delta_N(a_{i,j})]$ be the gap vector between $G_{N,:} - G_{a_{i,j},:}$ and $\mathbf{e}_N - \mathbf{e}_{a_{i,j}}$. In the following, we characterize four nice properties of $\boldsymbol{\delta}(a_{i,j})$, which are described in Lemma 4 to Lemma 7 for later analysis.

Lemma 4: The sum of all the elements in $\boldsymbol{\delta}(a_{i,j})$ is zero.

Proof: From the definition of $\boldsymbol{\delta}(a_{i,j})$, we have

$$\sum_i \delta_i(a_{i,j}) = \sum_i (G_{N,i} - G_{a_{i,j},i} - e_{N,i} + e_{a_{i,j},i}) = 0. \quad (33)$$

∎

Lemma 5: The first positive element in $\boldsymbol{\delta}(a_{i,j})$ is the $a_{i,j}^{th}$ element, and the last negative element is the Nth element.

Proof: From Lemma 1 and Lemma 2, we can see that $\forall j \leq a_{i,j} < N$, $G_{N,j} - G_{a_{i,j},j} < 0$, which means all the first

$a_{i,j}$ elements in vector $G_{N,:} - G_{a_{i,j},:}$ are negative. We can also see that the Nth element in vector $G_{N,:} - G_{a_{i,j},:}$ is positive and no larger than 1. After $G_{N,:} - G_{a_{i,j},:}$ subtracts $\mathbf{e}_N - \mathbf{e}_{a_{i,j}}$ and yields $\boldsymbol{\delta}(a_{i,j})$, the $a_{i,j}^{th}$ element in $\boldsymbol{\delta}(a_{i,j})$ becomes positive, the Nth element becomes negative and other elements remain unchanged as in vector $G_{N,:} - G_{a_{i,j},:}$. Therefore, the $a_{i,j}^{th}$ entry in $\boldsymbol{\delta}(a_{i,j})$ is the first positive element while the Nth entry is the last negative one. ∎

Lemma 6: Let $\boldsymbol{\xi} = [1, 2, ..., N]$, then $\boldsymbol{\delta}(a_{i,j})\boldsymbol{\xi}^T > (a_{i,j} - N)(1 - G_{a_{i,j},a_{i,j}} + \sum_{k=a_{i,j}}^{N-1} G_{N,k})$.

Proof: See proof in the appendix. ∎

Lemma 7: $\boldsymbol{\delta}(a_{i,j})\mathbf{W}\mathbf{q}^T > \frac{P_v g \lambda N_t}{\bar{\eta}(1-\gamma\lambda)}(a_{i,j} - N)(1 - G_{a_{i,j},a_{i,j}} + \sum_{k=a_{i,j}}^{N-1} G_{N,k})$.

Proof: See proof in the appendix. ∎

Next, let $F(a_{i,j}, a^*)$ be the payoff an SU can obtain when it deviates to action $a_{i,j}$ for one iteration while others always take the optimal action $a_{i,j}^*$, i.e.,

$$F(a_{i,j}, a^*) = -c(a_{i,j} - 1) + P_v g t(a_{i,j})$$
$$+ \gamma \sum_k \sum_l d_k(a_{i,j}) q_l W_{k,l}. \quad (34)$$

Then, according to the one shot deviation principle for MDP [38] and the definition of ESS, a^* is an ESS if the following inequality holds,

$$F(a_{i,j}^*, a^*) > F(a_{i,j}, a^*), \qquad \forall a_{i,j}. \quad (35)$$

According to (35), we can derive the stable condition for optimal action a_2^* as described in Theorem 2.

Theorem 2: The optimal action a_2^* is an ESS if the cost-to-gain ratio c/g satisfies $0 < c/g < \frac{P_v N_t (1-\lambda)}{\bar{\eta}(1-\gamma\lambda)} (\min_{a_{i,j} \in \mathcal{A}} G_{a_{i,j},a_{i,j}} + G_{N,N} - 1)$.

Proof: (1) If $j = 1$, $a_{i,j}^* = 1$, $\forall i$. Then $\forall a_{i,j} \neq a_{i,j}^*$, we have

$$\mathbf{d}(a_{i,j}) = (1 - \lambda) \mathbf{e}_1 + \lambda \mathbf{e}_i, \quad (36)$$

and

$$t(a_{i,j}) = \frac{(i - 1)\lambda N_t}{\bar{\eta}}. \quad (37)$$

By substituting (24), (25), (36) and (37) into (34), we have

$$F(a_{i,j}^*, a^*) - F(a_{i,j}, a^*) = -c(a_{i,j}^* - a_{i,j}). \quad (38)$$

Since $a_{i,j}^* - a_{i,j} < 0$, in order for (35) to be held, we should have

$$c > 0. \quad (39)$$

(2) If $j \neq 1$, $a_{i,j}^* = N$, $\forall i$. Then $\forall a_{i,j} \neq a_{i,j}^*$, we have

$$\mathbf{d}(a_{i,j}) = (1 - \lambda) G_{a_{i,j},:} + \lambda \mathbf{e}_i, \quad (40)$$

and

$$t(a_{i,j}) = N_t \frac{(1 - \lambda) \left(\sum_{n=1}^{N} n \cdot G_{a_{i,j},n} - 1\right) + \lambda(i - 1)}{\bar{\eta}}. \quad (41)$$

Since $\boldsymbol{\delta}(a_{i,j}) = (G_{N,:} - G_{a_{i,j},:}) - (\mathbf{e}_N - \mathbf{e}_{a_{i,j}})$, by substituting (28), (29), (40) and (41) into (34), we have (42), where the inequality above uses the results of Lemma 6 and Lemma 7.

