Learning to Be Energy-Efficient in Cooperative Networks

Daxin Tian, *Senior Member, IEEE,* Jianshan Zhou, Zhengguo Sheng, and Qiang Ni, *Senior Member, IEEE*

***Abstract*—Cooperative communication has great potential to improve the transmit diversity in multiple users environments. To achieve a high network-wide energy-efficient performance, this letter poses the relay selection problem of cooperative communication as a noncooperative automata game considering nodes' selfishness, proving that it is an ordinal game (OPG), and presents a game-theoretic analysis to address the benefit-equilibrium decision-making issue in relay selection. A stochastic learning-based relay selection algorithm is proposed for transmitters to learn a Nash-equilibrium strategy in a distributed manner. We prove through theoretical and numerical analysis that the proposed algorithm is guaranteed to converge to a Nash equilibrium state, where the resulting cooperative network is energy-efficient and reliable. The strength of the proposed algorithm is also confirmed through comparative simulations in terms of energy benefit and fairness performances.**

***Index Terms*—Cooperative networks, energy-efficiency, self-organized relay selection, decentralized learning.**

I. Introduction

IN general, decision-making algorithms for relay selection play an essential role in cooperative communication [1]. Many recent works focus on the best relay selection mechanism (BRS), in which the "best" relay is determined according to the minimum (or maximum) instantaneous value of a metric such as transmission power, average error probability, outage probability, cooperative diversity, etc. [1]–[3]. In addition, some variations of BRS such as [4] and buffer-aided relay selection policies [5] have also been developed.

The conventional cooperative decision-making mechanism to achieve its good performance gain is based on the fundamental assumption that individuals are socially responsible. However, in reality, users are more likely to behave in a selfish and greedy manner. This fact is further substantiated by absence of centralized control. Little work considers how the individual selfishness and greediness impacts the overall cooperative communication. It is still left to answer whether there potentially exists a desired energy-efficient network state and how to achieve such desired state with a good tradeoff between energy cost and system performance, if users behave selfishly in networking interactions. We believe that providing decentralized benefit-compatible and self-adaptive relay selection methods can open an appealing way to promote cooperative networks.

In this letter, we model the joint behaviors of selfish users in the decision-making process of relay selection as a noncooperative automata game. For this relay selection game, we propose a novel utility function specifying that nodes have enough incentive i) to establish and maintain reliable cooperative communication with low transmission power level and ii) to improve the energy consumption balance. This work provides a novel game-theoretic mapping from the relay selection of self-interest-driven nodes regarding selfish interactions in cooperative communication to a proper decentralized learning-based decision-making formulation, helping to better understand the desired strategic behaviors of selfish nodes and to induce an energy-efficient Nash-equilibrium cooperative network.

II. Problem Formulation and System Model

We consider a wireless ad-hoc network where multiple nodes co-exist, denoted by a set $\mathcal{N} = \{1, 2, \ldots, N\}$. Some nodes are the sources $s \in \mathcal{S} \subset \mathcal{N}$ who would like to transmit packets to specified destinations $d \in \mathcal{D} \subset \mathcal{N}$ using the two-time-slot repetition-coded DF cooperative communication [1]. Let the maximum transmission power of each node be $p_{\max}$. We denote by $\mathcal{A}_s \subset N$ the set of s's neighboring nodes. Each source s can select a relay a_s from $\mathcal{A}_s$, i.e., $a_s \in \mathcal{A}_s$. We operate the cooperative transmissions in discrete successive time periods $\{[t\Delta t, (t+1)\Delta t) | t = 0, 1, 2, \ldots\}$ where Δt is a specific time period. Additionally, every time period Δt can be further divided into a series of two successive time slots, in each of which different transmissions are active.

We employ a channel model incorporating effects of Rayleigh fading, shadow fading and path loss [1], [3], [4]. Given a specific data rate R in $bit/s/Hz$ specified according to the QoS requirement and a specific outage probability threshold β_s that should not be exceeded in order to guarantee the transmission reliability in the cooperative communication, we can derive the minimum power consumption of the direct ($P^{\min}_{sd_s,D}$) and the cooperative ($P^{\min}_{sd_s,C}$) transmissions by

$$P^{\min}_{sd_s,D} = SNR_{sd_s,D} N_0 BR \tag{1}$$

$$P^{\min}_{sd_s,C} = 2SNR_{sd_s,C} N_0 BR \tag{2}$$

where N_0 denotes the variance of the zero-mean mutually independent, circularly symmetric, complex Gaussian noises, and B denotes the bandwidth assigned to achieve per-unit spectrum effectiveness. $SNR_{sd_s,D}$ and $SNR_{sd_s,C}$ represent the

This research was supported in part by the National Natural Science Foundation of China under Grant no. 61672082.

