A Spectrum Access Game in Bidirectional Cognitive Relay Networks

Wei Zhong\textsuperscript{1,2}, Meixia Tao\textsuperscript{1}, Youyun Xu\textsuperscript{2,1}
\textsuperscript{1}Dept. of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China
\textsuperscript{2}Institute of Communications Engineering, PLAUST, Nanjing, China
Email: zhongwei@sjtu.edu.cn, mxtao@sjtu.edu.cn, yyxu@vip.sina.com

Abstract—We consider a cognitive relay network where two secondary users can access the unutilized spectrum and conduct bidirectional communication via the help of primary users. The primary users are regarded as relay nodes of the secondary users and adopt amplify-and-forward relaying strategy. Physical-layer network coding is used when both secondary users select the same relay. The spectrum access problem in the considered system model is defined as joint relay selection and discrete power control. We formulate the problem as a noncooperative game where the players are the secondary users and the common payoff function takes both transmission rate and power consumption into account. The proposed game falls into the framework of potential games where at least one pure strategy Nash equilibrium (NE) exists. Furthermore, the social optimal solution is also a pure strategy NE of the proposed game. We then propose a distributed algorithm based on learning automata to achieve the pure strategy NE. Numerical results show that the proposed algorithm has good convergence and near optimal performance.

I. INTRODUCTION

Recently, cognitive relay networks (CRLN), where primary users can cooperate with secondary users and help to relay the signals of the latter when they are idle, has attracted lot of research interests [1] [2]. In this work, we consider a bidirectional CRLN where two secondary users want to exchange information via the help of primary users. Suppose that there exist multiple primary users which can help the secondary users at the same time. Then, in bidirectional CRLN, an important problem for the secondary users is: how to select the best relay channel to exchange information? On the other hand, the secondary users usually have limited battery power, thus energy conservation is important for the life time of the secondary users. Then another important problem for the secondary users is: how to select the best transmission power level to maximize the energy efficiency? The two questions will be addressed jointly in this work through a game-theoretic based spectrum access.

Previous work on relay selection and power control has mainly focused on one-way communication (e.g. [3]- [6]). The similar problem (especially when the emerging physical-layer network coding (PNC) [7]- [9] or analog network coding (ANC) [10] is used at the relay) for two-way (bidirectional) communication in cognitive networks is however much less investigated. An initial attempt can be found in [2] where the authors studied the beamforming and power control for multi-antenna cognitive two-way relaying where there is only one relay node.

Since none of the secondary and primary users can have the knowledges of all spectrum utilization and channel conditions in advance, it is difficult to obtain the optimal spectrum access strategies (i.e. joint relay selection and power control strategies) in a centralized manner. Distributed spectrum access algorithm is thus of great practical interest. Game theory [11] is a useful tool to analyze the distributed optimization problem and has been successfully applied in power control [12] [13], relay selection [4], joint relay selection and power control in cooperative communication networks [14], distributed mult-mode precoding strategy selection [15] and network coding [16] [17] [18]. However, to the best of our knowledge, none of the previous works have studied the game theoretic analysis of the spectrum access problem for the bidirectional CRLN.

In this paper, we formulate the distributed spectrum access problem as a noncooperative game. We establish the game model and then design a specific utility function of both transmission rate and energy consumption. The Nash equilibrium (NE) of the proposed game is analyzed. Finally, a distributed learning algorithm based on learning automata is proposed to obtain the NE. Simulation results are also given to evaluate the performance.

The remainder of this paper is organized as follows. In Section II, the system model is introduced. In Section III, we study the noncooperative spectrum access game. In Section IV, a distributed learning algorithm is proposed. Simulation results are demonstrated in Section V. The conclusions are given in Section VI.

II. SYSTEM MODEL

Consider a wireless bidirectional CRLN depicted in Fig.1, where two secondary users, T1 and T2, attempt to exchange information, however, there is no direct communication link between them. It is assumed that T1 and T2 sense $n$ idle channels owned by primary users $1, 2, \cdots, n$ at the same moment. These primary users can help to relay the signals from T1 and T2. Thus the primary users can be looked as relays (RL) in this paper. Suppose that these idle channels are orthogonal and will not interfere with each other. Furthermore, the bandwidth of these channels are assumed to be $B_1, \cdots, B_n$, respectively. In Fig.1 (b), $h_{11}, \cdots, h_{1n}$, $h_{21}, \cdots, h_{2n}$ denote the corresponding complex-valued channel coefficients.