$$
\begin{aligned}
&F(a_{i,j}^*, a^*) - F(a_{i,j}, a^*) \\
={}& -c(N - a_{i,j}) + P_v g \frac{N_t(1-\lambda)}{\bar{\eta}}(\mathbf{e}_N - \mathbf{e}_{a_{i,j}})\boldsymbol{\xi}^T \\
&+ \gamma(1-\lambda)(\mathbf{e}_N - \mathbf{e}_{a_{i,j}})\mathbf{W}\mathbf{q}^T + P_v g \frac{N_t(1-\lambda)}{\bar{\eta}}\boldsymbol{\delta}(a_{i,j})\boldsymbol{\xi}^T + \gamma(1-\lambda)\boldsymbol{\delta}(a_{i,j})\mathbf{W}\mathbf{q}^T, \\
>{}& -c(N - a_{i,j}) + P_v g \frac{N_t(1-\lambda)}{\bar{\eta}}(N - a_{i,j}) + \gamma(1-\lambda)\frac{P_v g \lambda N_t}{\bar{\eta}(1-\gamma\lambda)}(N - a_{i,j}) \\
&+ P_v g \frac{N_t(1-\lambda)}{\bar{\eta}}(1 - G_{a_{i,j},a_{i,j}} + \textstyle\sum_{k=a_{i,j}}^{N-1} G_{N,k})(a_{i,j} - N) \\
&+ \gamma(1-\lambda)\frac{P_v g \lambda N_t}{\bar{\eta}(1-\gamma\lambda)}(1 - G_{a_{i,j},a_{i,j}} + \textstyle\sum_{k=a_{i,j}}^{N-1} G_{N,k})(a_{i,j} - N), \\
={}& -c(N - a_{i,j}) + \frac{P_v g N_t(1-\lambda)}{\bar{\eta}(1-\gamma\lambda)}(G_{a_{i,j},a_{i,j}} - \textstyle\sum_{k=a_{i,j}}^{N-1} G_{N,k})(N - a_{i,j}),
\end{aligned}
\tag{42}
$$

Since $1 \le a_{i,j} < N$, $\forall a_{i,j}$, in order for (35) to be held, we should have

$$
c < \frac{P_v g N_t(1-\lambda)}{\bar{\eta}(1-\gamma\lambda)}(G_{a_{i,j},a_{i,j}} - \sum_{k=a_{i,j}}^{N-1} G_{N,k}).
\tag{43}
$$

According to (39) and (43), the sufficient condition for a_2^* to be an ESS is

$$
0 < c/g < \frac{P_v N_t(1-\lambda)}{\bar{\eta}(1-\gamma\lambda)}(\min_{a_{i,j}\in\mathcal{A}} G_{a_{i,j},a_{i,j}} + G_{N,N} - 1).
\tag{44}
$$

■

From Theorem 2, we can see that in a society where most of the SUs have high reputation, i.e., $\bar{\eta} \approx N - 1$, if the BS has an acceptable power detection accuracy (e.g. $\min_{a_{i,j}\in\mathcal{A}} G_{a_{i,j},a_{i,j}} = 0.9$), the vacant channel probability of the PU system is high enough (e.g. $P_v = 0.8$), the SUs are very likely to continue staying in the system (e.g. $\gamma = 0.9$), an appropriate weight on the past reputation is set (e.g. $\lambda = 0.5$) and a large portion of SUs that can be accommodated in the vacant channels is guaranteed (e.g. $N_t = 0.8$), then c/g has a reasonable upper bound (e.g. 0.13 with N=5 power levels). In such a case, the condition for the optimal action a_2^* to be an ESS is easy to be satisfied in a practical system.

IV. SIMULATION

In this section, we first examine the evolutionary stability of the optimal action a_2^* as described in (21). Then, we evaluate the effectiveness of the proposed method by comparing with the state-of-the-art dynamic spectrum access methods. We finally study the characters of different social norms and compare their performance.

A. Evolutionary Stability of Optimal Action a_2^*

To evaluate the evolutionary stability of optimal action a_2^*, we study a scenario including $Y = 500$ PUs and $M = 1000$ SUs. At the beginning, each SU is assigned an initial reputation which is randomly chosen from $\{1, 2, ..., N\}$ with equally probability $\frac{1}{N}$. Mean while, each SU decides the action it will take. Taking into account the perturbation effect caused by the uncertainty of the system and/or noisy parameters, not all SUs will take the optimal action. In our simulation, 60% SUs will choose to use optimal action shown in (21) while the others choose other action rules randomly. Other parameters, i.e., discount factor λ, vacant channel probability p and the number of SUs allowed to access a channel simultaneously N_c, are set to be 0.5, 0.8 and 1, respectively.
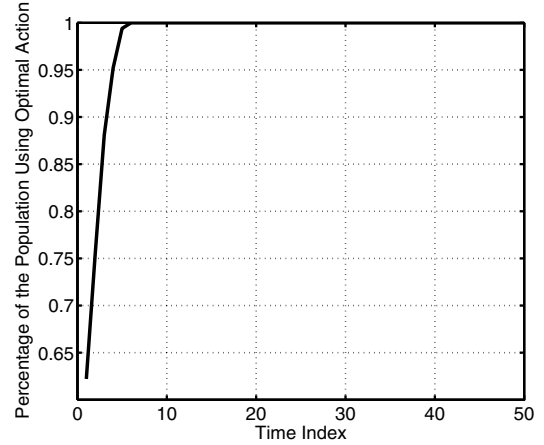
Fig. 3. The percentage of the population using optimal action a_2^* shown in (21).

Before the step of action updating, each SU performs the action it has chosen 10 times under different R-D channel conditions. After 10 actions, the SU gains an average payoff and gets a new reputation according to the reputation updating policy show in Fig. 2. We assume that all SUs in the system use a fix social norm show in (3) and agree on the reputation generated by the reputation updating policy. Then the SU stays in the system with probability $\gamma = 0.97$ or leaves the system with probability $1 - \gamma$. For every SU who leaves, a new SU enters the system to keep the total population size constant. The initial reputation of the new comer is randomly chosen from $\{1, 2, ..., N\}$ with equally probability $\frac{1}{N}$. All the SUs, including those staying in the system and those entering the system, need to choose their actions. Here, we adopt Wright-Fisher model [40], which is by far the most popular stochastic model for reproduction in population genetics, to investigate how the SUs update their actions. According to the Wright-Fisher model, the percentage of the population using action a_i at time index $t + 1$, y_i^{t+1}, is proportional to the total payoff of the users using a_i in time index t, i.e.,

$$
y_i^{t+1} = \frac{y_i^t U_i^t}{\sum_{i=1}^{M} y_i^t U_i^t},
\tag{45}
$$

where U_i^t is the total payoff of the users using a_i at time index t.