D. Tian and J. Zhou are with Beijing Key Laboratory for Cooperative Vehicle Infrastructure Systems & Safety Control, School of Transportation Science and Engineering, Beihang University, Beijing 100191, China (e-mail: dtian@buaa.edu.cn,jianshanzhou@foxmail.com).

Z. Sheng is with Department of Engineering and Design, the University of Sussex, Richmond 3A09, UK (e-mail: z.sheng@sussex.ac.uk).

Q. Ni is with School of Computing and Communications, Lancaster University, Lancaster LA1 4WA, UK (e-mail:q.ni@lancaster.ac.uk).

corresponding signal-noise-ratios, the formulas of which can be referred to in [1]. In addition, based on (1) and (2), the transmission power can also be normalized by $P^{\min}_{sd_s,D}/N_0RB$ and $P^{\min}_{sd_s,C}/N_0RB$, respectively. The detailed expressions of the normalized minimum transmission power can be found in [4]. We remark that in our experiments we use the normalized forms rather than the original to avoid introducing additional parameter settings.

III. Decentralized Learning Framework for Adaptive Relay Selection

A. Learning Automata Game Mapping

We formulate the relay selection as a dynamic noncooperative game where $\mathcal{S}$ is the set of the players. The action set of any player $s \in \mathcal{S}$ is defined as the neighbor set $\mathcal{A}_s$, where each relay candidate is treated as an action the node s can select. We further define $a_s(t) \in \mathcal{A}_s$, representing the relay of s in the t-th time period. An action profile can then be $\boldsymbol{a}_s(t) = (a_s(t), \boldsymbol{a}_{-s}(t))$ where $\boldsymbol{a}_{-s}(t)$ denotes the actions taken by the others $s' \in \mathcal{S}\backslash\{s\}$. Let $\mathcal{A} = \times_{s\in\mathcal{S}}\mathcal{A}_s$ be the space of all action vectors. For any $s \in \mathcal{S}$, a utility function $u_s(\boldsymbol{a}_s(t)):\ \mathcal{A} \to \mathbb{R}$ can model its reward or payoff depending on the action profile $\boldsymbol{a}_s(t)$. A vector collecting all the utilities is $\boldsymbol{u} = \left(u_{s_1}, \ldots, u_{s_{|\mathcal{S}|}}\right):\ \mathcal{A} \to \mathbb{R}^{|\mathcal{S}|}$. In addition, we define the energy residual of any node $i \in \mathcal{N}$ at the beginning of the t-th time period by $E_i(t)$. A collection of the residual energy of all the nodes, $\boldsymbol{E}(t) = \left(E_1(t), \ldots, E_{|\mathcal{N}|}(t)\right)$, is the external state. We present the game by the tuple

$$\Omega(t) = \langle \mathcal{S}, \boldsymbol{E}(t), \mathcal{A}, \boldsymbol{u} \rangle \tag{3}$$

Due to the selfishness and greediness, each node tends to maximize its payoff in the relay selection game $\Omega(t)$. Such a payoff is considered to consist of the benefit a node receives from the resulting cooperative network and the cost it incurs in cooperative communication. Thus, a utility function of each player captures the benefit-cost trade-off and maps its action profile to a payoff. For $s \in \mathcal{S}$, we can model $u_s\left(\boldsymbol{a}_s(t)\right)$ as

$$u_s\left(\boldsymbol{a}_s(t)\right) = f_s\left(\boldsymbol{a}_s(t)\right) - g_s\left(\boldsymbol{a}_s(t)\right) \tag{4}$$