Abstract—We consider a cognitive relay network where two secondary users can access the unutilized spectrum and conduct bidirectional communication via the help of primary users. The primary users are regarded as relay nodes of the secondary users and adopt amplify-and-forward relaying strategy. Physical-layer network coding is used when both secondary users select the same relay. The spectrum access problem in the considered system model is defined as joint relay selection and discrete power control. We formulate the problem as a noncooperative game where the players are the secondary users and the common payoff function takes both transmission rate and power consumption into account. The proposed game falls into the framework of potential games where at least one pure strategy Nash equilibrium (NE) exists. Furthermore, the social optimal solution is also a pure strategy NE of the proposed game. We then propose a distributed algorithm based on learning automata to achieve the pure strategy NE. Numerical results show that the proposed algorithm has good convergence and near optimal performance.

I. INTRODUCTION

Recently, cognitive relay networks (CRLN), where primary users can cooperate with secondary users and help to relay the signals of the latter when they are idle, has attracted lot of research interests [1] [2]. In this work, we consider a bidirectional CRLN where two secondary users want to exchange information via the help of primary users. Suppose that there exist multiple primary users which can help the secondary users at the same time. Then, in bidirectional CRLN, an important problem for the secondary users is: how to select the best relay channel to exchange information? On the other hand, the secondary users usually have limited battery power, thus energy conservation is important for the life time of the secondary users. Then another important problem for the secondary users is: how to select the best transmission power level to maximize the energy efficiency? The two questions will be addressed jointly in this work through a game-theoretic based spectrum access.

Previous work on relay selection and power control has mainly focused on one-way communication (e.g. [3]- [6]). The similar problem (especially when the emerging physical-layer network coding (PNC) [7]- [9] or analog network coding (ANC) [10] is used at the relay) for two-way (bidirectional) communication in cognitive networks is however much less investigated. An initial attempt can be found in [2] where the authors studied the beamforming and power control for multi-antenna cognitive two-way relaying where there is only one relay node.

Since none of the secondary and primary users can have the knowledges of all spectrum utilization and channel conditions in advance, it is difficult to obtain the optimal spectrum access strategies (i.e. joint relay selection and power control strategies) in a centralized manner. Distributed spectrum access algorithm is thus of great practical interest. Game theory [11] is a useful tool to analyze the distributed optimization problem and has been successfully applied in power control [12] [13], relay selection [4], joint relay selection and power control in cooperative communication networks [14], distributed mult-mode precoding strategy selection [15] and network coding [16] [17] [18]. However, to the best of our knowledge, none of the previous works have studied the game theoretic analysis of the spectrum access problem for the bidirectional CRLN.

In this paper, we formulate the distributed spectrum access problem as a noncooperative game. We establish the game model and then design a specific utility function of both transmission rate and energy consumption. The Nash equilibrium (NE) of the proposed game is analyzed. Finally, a distributed learning algorithm based on learning automata is proposed to obtain the NE. Simulation results are also given to evaluate the performance.

The remainder of this paper is organized as follows. In Section II, the system model is introduced. In Section III, we study the noncooperative spectrum access game. In Section IV, a distributed learning algorithm is proposed. Simulation results are demonstrated in Section V. The conclusions are given in Section VI.

II. SYSTEM MODEL

Consider a wireless bidirectional CRLN depicted in Fig.1, where two secondary users, T1 and T2, attempt to exchange information, however, there is no direct communication link between them. It is assumed that T1 and T2 sense $n$ idle channels owned by primary users $1, 2, \cdots, n$ at the same moment. These primary users can help to relay the signals from T1 and T2. Thus the primary users can be looked as relays (RL) in this paper. Suppose that these idle channels are orthogonal and will not interfere with each other. Furthermore, the bandwidth of these channels are assumed to be $B_1, \cdots, B_n$, respectively. In Fig.1 (b), $h_{11}, \cdots, h_{1n}$, $h_{21}, \cdots, h_{2n}$ denote the corresponding complex-valued channel coefficients.
Assume that all the nodes operate only in half-duplex mode, i.e., cannot transmit and receive at the same time. The traffic pattern for relaying is divided into two phases (or time slots). The channels are assumed to be constant over both phases. It is further assumed that the primary users do not attempt to decode the information of secondary users and thus the amplify-and-forward (AF) relaying is adopted. If T1 and T2 select the same RL, then the selected RL will use physical-layer network coding to relay the signals from T1 and T2 [7]-[9]. Otherwise, when T1 and T2 select different RLs, the selected RLs will relay the signals individually on different channels. For simplicity, we assume that T1 and T2 select the same RL, e.g. RL 1, then, at phase 1, the signal received at RL 1 is given by