We show the evolutionary results of the percentage of population using optimal action in Fig. 3. From Fig. 3, we can see that the optimal action a_2^* spreads over the whole system within 10 iterations. Once the whole population adopt
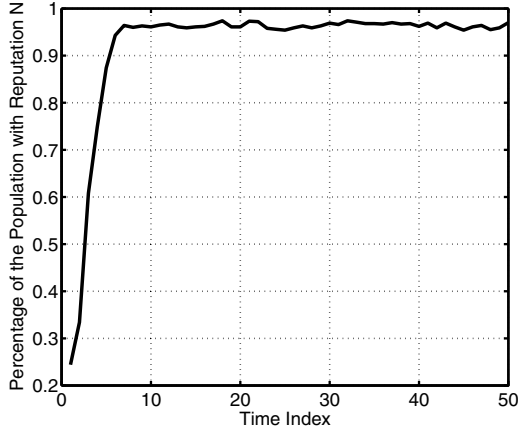
Fig. 4. The percentage of the population with reputation N=4 under social norm in (3).

Fig. 5. Throughput of PU under social norm in (3).

the optimal action, no one will deviate, which verifies that the optimal action a_2^* is an evolutionarily stable strategy under the simulation condition. Accordingly, the percentage of population with reputation N also converges within 10 iterations, as demonstrated in Fig. 4, which proves that action a_2^* is a desirable strategy since it leads to a "good" society with most of the SUs having good reputation.

According to the results we have obtained above, in the following, we only consider the performance of our system when it is in steady stage where most of SUs have the highest reputation.

B. System Performance

To give more insights into our proposed algorithm, we consider the performance of a DSA network with one BS, a single PU and multiple SUs. We assume the PU and BS locate at coordinate (0,0) and (1,0) respectively. There are five SUs randomly located in a circle centered at coordinate (0.5,0) with a radius of 0.25. An SU's receiver is randomly located in a circle centered at the corresponding SU with a radius between 0.1 and 0.25.

The channels are complex Gaussian channels with average power gains: $\sigma_{sd}^2 = 1$, $\sigma_{sr,i}^2 = 1/d_{sr}^\theta$, $\sigma_{rd,i}^2 = 1/d_{rd}^\theta$ and $\sigma_{ss,i}^2 = 1/d_{ss}^\theta$. The path loss factor θ is set to 4 and d_{sr}, d_{rd} and d_{ss} are distances between PU and the selected SU, the selected SU and BS, and SU and its receiver, respectively. The maximum transmission power of PU and SUs are $P_s = P_{r_i} = 10$, and the signal-to-noise ratio (SNR) is $SNR = P_s/\sigma_n = 10$dB.

We evaluate the system performance in terms of total throughput and outage probability by comparing the proposed indirect reciprocity game theoretic scheme with two existing schemes. One is "NCPC", where there is no cooperation between PU and SUs and all SUs can only access unoccupied channels randomly and transmit their own data via a non-cooperative power control game. The second approach is spectrum leasing proposed in [13]. It is a direct reciprocity based scheme under the Stackelberg game theoretic framework where the PU leases a fraction of its spectrum for the cooperation of a group of SUs, and after finishing relaying,
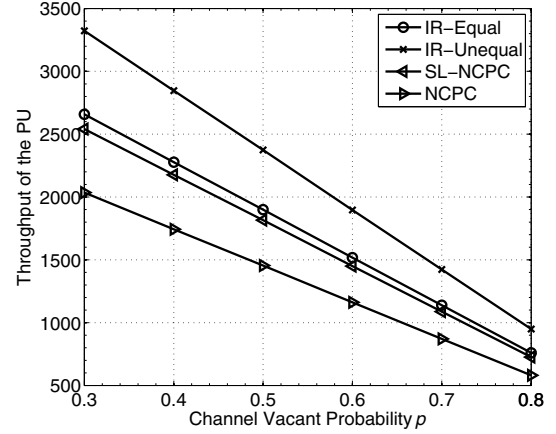
this group of SUs compete with each other for their own data transmission within the leased time slot. However, the authors in [13] only address how to utilize channels occupied by a PU. For fair comparison, we assume all SUs randomly access those vacant channels and transmit their own data via a non-cooperative power control game. We denote this scheme as "SL-NCPC". In SL-NCPC, the SUs with the best performance are chosen. Note that, in the spectrum leasing approach in [13], the SU may use less power to relay than it had promised after it had gained the chance to use the spectrum since there is no supervision and evaluation for its action. Nevertheless, here we assume all SUs are honest and use the power they promise for relaying in SL-NCPC.

In the first simulation, we evaluate the achievable end-to-end throughput of PU versus vacant channel probability p for different schemes. In our scheme, only one SU is selected as relay based on the relay selection strategy in [32], and we assume the signals broadcasted by PUs are modulated with BPSK. Since the selected SU uses the same channel as PU to relay data, we should consider how to allocate the transmission time between the PU and the selected SU. Here we propose two strategies, one is "IR-Equal", where one time frame is equally divided into two time slot, and the other one is "IR-Unequal", where the length of time slot, α, allocated to a PU's transmission is given by the following strategy. Let $R_{sr} = \log_2(1 + P_s|h_{sr}|^2/\sigma_n^2)$, $R_{rd} = \log_2(1 + P_r|h_{rd}|^2/\sigma_n^2)$ and $R_{sd} = \log_2(1 + P_s|h_{sd}|^2/\sigma_n^2)$ denote the achievable rate of S-R, R-D and S-D channel, respectively. To achieve better performance, α is chosen in such a way that the rate at the relay is equal to that at the BS, i.e., $\alpha = R_{rd}/(R_{sr} + R_{sd} - R_{rd})$.

The results of the first simulation are shown in Fig. 5. From Fig. 5, we can see that NCPC has the worst performance. This is because there is no cooperation between the PU and SUs, the PU can only obtain its throughput by direct transmission without any QoS improvement. Through cooperation with SUs using spectrum leasing, SL-NCPC can improve the throughput of PU. However, due to the selfish nature, SUs will use only part of their transmission power for relaying data when playing the spectrum leasing game, which leads to an inefficient cooperation. With the proposed indirect reciprocity game theoretic
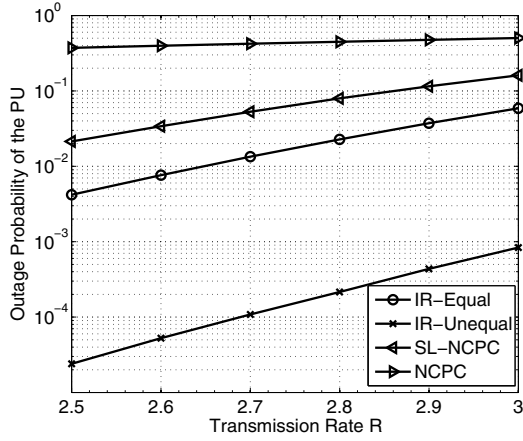
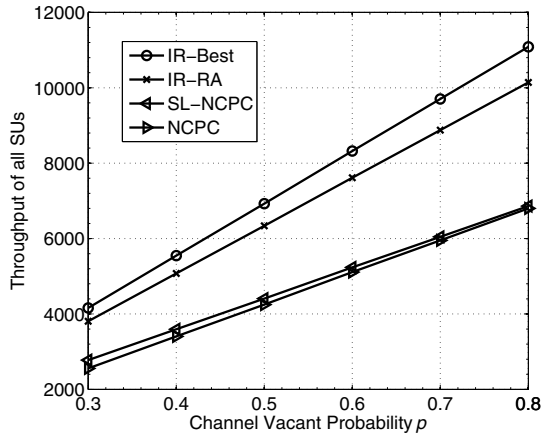Fig. 6. Outage probability of PU under social norm in (3).