where $f_s:\ \mathcal{A} \to \mathbb{R}$ denotes the benefit s can gain when the action profile $\boldsymbol{a}_s(t)$ is deployed, and $g_s:\ \mathcal{A}_s \to \mathbb{R}$ represents the cost incurred. To be specific, in the context of cooperative communication, each player can receive a benefit in establishing a high-reliable communication and balancing the energy utilization. For any $i \in \mathcal{S}$, the reliability in i's cooperative transmission is quantified by the outage probability, denoted by $Pr\left(P_{id_i,C}(t)\right)$, which is the possibility that the maximum average mutual information ($Inf_{sd_i,C}$) between i and d_i at the power level $P_{id_i,C}(t)$ is less than the required spectral efficiency R, i.e., $Pr\left(P_{id_i,C}(t)\right) = \text{Prob}\left\{Inf_{sd_i,C} < R\right\}$. It should be noted that $Inf_{sd_i,C}$ is a non-decreasing function of the power level $P_{id_i,C}(t)$, and the outage probability of the DF cooperative communication can be found in [1]–[4]. A QoS-oriented reliability usually imposes a constraint, $Pr\left(P_{id_i,C}(t)\right) \le \beta_i$. To capture the transmission reliability of the overall network, an indicator function is given by

$$I_s(\boldsymbol{a}_s(t)) = \begin{cases} 0, & \text{if } \exists i \in \mathcal{S}, \beta_i < Pr(P_{id_i,C}(t)) \\ 1, & \text{otherwise} \end{cases} \tag{5}$$

Accordingly, if and only if $Pr(P_{id_i,C}(t)) \le \beta_i$ for all $i \in \mathcal{S}$, $I_s(\boldsymbol{a}_s(t)) = 1$, indicating that a QoS-oriented reliable cooperative network is established.

To quantify the energy utilization, the dynamics of the average energy residual of each s's neighbors is given by

$$\overline{W_s(\boldsymbol{a}_s(t))} = \frac{1}{|\mathcal{A}_s|+1}\left(\frac{E_s(t)}{E_s(0)} + \sum_{i\in\mathcal{A}_s}\frac{E_i(t)}{E_i(0)}\right) \tag{6}$$

Due to the fact that a transmitter can benefit from its usable maximum power level $p_{\max}$ (a higher $p_{\max}$ implies a larger transmission capability), a specific $f_s\left(\boldsymbol{a}_s(t)\right)$ is proposed based on (5) and (6):

$$f_s\left(\boldsymbol{a}_s(t)\right) = I_s(\boldsymbol{a}_s(t))\left(2\omega_1 p_{\max}\frac{E_s(0)}{E_s(t)} + \omega_2\overline{W_s(a_s(t))}\right) \tag{7}$$

which signifies the comprehensive benefit of a player in terms of transmission reliability, energy utilization and power capability. ω_1 and ω_2 are two positive weights. It is noted that since a cooperative transmission involves a source and a relay, the total maximum usable power is $2p_{\max}$.

On the other side, the cost incurred in cooperative communication can be naturally represented by the energy consumption and residual energy. Thus, the cost component is given by

$$g_s\left(\boldsymbol{a}_s(t)\right) = \omega_1 P_{sd_s,C}(t)\frac{E_s(0)}{E_s(t)} \tag{8}$$

In a normal form, we express the relay selection game by

$$\Omega\left(t\right):\ \max_{a_s(t)\in\mathcal{A}_s}\left\{u_s\left(a_s(t), \boldsymbol{a}_{-s}(t)\right)\right\} \tag{9}$$

The game is played repeatedly, so that the cooperative network can evolve dynamically via iterative process.

B. Theoretic Properties of Proposed Model

To facilitate the theoretical analysis, we modify the action set of any s, $\mathcal{A}_s = \left\{a_s^k \middle| k = 1, \ldots, |\mathcal{A}_s|\right\}$, as an ordered set where the elements a_s^k are arranged in order of their corresponding power consumption levels $P^k_{sd_s,C}$ (k can be treated as the order number). That is, in this ordered set, $a_s^{k-1} > a_s^k$ can correspondingly indicate $P^{k-1}_{sd_s,C} \ge P^k_{sd_s,C}$. Thus, I_s is monotonically non-decreasing with respect to $a_s \in \mathcal{A}_s$. We first prove that $\Omega(t)$ is a ordinal potential game.