\[
y_{R_1} = \sqrt{P_1}h_{1\eta}x_1 + \sqrt{P_2}h_{2\eta}x_2 + z_{R_1}
\]

where \(x_1\) and \(x_2\) denote the signals to be exchanged from T1 to T2, and from T2 to T1, respectively, \(P_1\) and \(P_2\) denote the transmit power of T1 and T2, respectively, and \(z_{R_1}\) denotes the additive Gaussian noise which follows the distribution \(\mathcal{CN}(0, \sigma^2)\). In this paper, we assume that the transmission powers of T1 and T2 are subject to peak power constraints \(P_1\) and \(P_2\) respectively, i.e., \(P_1 \leq P_1, P_2 \leq P_2\).

At RL \(\eta\), the received signal \(y_{R_\eta}\) is amplified to a signal \(\beta_\eta y_{R_\eta}\), where \(\beta_\eta\) is the amplification coefficient. The power of \(\beta_\eta y_{R_\eta}\) needs to satisfy the power budget \(P_{R_\eta}\) at the RL, i.e.,

\[
\beta_\eta^2P_1|h_{1\eta}|^2 + \beta_\eta^2P_2|h_{2\eta}|^2 + \beta_\eta^2\sigma^2 \leq P_{R_\eta}
\]

At phase 2, RL \(\eta\) broadcasts the signal \(\beta_\eta y_{R_\eta}\) to both T1 and T2. The signals received at T1 and T2 can be expressed as

\[
\hat{y}_1 = h_{1\eta}\beta_\eta y_{R_\eta} + z_{1\eta}
\]

\[
= \beta_\eta h_{1\eta}^2\sqrt{P_1}x_1 + \beta_\eta h_{1\eta}h_{2\eta}\sqrt{P_2}x_2 + \beta_\eta h_{1\eta}z_{R_\eta} + z_{1\eta}
\]

\[
\hat{y}_2 = h_{2\eta}\beta_\eta y_{R_\eta} + z_{2\eta}
\]

\[
= \beta_\eta h_{2\eta}^2\sqrt{P_2}x_2 + \beta_\eta h_{1\eta}h_{2\eta}\sqrt{P_1}x_1 + \beta_\eta h_{2\eta}z_{R_\eta} + z_{2\eta}
\]

where \(z_{1\eta}\) and \(z_{2\eta}\) are additive Gaussian noise at T1 and T2. The principle of PNC is then applied at T1 and T2 for decoding. Assume that the channel state information (CSI) is perfectly known at both T1 and T2, the self interference components can be completely canceled out, yielding

\[
y_1 = \beta_\eta h_{1\eta}h_{2\eta}\sqrt{P_2}x_2 + \beta_\eta h_{1\eta}z_{R_\eta} + z_{1\eta}
\]

\[
y_2 = \beta_\eta h_{1\eta}h_{2\eta}\sqrt{P_1}x_1 + \beta_\eta h_{2\eta}z_{R_\eta} + z_{2\eta}
\]
If T1 and T2 select different relays (e.g., T1 selects RL $l$ and T2 $d$), the signals received at the selected RLs can be given by
\[
y_{RL} = \sqrt{P_l h_{Rl} x_1} + z_{RL}
\]
where $z_{RL}$ denotes the received Gaussian noise and follow the same distribution $CN(0, \sigma^2)$.