Fig. 7. Throughput of all SUs under social norm in (3).

method, the selected SU is motivated to use its full power for relaying the PU's data. In such a case, the PU's throughput can be greatly improved. Furthermore, from Fig. 5, we can see that the performance of the proposed scheme can be further improved with the IR-unequal time slot allocation. We can also see when the vacant channel probability decreases, i.e., the PU has more information to transmit, the performance gap among different schemes increases, which means that cooperation stimulation becomes more important as the load of PU becomes heavier.

We also study the performance of PU system in terms of outage probability. From Fig. 6, we can see that without cooperation, the outage probability of PU with NCPC is as high as 0.5 when R=2.5. With cooperation, the outage probability in SL-NCPC can drop to around 10^{-2}. With the proposed cooperation stimulation framework, the PU system can achieve the lowest outage probability.

In the third simulation, we evaluate the throughput of all SUs versus vacant channel probability and the results are shown in Fig. 7. In our proposed scheme, when SUs apply for the usage of vacant channels, only one SU is selected to access one channel. There are two ways for the BS to select the SU: "IR-RA" where the SU is randomly chosen, and "IR-Best" where the SU with the best channel gain is chosen.

From Fig. 7, we can see that the throughput of all SUs with the proposed scheme is up to 45% larger than that of SL-NCPC and NCPC. This is mainly because in our scheme, only one SU is assigned to access one channel. In such a case, the interference introduced by multiple SUs' transmission over one channel in SL-NCPC and NCPC can be avoided in our scheme. Fig. 7 also illustrates that the throughput of all SUs increases as the vacant channel probability increases. Such a phenomenon is due to the fact that there are more opportunities for SUs to access the vacant channel as the vacant channel probability increases.

C. Different Social Norms

In this subsection, we study the impacts of different social norms on the equilibrium of the indirect reciprocity game and the system performance.

The first social norm, denoted as $\mathbf{\Omega}_1$, is the one we propose in (3). Recalling that $\mathbf{\Omega}_1$ is designed to encourage an SU to relay the PU's signal using a higher power by assigning him a larger reputation when the channel is not in an outage. To avoid blindly penalizing SUs with poor channel quality, a revised social norm, $\mathbf{\Omega}_2$, is proposed as follow, with which the reputation is assigned purely based on the SU's action.

$$\mathbf{\Omega}_2 = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 2 & 2 & \cdots & 2 \\ \vdots & \vdots & \ddots & \vdots \\ N & N & \cdots & N \end{pmatrix}. \tag{46}$$

To avoid using excessive transmission powers for relaying, the following social norm, $\mathbf{\Omega}_3$, is proposed

$$\mathbf{\Omega}_3 = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 2 & 2 & \cdots & 2 \\ \vdots & \vdots & \ddots & \vdots \\ N-2 & N-2 & \cdots & N-2 \\ N & N & \cdots & N \\ N & N & \cdots & N \end{pmatrix}. \tag{47}$$

Another social norm we study is the channel quality dependent social norm, $\mathbf{\Omega}_4$, defined as follows

$$\mathbf{\Omega}_4 = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \\ 1 & 2 & 3 & \cdots & N-1 & N \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ N-2 & N-1 & N & \cdots & N & N \\ N-1 & N & N & \cdots & N & N \end{pmatrix}. \tag{48}$$

With $\mathbf{\Omega}_4$, an SU with bad channel condition may not be able to receive a good reputation even relaying with maximal power. On the other hand, if the channel condition is good enough, a low power level can lead to a high reputation.

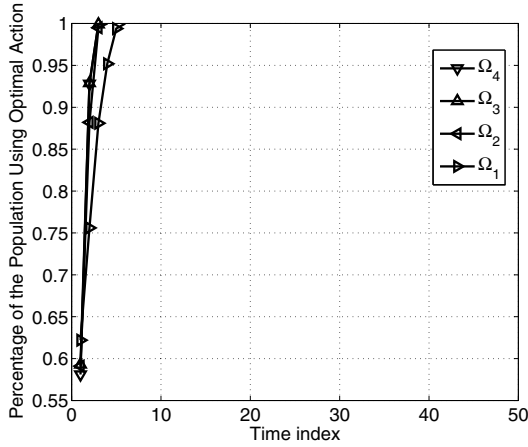With the Algorithm 1 proposed in section III-D, the optimal action rules for SUs with $\mathbf{\Omega}_i$, $i = 2, 3, 4$, can be derived as follows

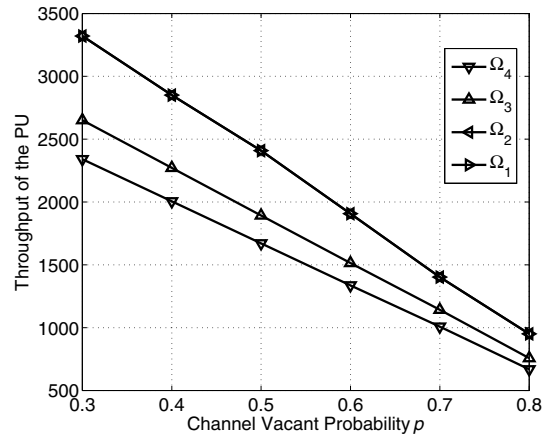Fig. 8.   The percentage of the population using optimal action shown in (49) with different social norms.