Theorem 1: Given (4), $\Omega(t)$ is an ordinal potential game, a potential function of which can be constructed by

$$\begin{aligned}\mathcal{U}\left(a_s(t), \boldsymbol{a}_{-s}(t)\right) = &\sum_{s\in\mathcal{S}} f_s\left(a_s(t), \boldsymbol{a}_{-s}(t)\right) \\ &- \sum_{s\in\mathcal{S}} g_s\left(a_s(t), \boldsymbol{a}_{-s}(t)\right)\end{aligned} \tag{10}$$

Proof: For any player s, the change in the utility function of s when it switches from the action $a_s(t)$ to $a'_s(t)$ is

$$\begin{aligned}\Delta u_s &= u_s\left(a_s(t),\boldsymbol{a}_{-s}(t)\right)-u_s\left(a'_s(t),\boldsymbol{a}_{-s}(t)\right)\\ &=2\omega_1 p_{\max}\frac{E_s(0)}{E_s(t)}\begin{pmatrix}I_s\left(a_s(t),\boldsymbol{a}_{-s}(t)\right)\\ -\,I_s\left(a'_s(t),\boldsymbol{a}_{-s}(t)\right)\end{pmatrix}\\ &\quad-\omega_1\frac{E_s(0)}{E_s(t)}\left(P_{sd_s,C}(t)-P'_{sd_s,C}(t)\right)\\ &\quad+\omega_2\begin{pmatrix}I_s\left(a_s(t),\boldsymbol{a}_{-s}(t)\right)\overline{W_s\left(a_s(t),\boldsymbol{a}_{-s}(t)\right)}\\ -\,I_s\left(a'_s(t),\boldsymbol{a}_{-s}(t)\right)\overline{W_s\left(a'_s(t),\boldsymbol{a}_{-s}(t)\right)}\end{pmatrix}\end{aligned}\tag{11}$$

Without loss of generality, we can assume $a_s(t) > a'_s(t)$, i.e., $P_{sd_s,C}(t) \geq P'_{sd_s,C}(t)$, so that it sees

$$\Delta u_s\begin{cases}\geq 0, & \text{if } I_s\left(a_s(t),\boldsymbol{a}_{-s}(t)\right)>I_s\left(a'_s(t),\boldsymbol{a}_{-s}(t)\right)\\ \geq 0 \text{ or } <0, & \text{if } I_s\left(a_s(t),\boldsymbol{a}_{-s}(t)\right)=I_s\left(a'_s(t),\boldsymbol{a}_{-s}(t)\right)\end{cases}\tag{12}$$

In a similar way, we can also get

$$\begin{aligned}\Delta\mathcal{U} &=\mathcal{U}\left(a_s(t),\boldsymbol{a}_{-s}(t)\right)-\mathcal{U}\left(a'_s(t),\boldsymbol{a}_{-s}(t)\right)\\ &=\Delta u_s\\ &\quad+\sum_{i\in\mathcal{S},i\neq s}\left\{\begin{array}{c}\left(I_i\left(a_s(t),\boldsymbol{a}_{-s}(t)\right)-I_i\left(a'_s(t),\boldsymbol{a}_{-s}(t)\right)\right)\\ \times\left(2\omega_1 p_{\max}\dfrac{E_i(0)}{E_i(t)}+\omega_2\overline{W_i(a_i(t))}\right)\end{array}\right\}\end{aligned}\tag{13}$$

For the first case presented in (12), the sign of $\left(I_i\left(a_s(t),\boldsymbol{a}_{-s}(t)\right)-I_i\left(a'_s(t),\boldsymbol{a}_{-s}(t)\right)\right)$ in the second term of (13) is the same as that of Δu_s. For the second case in (12), since the transmission reliability of the overall cooperative transmission links is left unchanged, $\left(I_i\left(a_s(t),\boldsymbol{a}_{-s}(t)\right)-I_i\left(a'_s(t),\boldsymbol{a}_{-s}(t)\right)\right)=0$ is held for any $i\in\mathcal{S}$. This indicates $\Delta\mathcal{U}=\Delta u_s$. It sees that the sign of $\Delta\mathcal{U}$ is always the same as that of Δu_s. This result can also be proven when $a_s(t) < a'_s(t)$. To sum up, $\Omega(t)$ is an OPG where $\mathcal{U}$ is its ordinal potential function [6]. ■

Based on **Theorem** 1, the existence of a Nash Equilibrium in $\Omega(t)$ can be always guaranteed and it coincides with a maximizer of the ordinal potential function $\mathcal{U}$ [6].