At phase 2, the RL $l$ and RL $d$ transmit the amplified received signals $\beta_l y_{RL}$ and $\beta_d y_{RD}$ to T1 and T2 respectively on their own channels without interfering to each other. Note that, the power of $\beta_l y_{RL}$ and $\beta_d y_{RD}$ need to satisfy the power budget $P_{Rl}$ and $P_{Rd}$ respectively, i.e.,
\[
\beta_l^2 P_1 h_{Rl}^2 + \beta_l^2 \sigma^2 \leq P_{Rl}
\]
and
\[
\beta_d^2 P_2 h_{Rd}^2 + \beta_d^2 \sigma^2 \leq P_{Rd}
\]
The signals received at T1 and T2 can then be expressed as
\[
y_1 = \beta_d h_{Rd} \left( \sqrt{P_2 h_{2d} x_2} + z_{1d} \right) + z_{1d}
\]
and
\[
y_2 = \beta_l h_{Rl} \left( \sqrt{P_1 h_{1l} x_1} + z_{2l} \right) + z_{2l}
\]
where $\beta_l$ and $\beta_d$ denote the amplification coefficients at RL $l$ and RL $d$, respectively, $z_{1d}$ and $z_{2l}$ denote the received Gaussian noise and follow the same distribution $CN(0, \sigma^2)$.

In this paper, we assume that all additive noises are independent and that the transmission powers of the RLs are the same and fixed, i.e., $P_{Rl} = P_{Rd}, \forall i \in V = \{1, \cdots, n\}$. Furthermore, the transmission powers of T1 and T2 are assumed to be chosen from the finite sets $\mathcal{P}_1$ and $\mathcal{P}_2$ respectively, where $\mathcal{P}_1 = \{ P_{11} < P_{12} < \cdots < P_{1\Omega_1} = P_1 \}$, $\mathcal{P}_2 = \{ P_{21} < P_{22} < \cdots < P_{2\Omega_2} = P_2 \}$, $\Omega_1$ is the cardinality of $\mathcal{P}_1$, and $\Omega_2$ is the cardinality of $\mathcal{P}_2$. The discrete power control is due to that the power level in practical wireless communication systems is usually quantized into discrete values [13].

III. NONCOORDERATIVE SPECTRUM ACCESS GAME

In this section, we first establish a general framework of the noncooperative spectrum access game. After that, we design a specific utility function for the considered bidirectional CRLN and analyze the NE of the proposed game.

A. Game Theoretic Formulation

Since all nodes cannot know the global information in advance, it is difficult for the secondary users to cooperate with each other. In order to optimize a certain performance metric, the decision made by each secondary user should depend on the decision made by the other one. This leads to a noncooperative game between the T1 and T2 as follows.

Noncooperative spectrum access game (NSAG):
- Players $\iff$ Secondary user T1 and T2;
- Pure strategies $\iff$ The relay selection and the transmit power levels of the secondary users;
- Payoff $\iff$ The degree of satisfaction. Formally, this game can be denoted as
\[
(G) \quad \mathcal{G} = [N, \{\mathcal{A}_i\}_{i \in N}, \{u_i\}_{i \in N}]
\]
where $N = \{1, 2\}$ is the set of secondary users, $\mathcal{A}_i = V \times \mathcal{P}_i$ is the set of the pure strategies (actions) and $u_i$ is the utility function of secondary user $i$.

Let $a_i$ denote an action from the action set $\mathcal{A}_i$ for secondary user $i$ and let $K_i$ be the cardinality of $\mathcal{A}_i$. The pure strategy profile is denoted as $(a_1, a_2)$. To single out the pure strategy of secondary user $i$, let $a_{-i}$ denote the set of actions of secondary user $j \neq i$. In this paper, $a_i$ is defined as the integer variable $a_i = (v_i, P_i)$, where $v_i \in V$ denotes the index of the relay selected by secondary user $i$ and $P_i$ denotes the power level of secondary user $i$. That is, the pure strategy selection is essentially the joint relay selection and discrete power control.

The utility is a measure of the relative satisfaction of the decision maker and is not unique. Any function that puts the elements of $\{A_i\}_{i \in N}$ in the desired order is a candidate for a utility function [12]. Normally, different utility functions may result in different analysis and conclusions. How to design the utility function depends on the preference of the terminals in the system.

For CRLN, the satisfaction is naturally related to the total information bits that the secondary users can transfer. On the other hand, since secondary users are usually energy limited, the power consumption is also a major concern. However, there is a fundamental tradeoff between the rate efficiency and power efficiency, since decreasing the transmit power usually degrades the transmission rate. In this work, we introduce a novel utility of both total information rates and power consumption in the following.