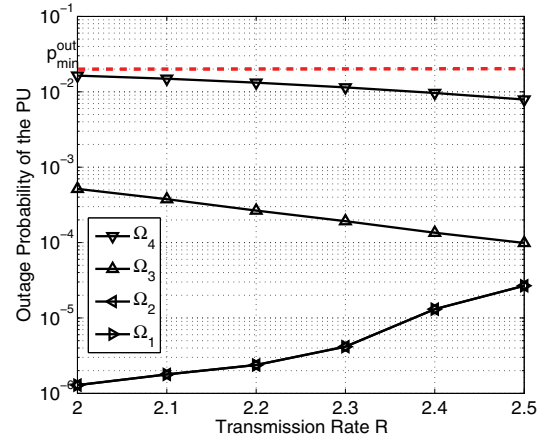Fig. 9.   The percentage of the population with reputation N=5 with different social norms.

$$a^*_{\boldsymbol{\Omega}_2} = \begin{pmatrix} N & \cdots & N \\ N & \cdots & N \\ \vdots & \ddots & \vdots \\ N & \cdots & N \end{pmatrix},$$

$$a^*_{\boldsymbol{\Omega}_3} = \begin{pmatrix} N-1 & \cdots & N-1 \\ N-1 & \cdots & N-1 \\ \vdots & \ddots & \vdots \\ N-1 & \cdots & N-1 \end{pmatrix}$$

$$\text{and} \quad a^*_{\boldsymbol{\Omega}_4} = \begin{pmatrix} N & N & N-1 & \cdots & 2 \\ N & N & N-1 & \cdots & 2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ N & N & N-1 & \cdots & 2 \end{pmatrix}. \quad (49)$$

We then examine the evolutionary stability of the optimal action rules given in (49) with the same setting described in Section IV-A. The simulation results are shown in Fig. 8 and Fig. 9. We can see that both the optimal actions and the percentage of population with reputation N converge within 10 iterations, which verify that all these optimal action rules are ESSs.

Fig. 10.   Throughput of the PU with different social norms.

Fig. 11.   Outage of the PU with different social norms.

Finally, we compare the system performance of the same DSA network as described in Section IV-B under different social norms. In the following, we assume that the IR-unequal strategy is used and the PU's QoS requirement is set to $p^{out}_{min} = 0.02$ and $p^{out}_{max} = 0.2$.

We first evaluate the end-to-end throughput of PU versus vacant channel probability. From Fig. 10, we can see that with $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$, the PU has the highest throughput. That's because under both social norms, SUs use their maximal powers for relaying. With $\boldsymbol{\Omega}_3$, since it is not necessary for SUs to use full power to gain the highest reputation, they decrease their relay powers to the second highest power level, which leads to a degradation of the PU's throughput. With $\boldsymbol{\Omega}_4$, SUs adjust their power levels according to the channel condition. Since the channel quality of most selected SUs is good enough, they will further lower their power levels, due to which the PU gains the lowest throughput.

The simulation results of PU's outage probability versus transmission rate threshold are shown in Fig. 11. From Fig. 11, we can see that with $\boldsymbol{\Omega}_4$, the PU can already meet the QoS requirement with the help of SUs. With $\boldsymbol{\Omega}_3$, the PU's outage probability can be further improved. This phenomenon is mainly due to the fact that the power level of this social norm is quantized based on the average channel condition
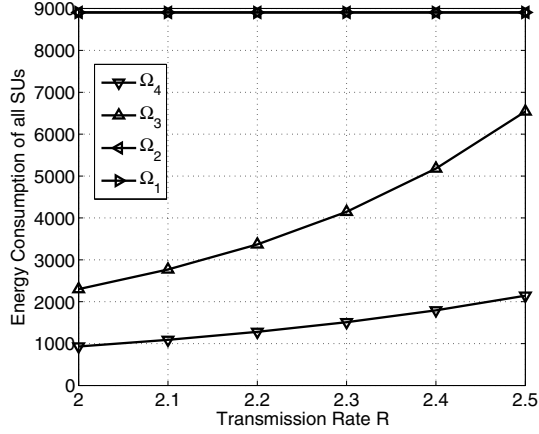
Fig. 12. Energy consumption of all SUs with different social norms.

while the selected SU always has a better channel condition. With the largest relay power, the PU gains the lowest outage probability with Ω_1 and Ω_2.

In the last simulation, we study the performance of the secondary system. Considering that the throughput of all SUs will be the same given a specific vacant channel probability, here we only compare the total energy consumption of all SUs versus transmission rate threshold with the vacant channel probability being $p = 0.55$. From Fig. 12, we can see that to gain the same throughput, SUs need to spend the most energy on relaying with social norms Ω_1 and Ω_2. With Ω_3, they can save at least 26% of the energy. With Ω_4, the energy saving can be up to 90%.

Furthermore, from Fig. 8 to Fig. 12, we can see that there is almost no performance difference between Ω_1 and Ω_2. That's because the probability of the selected SU's channel quality falling into the lowest channel quality level is so low that it is hard for the reputation assignment on this level to make a noticeable difference.

Therefore, among the four social norms, Ω_1 and Ω_2 are preferable to the PU since it can bring better performance, while Ω_4 is the most energy efficient social norm for SUs. Moreover, Ω_1, Ω_2 and Ω_3 are easier for a real system to implement than Ω_4 since the latter requires fine quantization of the channel quality and relay power as well as accurate relationship among them, which is not trivial and needs more theoretical analysis in the future.

V. CONCLUSION

In this paper, we propose a novel spectrum access scheme to efficiently utilize channels based on the indirect reciprocity game modeling. With the proposed indirect reciprocity model, the SUs are stimulated through reputation to use the optimal strategy, i.e., maximal power to help PUs relay data when the channel is not in an outage, which leads to the great improvement of PUs' QoS and the spectrum utilization efficiency. Through theoretical analysis, we derive the cost-to-gain ration under which such an optimal strategy is stable. Simulation results show that the optimal strategy will lead to a "good" society where most of the SUs have the highest reputation, and the proposed scheme achieves much better performance

compared with the state-of-the-art direct reciprocity based spectrum access schemes.