C. Stochastic Learning based Relay Selection Adaptation

To learn the Nash-equilibrium optimal strategies, we design a decentralized learning-based algorithm. Let the optimal power level associated with $a_s^k\in\mathcal{A}_s$ be $P_{sd_s,C}^{k,\min}$. We present a strategy of s as $\boldsymbol{x}_{sd_s}(t)=\left(x_{sd_s,1}(t),\ldots,x_{sd_s,|\mathcal{A}_s|}(t)\right)$, i.e., the selection probability distribution over $\mathcal{A}_s$. Then, we can derive the update of $\boldsymbol{x}_{sd_s}(t)$ based on the linear reward-inaction approach [7]:

$$\begin{aligned}&x_{sd_s,k}(t+1)=\\ &\begin{cases}x_{sd_s,k}(t)+\delta\widetilde{r}_s\left(a_s(t)\right)\left(1-x_{sd_s,k}(t)\right), & \text{if } a_s^k=a_s(t)\\ x_{sd_s,k}(t)-\delta\widetilde{r}_s\left(a_s(t)\right)x_{sd_s,k}(t), & \text{otherwise}\end{cases}\end{aligned}\tag{14}$$

where $\delta\in(0,1)$ is the learning rate and $\widetilde{r}_s\left(a_s(t)\right)$ is the instantaneous reward that the player s perceives when it currently takes the action $a_s(t)$, which is normalized in the interval $(0,1)$ and determined based on its utility function (4):

$$\widetilde{r}_s\left(a_s(t)\right)=\frac{u_s\left(a_s(t),\boldsymbol{a}_{-s}(t)\right)-u_s^{\text{lower}}(t)}{u_s^{\text{upper}}(t)-u_s^{\text{lower}}(t)}\tag{15}$$

where $u_s^{\text{lower}}(t)=\min_{0\leq\tau\leq t}\left\{u_s\left(a_s(\tau),\boldsymbol{a}_{-s}(\tau)\right)\right\}$ and $u_s^{\text{upper}}(t)=\max_{0\leq\tau\leq t}\left\{u_s\left(a_s(\tau),\boldsymbol{a}_{-s}(\tau)\right)\right\}$. The proposed decentralized learning-based adaptive relay selection (DLbARS) algorithm is described in **Algorithm** 1.

Algorithm 1 DLbARS

```
1: Initialization: Let $t = 0$; for any $s$, set $a_s(t) = \mathrm{argmin}_{a_s^k\in\mathcal{A}_s}\left\{p_{sd_s,C}^{k,\min}\right\}$; $s$ performs transmission assisted by $a_s(t)$; set $x_{sd_s,k}(t) = 1/|\mathcal{A}_s|$ for $k = 1,\ldots,|\mathcal{A}_s|$;
2: Adaptation: For $t \geq 1$, each $s$ updates its $\boldsymbol{x}_{sd_s}(t)$ by (14), and selects a new $a_s(t)$ by stochastic experiment with this new $\boldsymbol{x}_{sd_s}(t)$; determine the optimal power level $P_{sd_s,C}^{\min}$ by (2) to perform the cooperative transmission;
3: Update: The sources and relays update their energy residuals $E_s(t)$, $E_{a_s(t)}(t)$; each source derives a new instantaneous normalized reward $\widetilde{r}_s(t)$ by (15);
4: Set $t = t + 1$ and repeat Adaptation and Update in turn.
```

Next, we show the convergence of the DLbARS as follows:

***Theorem** 2:* The LbDARS guarantees the transmission reliability of each cooperation link and converges to a Nash Equilibrium of the game $\Omega(t)$ that is locally energy efficient when the learning rate δ is sufficiently small.

Proof: In the LbDARS, once a source $s\in\mathcal{S}$ chooses a certain relay a_s^k (an action), it can determine an optimal power level corresponding to a_s^k by (2), $P_{sd_s,C}^{k,\min}$, such that $Pr\left(P_{sd_s,C}^{k,\min}\right)=\beta_s$ is satisfied. Thus, the transmission reliability constraint is always held at every iteration stage.