Let $c_{\text{sum}}$ denote the weighted sum rate of the system, $c_1$ and $c_2$ denote the rate of secondary user T1 and T2 respectively. Then we define the utility function $u_i$ of secondary user $i$, which is common for both users, as
\[
u_i = \frac{c_{\text{sum}}}{f(P_1, P_2)} = \frac{w_1 c_1 + w_2 c_2}{f(P_1, P_2)}
\]
where $w_1 > 0$ and $w_2 > 0$ are weights of T1 and T2 respectively and $f(P_1, P_2)$ is a monotonic function of $P_1$ and $P_2$. Such utility function characterizes the energy-aware information rate of the system. Thus, we can see that the NSAG is a common payoff game with the utility function defined in (14). That is, $u_1 = u_2 = U$.

Assume that T1 selects RL $v_1$ and T2 selects RL $v_2$. If $v_1 = v_2 = v$, then the transmission rates of T1 and T2 are given by, following the system equations (5) and (6), respectively,
\[
c_1 = B_v \log \left( 1 + \frac{\beta_v^2 P_1 |h_{1v}|^2 |h_{2v}|^2}{\beta_v^2 |h_{2v}|^2 \sigma^2 + \sigma^2} \right)
\]
and
\[
c_2 = B_v \log \left( 1 + \frac{\beta_v^2 P_2 |h_{1v}|^2 |h_{2v}|^2}{\beta_v^2 |h_{1v}|^2 \sigma^2 + \sigma^2} \right)
\]
where
\[
\beta_v^2 = \frac{P_R}{P_1 |h_{1v}|^2 + P_2 |h_{2v}|^2 + \sigma^2}
\]
If \( v_1 \neq v_2 \), then we have
\[
\begin{align*}
    c_1 &= B_{v_1} \log \left( 1 + \frac{\beta_{v_1}^2 P_1 |h_{v_1}|^2 |h_{v_2}|^2}{\beta_{v_1} |h_{v_1}|^2 \sigma^2 + \sigma^2} \right) \quad (18) \\
    c_2 &= B_{v_2} \log \left( 1 + \frac{\beta_{v_2}^2 P_2 |h_{v_1}|^2 |h_{v_2}|^2}{\beta_{v_2} |h_{v_2}|^2 \sigma^2 + \sigma^2} \right) \quad (19)
\end{align*}
\]
where
\[
\begin{align*}
    \beta_{v_1} &= \frac{P_R}{P_1 |h_{v_1}|^2 + \sigma^2} \quad (20) \\
    \beta_{v_2} &= \frac{P_R}{P_2 |h_{v_2}|^2 + \sigma^2} \quad (21)
\end{align*}
\]
Substituting the rate expressions (15) and (16) or (18) and (19) into (14), we can obtain the common payoff functions of the spectrum access game under the two different relay selection strategies.

B. Nash Equilibrium

Nash equilibrium (NE) is one of the most important solution to game-theoretic problem [11]. In this subsection, we analyze the NE of the proposed game.

**Theorem 1:** With the utility function defined in (14), NSAG is a potential game.

**Proof:** Assume that \( a_i \in A_i \) is an arbitrary strategy of secondary user \( i \), and \( a'_i \in A_i \) is an alternate strategy of secondary user \( i \) and the strategy of the other secondary user remains unchanged.

As we stated before, the NSAG is a game with common payoff. Then we can have
\[
\begin{align*}
    \forall a_i, a'_i, \quad u_i(a_i, a_{-i}) - u_i(a'_i, a_{-i}) &= U(a_i, a_{-i}) - U(a'_i, a_{-i}) \quad (22)
\end{align*}
\]
In (22), by considering the function \( U \) as a potential function \( \Phi \), we find that NSAG satisfies potential game definition [19]:
\[
\begin{align*}
    \forall a_i, a'_i, \quad u_i(a_i, a_{-i}) - u_i(a'_i, a_{-i}) &= \Phi(a_i, a_{-i}) - \Phi(a'_i, a_{-i}) \quad (23)
\end{align*}
\]

Define an optimization problem as
\[
\begin{align*}
    (P1) \quad \max_{\{a_i \in A_i\} \in \mathcal{N}} U = \frac{\ell_{\text{sum}}}{f(P_1, P_2)} \quad (24)
\end{align*}
\]
and assume that \( (a^n, a^0) \) is the optimal solution to P1. Then we can have the following theorem.