APPENDIX A
PROOF OF LEMMA 1

Lemma 1: $G_{N,j} < G_{i,j}, \forall 1 \le i < N$ and $1 \le j < i$.
   Proof:

- If $j = 1, \forall i < N$, we have

$$G_{N,1} - G_{i,1} = Q(\frac{\mu_1 + \mu_2 - 2\mu_i}{2\sigma}) - Q(\frac{\mu_1 + \mu_2 - 2\mu_N}{2\sigma}). \quad (50)$$

Since $\mu_1 < \mu_2 < ... < \mu_N$, we have $\frac{\mu_1 + \mu_2 - 2\mu_i}{2\sigma} > \frac{\mu_1 + \mu_2 - 2\mu_N}{2\sigma}$. Since $Q(x)$ is monotonically decreasing in x, we have

$$G_{N,1} - G_{i,1} < 0. \quad (51)$$

- If $1 < j < i, \forall i < N$, we have

$$
\begin{aligned}
&G_{N,j} - G_{i,j} \\
&= Q(\frac{\mu_{j-1} + \mu_j - 2\mu_N}{2\sigma}) - Q(\frac{\mu_j + \mu_{j+1} - 2\mu_N}{2\sigma}) \\
&\quad - Q(\frac{\mu_{j-1} + \mu_j - 2\mu_i}{2\sigma}) + Q(\frac{\mu_j + \mu_{j+1} - 2\mu_i}{2\sigma}). \quad (52)
\end{aligned}
$$

Let $a_1 = \frac{\mu_{j-1} + \mu_j - 2\mu_N}{2\sigma}$, $a_2 = \frac{\mu_j + \mu_{j+1} - 2\mu_N}{2\sigma}$, $a_3 = \frac{\mu_{j-1} + \mu_j - 2\mu_i}{2\sigma}$ and $a_4 = \frac{\mu_j + \mu_{j+1} - 2\mu_i}{2\sigma}$. Note $a_1 < a_2 < a_3 < a_4 < 0$ and $a_3 - a_1 = a_4 - a_2$. From the characteristic of Q-function, we have $Q(a_1) - Q(a_3) < Q(a_2) - Q(a_4)$. Hence

$$G_{N,j} - G_{i,j} < 0. \quad (53)$$

In all, $G_{N,j} < G_{i,j}, \forall 1 \le i < N$ and $1 \le j < i$. ∎

APPENDIX B
PROOF OF LEMMA 2

Lemma 2: $G_{j,j} > G_{i,j}, \forall 1 \le i, j \le N$ and $i \ne j$.
   Proof:

- If $j = 1, \forall i \ne j$, we have

$$G_{1,1} - G_{i,1} = Q(\frac{\mu_1 + \mu_2 - 2\mu_i}{2\sigma}) - Q(\frac{\mu_1 + \mu_2 - 2\mu_1}{2\sigma}). \quad (54)$$

Since $\frac{\mu_1 + \mu_2 - 2\mu_i}{2\sigma} < \frac{\mu_1 + \mu_2 - 2\mu_1}{2\sigma}$, we have

$$G_{1,1} - G_{i,1} > 0. \quad (55)$$

- If $1 < j < N, \forall i \ne j$, we have

$$
\begin{aligned}
&G_{j,j} - G_{i,j} \\
&= Q(\frac{\mu_{j-1} + \mu_j - 2\mu_j}{2\sigma}) - Q(\frac{\mu_j + \mu_{j+1} - 2\mu_j}{2\sigma}) \\
&\quad - Q(\frac{\mu_{j-1} + \mu_j - 2\mu_i}{2\sigma}) + Q(\frac{\mu_j + \mu_{j+1} - 2\mu_i}{2\sigma}). \quad (56)
\end{aligned}
$$

Let $b_1 = \frac{\mu_{j-1} + \mu_j - 2\mu_j}{2\sigma}$, $b_2 = \frac{\mu_j + \mu_{j+1} - 2\mu_j}{2\sigma}$, $b_3 = \frac{\mu_{j-1} + \mu_j - 2\mu_i}{2\sigma}$ and $b_4 = \frac{\mu_j + \mu_{j+1} - 2\mu_i}{2\sigma}$. Then we have $b_3 - b_1 = b_4 - b_2$.
If $i > j$, $b_3 < b_1 < b_4 < 0 < b_2$, then $Q(b_4) - Q(b_2) > Q(b_3) - Q(b_1)$, and we have

$$G_{j,j} - G_{i,j} > 0. \quad (57)$$

If $i < j$, $b_1 < 0 < b_3$ and $0 < b_2 < b_4$, then $Q(b_1) - Q(b_3) > Q(b_2) - Q(b_4)$, and we have

$$G_{j,j} - G_{i,j} > 0. \tag{58}$$

- If $j = N$, $\forall i \neq j$

$$G_{N,N} - G_{i,N} = Q(\frac{\mu_{N-1} + \mu_N - 2\mu_N}{2\sigma})$$
$$- Q(\frac{\mu_{N-1} + \mu_N - 2\mu_i}{2\sigma}). \tag{59}$$

Since $\frac{\mu_{N-1}+\mu_N-2\mu_N}{2\sigma} < \frac{\mu_{N-1}+\mu_N-2\mu_i}{2\sigma}$, we have

$$G_{N,N} - G_{i,N} > 0. \tag{60}$$

To summarize, $\forall 1 \leq i$, $j \leq N$ and $i \neq j$, we have $G_{j,j} > G_{i,j}$, i.e., the diagonal entry of matrix G is the maximum element in its column. ∎

APPENDIX C
PROOF OF LEMMA 6

Lemma 6: Let $\boldsymbol{\xi} = [1, 2, ..., N]$, then $\boldsymbol{\delta}(a_{i,j})\boldsymbol{\xi}^T > (a_{i,j} - N)(1 - G_{a_{i,j},a_{i,j}} + \sum_{k=a_{i,j}}^{N-1} G_{N,k})$.

Proof: Let $S_+ = \{k | \delta_k > 0, \delta_k \in \boldsymbol{\delta}(a_{i,j}), k = 1, 2, ..., N\}$ and $S_- = \{k | \delta_k < 0, \delta_k \in \boldsymbol{\delta}(a_{i,j}), k = 1, 2, ..., N\}$. Then $\boldsymbol{\delta}(a_{i,j})\boldsymbol{\xi}^T$ can be calculated as follows

$$\boldsymbol{\delta}(a_{i,j})\boldsymbol{\xi}^T$$
$$= \sum_{k \in S_+} k\delta_k + \sum_{k \in S_-} k\delta_k > \sum_{k \in S_+} a_{i,j}\delta_k + \sum_{k \in S_-} N\delta_k,$$
$$= (a_{i,j} - N)\sum_{k \in S_+} \delta_k > (a_{i,j} - N)(\delta_{a_{i,j}} + \sum_{k=a_{i,j}+1}^{N-1} G_{N,k}),$$
$$= (a_{i,j} - N)(1 - G_{a_{i,j},a_{i,j}} + \sum_{k=a_{i,j}}^{N-1} G_{N,k}), \tag{61}$$

where the first inequality and the second equality come from Lemma 5 and Lemma 4, respectively. The second inequality comes from the fact that $\forall k \in S_+ \backslash a_{i,j}, G_{N,k} > \delta_k$ and $a_{i,j} < N$. ∎

APPENDIX D
PROOF OF LEMMA 7

Lemma 7: $\boldsymbol{\delta}(a_{i,j})\mathbf{W}\mathbf{q}^T > \frac{P_v g\lambda N_t}{\eta(1-\gamma\lambda)}(a_{i,j} - N)(1 - G_{a_{i,j},a_{i,j}} + \sum_{k=a_{i,j}}^{N-1} G_{N,k})$.

Proof: For each positive element δ_k, $k \in S_+$, it can be decomposed into n_k parts with each part having the same value as the negative element in $\boldsymbol{\delta}(a_{i,j})$. Let S_-^k be the set of subscript of those negative elements having the same value of n_k parts of δ_k, then $\boldsymbol{\delta}(a_{i,j})\mathbf{W}\mathbf{q}^T$ can be computed as shown in (62), where the first inequality come from Lemma 5, the last equality comes from Lemma 4, and the last inequality comes from the fact that $\forall k \in S_+ \backslash a_{i,j}, G_{N,k} > \delta_k$ and $a_{i,j} < N$. ∎

REFERENCES

[1] Federal Communications Commission, "Spectrum policy task force, report," FCC Document ET Docket, No. 02-155, Nov. 2002.

[2] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 1, pp. 201–220, Jan. 2005.

[3] J. M. Peha, "Approaches to spectrum sharing," *IEEE Commun. Mag.*, vol 43, no. 2, pp. 10–12, Feb. 2005.

[4] B. Wang and K. J. R. Liu, "Advances in cognitive radios: a survey," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 1, pp. 5–23, Feb. 2011.

[5] A. K. Sadek, K. J. R. Liu, and A. Ephremides, "Cognitive multiple access via cooperation: protocol design and performance analysis," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3677–3695, Oct. 2007.

[6] 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.

[7] L. Cao and H. Zheng, "Distributed rule-regulated spectrum sharing," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 130–145, Jan. 2008.

[8] B. Wang, Z. Ji, K. J. R. Liu, and T.C. Clancy, "Primary-prioritized Markov approach for efficient and fair dynamic spectrum sllocation," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1854–1865, Apr. 2009.

[9] X. Zhou and H. Zheng, "TRUST: a general framework for truthful double spectrum Auctions," in *Proc. 2009 IEEE INFOCOM*, pp. 999–1007.

[10] Y. Wu, B. Wang, K. J. R. Liu, and T. C. Clancy, "A scalable collusion-resistant multi-winner cognitive spectrum auction game," *IEEE Trans. Commun.*, vol. 57, no. 12, pp. 3805–3816, Dec. 2009.

[11] D. Niyato, E. Hossain, and Z. Han, "Dynamics of multiple-seller and multiple-buyer spectrum trading in cognitive radio networks: a game-theoretic modeling approach," *IEEE Trans. Mobile Comp.*, vol. 8, no. 8, pp. 1009–1022, Aug. 2009.

[12] G. E. Howayek and S. K. Jayaweera, "Distributed dynamic spectrum leasing (D-DSL) for spectrum sharing over multiple primary channels," *IEEE Trans. Wireless. Commun.*, vol. 10, no. 1, pp. 55–60, Jan. 2011.

[13] O. Simeone, I. Stanojev, S. Savazzi, Y. Bar-Ness, U. Spagnolini, and R. Pickholtz, "Spectrum leasing to cooperating secondary ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 1–11, Jan. 2008.

[14] M. J. Osborne and A. Rubenstein, *A Course in Game Theory*. MIT Press, 1994

[15] Z. Ji and K. J. R. Liu, "Dynamic spectrum sharing: a game theoretical overview," *IEEE Commun. Mag.*, vol. 45, no. 5, pp. 88–94, May 2007.

[16] K. J. R. Liu and B. Wang, *Cognitive Radio Networking and Security: A game Theoretical View*. Cambridge University Press, 2010.

[17] B. Wang, Y. Wu, Z. Ji, K. J. R. Liu, and T. C. Clancy, "Game theoretical mechanism design methods: suppressing cheating in cognitive radio networks," *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 74–84, Nov. 2008.

[18] 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, Apr. 2009.

[19] J. E. Suris, L. A. DaSilva, Z. Han, A. B. MacKenzie, and R. S. Komali, "Asymptotic optimality for distributed spectrum sharing using bargaining solutions," *IEEE Trans. Wireless Commun*, vol. 8, no. 10, pp. 5225–5237, Oct. 2009.

[20] D. Li, Y. Xu, X. Wang, and M. Guizani, "Coalitional game theoretic approach for secondary spectrum access in cooperative cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 3, pp. 844–856, Mar. 2011.

[21] Z. Ji and K. J. R. Liu, "Multi-Stage pricing game for collusion-resistant dynamic spectrum allocation," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 182–191, Jan. 2008.

[22] Y. Chen, Y. Wu, B. Wang, and K. J. R. Liu, "Spectrum auction games for multimedia streaming over cognitive radio networks," *IEEE Trans. Commun.*, vol. 58, no. 8, pp. 2381–2390, Aug. 2010.

[23] S. K. Jayaweera, M. Bkassiny, and K. A. Avery, "Asymmetric cooperative communications based spectrum leasing via auctions in cognitive radio networks," *IEEE Trans. Wireless Commun.*, accepted, 2011.

[24] S. Vassilaras, D. Vogiatzis, and G. S. Yovanof, "Security and cooperation in clustered mobile ad hoc networks with centralized supervision," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 2, pp. 329–342, Feb. 2006.

[25] S. Buchegger and J.-Y. L. Boudec, "Self-policing mobile ad hoc networks by reputation systems," *IEEE Commun. Mag.*, pp. 101–107, July 2005.

[26] A. Satsiou and L. Tassiulas, "Reputation-based resource allocation in P2P systems of rational users," *IEEE Trans. Par. Dist. Sys.*, vol. 21, no. 4, pp. 466–479, Apr. 2010.