**Theorem 2:** With the utility function defined in (14), the optimal solution to P1 is a pure strategy NE of NSAG and NSAG has at least one pure strategy NE.

**Proof:** This theorem can be easily proved based on Theorem 1 and [15, Lemma 2.1]

IV. DISTRIBUTED LEARNING ALGORITHM

Since all the terminals in the system have no knowledge of others in advance, the best response dynamic algorithm which is usually used to obtain the NE of a potential game is impractical due to large amount of overheads. In this section, we propose a distributed stochastic learning algorithm to obtain the NE using the concept of learning automata. The stochastic learning technique based on learning automata has been successfully used for discrete power control [13], multimode precoding strategy selection [15], and so on. The stochastic learning is a distributed algorithm involving only limited amount of feedback and is also computationally simple and efficient.

To characterize the learning algorithm, we extend the NSAG to a mixed strategy form. Let \( p_i = \{p_{i1}, \cdots, p_{iK_i}\} \) be the mixed strategy of the secondary user \( i \), where \( p_{ik} \) denotes the probability with which the \( i \)th secondary user chooses the \( k \)th pure strategy. The expected payoff \( g^i \) for secondary user \( i \) is given by
\[
g^i(p_1, p_2) = E[u_i \mid j \text{th secondary user employs strategy}]
\]
\[
p_j, 1 \leq j \leq 2 = \sum_{j_1, j_2} d^j(j_1, j_2) \prod s \pi_{j_2} \quad (25)
\]
where \( d^j(j_1, j_2) = E[u_i \mid s \text{th user employs strategy}]
\]
\[
j_s, 1 \leq s \leq 2.
\]
If the NSAG is played successively, it could be modeled as a stochastic game of learning automata. Each secondary user (i.e. player) is represented by a learning automaton and the actions of the automaton are the pure strategies of the secondary user. The mixed strategy \( p_i(t) = \{p_{i1}(t), \cdots, p_{iK_i}(t)\} \) is the action probability distribution of the \( i \)th automaton at instant \( t \). \( p_{ik}(t) \) denotes the probability with which the \( i \)th automaton chooses the \( k \)th pure strategy at instant \( t \). The normalized payoff to the \( i \)th user will be the reaction \( r_i(t) \) to the \( i \)th automaton. We let \( r_i(t) = \alpha u_i(t) \) be the normalized \( u_i(t) \), where \( 0 < \alpha < 1 \) is a normalized parameter to guarantee the value of \( r_i(t) \) in the interval \([0, 1]\). Then
\[
r_1(t) = \cdots = r_n(t) = r(t) = \alpha U(t).
\]
In this paper, we use an adaptive normalize parameter mechanism, since the secondary users usually have no prior knowledge of their utilities. That is, if at instant \( t \), \( U(t) > \frac{1}{\alpha} \), we let \( \alpha = \frac{1}{U(t) + \tau} \), where \( \tau \) is a positive scalar; otherwise, \( \alpha \) remains unchanged.

When one of the automata chooses an action independently according to its current action probabilities, the game is said to be played for once. The game is played repeatedly to learn the NE. Let \( p_i(0) \) denote the initial mixed strategy of secondary user \( i \). The learning algorithm used by each of the secondary users is given as below.

**Distributed Spectrum Access Learning Algorithm (DSALA):**

1. Set the initial probability vector \( p_i(0) \) as: \( p_{ik}(0) = \frac{1}{K_i}, \ i = 1, 2; \ k = 1, \cdots, K_i \).
2. At every time instant \( t \), each secondary user chooses an action \( a_i^t \) according to its action probability vector \( p_i(t) \).
3 Each secondary user computes a reaction \( r_i(t) \) based on the feedback from the relays. Note that each selected relay evaluates the transmission rate of the corresponding secondary user and broadcasts it to all secondary users. Furthermore, the secondary users send the power levels to the relays, and then the relays broadcast the power levels to both secondary users. The secondary users then compute the reaction based on the feedback information from all relays.

4 Each secondary user updates its action probability through the rule

\[
\begin{align*}
p_{ik}(t + 1) &= p_{ik}(t) - b r_i(t) p_{ik}(t), & i_k \neq a_i, \\
p_{ik}(t + 1) &= p_{ik}(t) + b r_i(t)(1 - p_{ik}(t)), & i_k = a_i,
\end{align*}
\]

where \( i = 1, 2; \ k = 1, \cdots, K_i \).

5 If \( \forall i \in \mathcal{N} \), there exists a component of \( p_i(t) \) which is larger than a value approaching one, say 0.99, stop. Otherwise, go to step 2.

The DSALA only needs a small amount of signalling overhead and does not require the terminals to know any prior information of the system. In addition, through the same theoretic analysis process as in [15], we can know that DSALA always converges to a pure strategy NE of NSAG with the utility function defined in (14).

Note that the NE of NSAG is only a local optimal solution. If we want to obtain a near-optimal solution of P1, we can repeat the DSALA several times and then choose the pure strategy NE with the highest utility.

V. NUMERICAL RESULTS

In this section, we present some numerical results to illustrate the performance of the proposed game. Here we assume that the channel coefficients follow the distribution of circularly symmetric complex Gaussian (CSCG) with zero mean and unity variance, and define \( n = 2 \), \( f(P_1, P_2) = P_1 + P_2 \), \( B_1 = B_2 = 1 \). The weights are given by \( w_1 = 2 \), \( w_2 = 1 \). The transmission power set is \( P_1 = P_2 = \{0.05, 0.1, 1\} \) Watts with \( \sigma^2 = 5 \times 10^{-15} \) Watts and \( P_R = 1 \) Watts. Then, we can have \( K_1 = K_2 = 6 \). That is, each secondary user has six pure strategies. The parameters for the learning algorithm are set as \( b = 0.2 \), \( \tau = 10 \), with the initial value of \( \alpha \) being 0.01.

In Fig.3 (a), without loss of generality, we plot the evolution of the choice probabilities of the actions (i.e. mixed strategy) \( p_{11}, p_{12}, p_{21}, p_{22} \) of T1 and T2. In Fig.3 (b), we plot the trace of the utility function \( U \). It is seen that DSALA converges after about 35 iterations. Both T1 and T2 converges to the first pure strategy \( (p_{11} = 1, p_{21} = 1) \).

In Table I, we compare the average utility \( U \) (i.e. the total information rate per unit energy consumption) of different algorithms for 500 channel realizations. Here, DSALA-Repeat means that the DSALA is repeated for 6 times. It is found that our proposed DSALA algorithm significantly outperforms the random selection algorithm. Furthermore, DSALA-Repeat can obtain the same performance as the brute-force search algorithm and slightly outperform the DSALA.

VI. CONCLUSIONS

In this paper, we considered a cognitive relay network where the primary users act as relay nodes to assist bidirectional communication between two secondary users. A noncooperative spectrum access game was established for relay selection and discrete power control of the secondary users. Our designed common utility function represents both the total information rate and energy efficiency of the system and takes the physical layer network coding into account. With this utility function, we proved that the proposed game is a potential game and a distributed stochastic learning algorithm can be used to achieve the pure strategy NE of the game. Numerical results showed that the proposed algorithm obtains near optimal performance and significantly outperforms the random selection algorithm. The proposed algorithm only needs a small amount of overheads and can always converge in finite steps. Moreover, it was found that repeating the proposed algorithm can further improve the energy efficiency and obtain almost optimal performance.

Several issues can be further investigated in the future work. For example, the price of anarchy of the DSALA, i.e., the ratio of the utility obtained by the worst case NE to highest utility, can be discussed. The network with multiple (more than two) secondary users can also be considered.
ACKNOWLEDGMENT

This work was supported by the Inter-Government Technology Collaboration Research Project of Shanghai under grant 10220712000, by the National High Technology Research and Development Program of China (863 Program) under Grant number 2009AA01Z249, by the Major National Basic Research Development Program of China (973 Program) under Grant number 2009CB3020402, by the Important National Science & Technology Specific Projects under Grant number 2010ZX03002-007, and by Projects of International Cooperation and Exchanges of the Ministry of Science and Technology of China under Grant number S2010GR0445 and EP/G042713/1.

REFERENCES