[27] Y. Zhang and Y. Fang, "A fine-grained reputation system for reliable service selection in peer-to-peer networks," *IEEE Trans. Par. and Dist. Sys.*, vol. 18, no. 8, pp. 1134–1145, Aug. 2007.

$$
\begin{aligned}
\boldsymbol{\delta}(a_{i,j})\mathbf{W}\mathbf{q}^T &= \sum_{k \in S_+} \sum_{m \in S_-^k} |\delta_m|(\mathbf{e}_k - \mathbf{e}_m)\mathbf{W}\mathbf{q}^T = \frac{P_v g \lambda N_t}{\bar{\eta}(1-\gamma\lambda)}\Big(\sum_{k \in S_+} k \sum_{m \in S_-^k} |\delta_m| - \sum_{k \in S_+} \sum_{m \in S_-^k} |\delta_m| m\Big), \\
&> \frac{P_v g \lambda N_t}{\bar{\eta}(1-\gamma\lambda)}\Big(a_{i,j} \sum_{k \in S_+} \sum_{m \in S_-^k} |\delta_m| - N \sum_{k \in S_+} \sum_{m \in S_-^k} |\delta_m|\Big) = \frac{P_v g \lambda N_t}{\bar{\eta}(1-\gamma\lambda)}(a_{i,j}-N) \sum_{k \in S_+} \delta_k, \\
&> \frac{P_v g \lambda N_t}{\bar{\eta}(1-\gamma\lambda)}(a_{i,j}-N)(1 - G_{a_{i,j},a_{i,j}} + \sum_{j=a_{i,j}}^{N-1} G_{N,j}),
\end{aligned} \tag{62}
$$

[28] H. Ohtsuki, Y. Iwasa, and M. A. Nowak, "Indirect reciprocity provides only a narrow margin for efficiency for costly punishment," *Nature*, vol. 457, pp. 79-82, 2009.

[29] M. A. Nowak and K. Sigmund, "Evolution of indirect reciprocity," *Nature*, vol. 437, pp. 1291–1298, 2005.

[30] Y. Chen and K. J. R. Liu, "Indirect reciprocity game modelling for cooperation stimulation in cognitive networks," *IEEE Trans. Commun.*, vol. 59, no. 1, pp. 159–168, Jan. 2011.

[31] K. J. R. Liu, A. K. Sadek, W. Su, and A. Kwasinski, *Cooperative Communications and Networking*. Cambridge University Press, 2008.

[32] A. S. Ibrahim, A. K. Sadek, W. Su, and K. J. R. Liu, "Cooperative communications with relay selection: when to cooperate and whom to cooperate with?" *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2814–2827, July 2008.

[33] O. Perron, "Zur Theorie der Matrices," Mathematische Annalen 64, no. 2, pp. 248–263.

[34] G. Frobenius, "Ueber Matrizen aus nicht negativen Elementen," Sitzungsber. Konigl. Preuss. Akad. Wiss.: 456-477.

[35] D. P. Bertsekas, *Dynamic Programming: Deterministic and Stochastic Model*. Prentic-Hall, 1987.

[36] J. M. Smith, *Evolutionary and the Theory of Games*. Cambridege University Press, 1982.

[37] Y. Chen, B. Wang, W. S. Lin, Y. Wu, and K. J. R. Liu, "Cooperative peer-to-peer streaming: an evolutionary game-theoretic approach," *IEEE Trans. Cir. and Sys. for Video Technol.*, vol. 20, no. 10, pp. 1346–1357, Oct. 2010.

[38] D. Blackwell, "Discounted dynamic programming," *Annals of Mathematical Statistics*, vol. 36, pp. 226–235, 1965.

[39] J. Laneman, D. Tse, and G. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, 2004.

[40] R. Fisher, *The Genetical Theory of Natural Selection*. Clarendon Press, 1930.

[41] Y. Tan, S. Sengupta, and K. P. Subbalakshmi, "Analysis of coordinated denial-of-service attacks in IEEE 802.22 networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 4, pp. 890–902, Apr. 2011.

[42] S. Sengupta and M. Chatterjee, "An economic framework for dynamic spectrum access and service pricing," *IEEE/ACM Trans. Netw.*, vol. 17, no. 4, pp. 1200–1213, Aug. 2009.

[43] Y. Tan, S. Sengupta, and K. P. Subbalakshmi, "Competitive spectrum trading in dynamic spectrum access markets: a price war," in *Proc. 2010 GLOBECOM*.

Biling Zhang (M'12) received the B.S. and M.S. degrees from Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2000 and 2003, respectively. Now she is an Assistant Professor as well as a Ph.D. candidate at BUPT. She received a scholarship from China Scholarship Council (CSC) in 2010. From February 2011 to March 2012, she was a visiting scholar in the Department of Electrical and Computer Engineering at the University of Maryland, College Park, MD, USA. Her current research interests are in the areas of cognitive radio networks, wireless communication and cooperative communications.
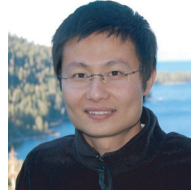
Yan Chen (S'06-M'11) received the Bachelor's degree from University of Science and Technology of China in 2004, the M. Phil degree from Hong Kong University of Science and Technology (HKUST) in 2007, and the Ph.D. degree from University of Maryland College Park in 2011. He is currently a research associate in the Department of Electrical and Computer Engineering at University of Maryland College Park. His current research interests are in social learning and networking, smart grid, cloud computing, crowdsourcing, network economics, multimedia signal processing and communication. Dr. Chen received the University of Maryland Future Faculty Fellowship in 2010, Chinese Government Award for outstanding students abroad in 2011, University of Maryland ECE Distinguished Dissertation Fellowship Honorable Mention in 2011, and was the Finalist of A. James Clark School of Engineering Dean's Doctoral Research Award in 2011.

K. J. Ray Liu (F'03) was named a Distinguished Scholar-Teacher of University of Maryland, College Park, in 2007, where he is Christine Kim Eminent Professor of Information Technology. He leads the Maryland Signals and Information Group conducting research encompassing broad areas of signal processing and communications with recent focus on cooperative communications, cognitive networking, social learning and networks, and information forensics and security. Dr. Liu is the recipient of numerous honors and awards including IEEE Signal Processing Society Technical Achievement Award and Distinguished Lecturer. He also received various teaching and research recognitions from University of Maryland including university-level Invention of the Year Award; and Poole and Kent Senior Faculty Teaching Award and Outstanding Faculty Research Award, both from A. James Clark School of Engineering. An ISI Highly Cited Author, Dr. Liu is a Fellow of IEEE and AAAS. Dr. Liu is President of IEEE Signal Processing Society where he has served as Vice President - Publications and Board of Governor. He was the Editor-in-Chief of *IEEE Signal Processing Magazine* and the founding Editor-in-Chief of *EURASIP Journal on Advances in Signal Processing*.