Additionally, since the learning rate δ is assumed sufficiently small, it follows the analysis of [7] that the update formation (14) can reduce to an ordinary differential equation (ODE) system (noting $\sum_{l=1}^{|\mathcal{A}_s|}x_{sd_s,l}(t)=1$)

$$\frac{dx_{sd_s,k}(t)}{dt}=x_{sd_s,k}(t)\sum_{l=1}^{|\mathcal{A}_s|}x_{sd_s,l}(t)\begin{pmatrix}\widetilde{r}_s\left(a_s^k\right)\\ -\,\widetilde{r}_s\left(a_s^l\right)\end{pmatrix}\tag{16}$$

We define the expected ordinal potential function with respect to the mixed strategy profile of s as $\overline{\mathcal{U}(t)}=\mathbb{E}_{\boldsymbol{x}_{sd_s}(t)}\left[\mathcal{U}(a_s(t),\boldsymbol{a}_{-s}(t))\right]=\sum_{l=1}^{|\mathcal{A}_s|}x_{sd_s,l}(t)\mathcal{U}(a_s^l,\boldsymbol{a}_{-s}(t))$, and then get $\partial\overline{\mathcal{U}(t)}/\partial x_{sd_s,k}(t)=\mathcal{U}(a_s^k,\boldsymbol{a}_{-s}(t))$. Combining this result and (16) further leads to

$$\frac{d\overline{\mathcal{U}(t)}}{dt}=\frac{1}{2}\sum_{s\in\mathcal{S}}\sum_{k=1}^{|\mathcal{A}_s|}\sum_{l=1}^{|\mathcal{A}_s|}x_{sd_s,k}(t)x_{sd_s,l}(t)\Delta\mathcal{U}\Delta\widetilde{r}_s\geq 0\tag{17}$$

where $\Delta\widetilde{r}_s=\widetilde{r}_s(a_s^k)-\widetilde{r}_s(a_s^l)$ and $\Delta\mathcal{U}=\mathcal{U}(a_s^k,\boldsymbol{a}_{-s}(t))-\mathcal{U}(a_s^l,\boldsymbol{a}_{-s}(t))$. Since $\mathrm{sgn}\left(\Delta U\right)=\mathrm{sgn}\left(\Delta u_s\right)=\mathrm{sgn}\left(\Delta\widetilde{r}_s\right)$ as shown in **Theorem** 1. It follows (17) that $\overline{\mathcal{U}(t)}$ is nondecreasing in the phase space of the ODE system. Because $\overline{\mathcal{U}(t)}$ is bounded, it is expected to converge to a Nash equilibrium. ■

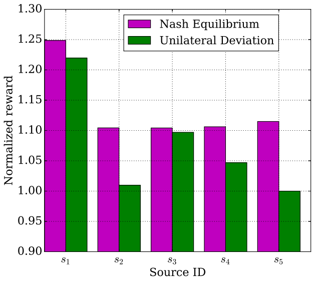

Fig. 1. Unilateral deviation from each player's converging strategy profile.

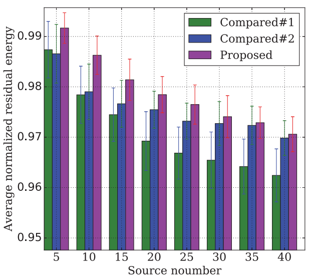

Fig. 2. The average normalized residual energy of different schemes.

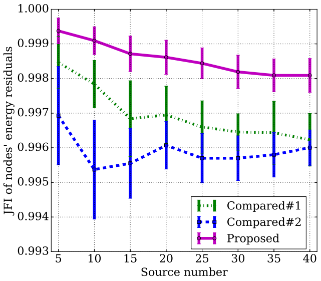

Fig. 3. JFI of nodes' energy residuals of different schemes.

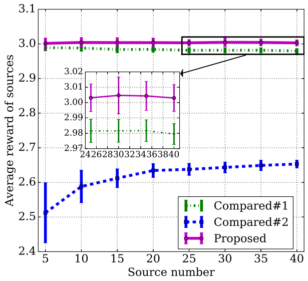

Fig. 4. The sources' average reward of different schemes.

IV. Numerical Results

We carry out a series of experiments to show the performance of the proposed method. We adopt the settings in Table I throughout the experiments. Additionally, $p_{\max}$ is set to be the direct transmission power level when the distance reaches $dist_{\max}$. Each source transmits $packetNum$ packets during $packetNum$ iterations. The initial energy of each node is identically set to be $(packetNum + 10)p_{\max}\Delta t$ (for the sake of example, let $\Delta t = 1\text{s}$).

TABLE I
The basic parameter settings.

The maximum transmission distance $dist_{\max}$	150m
The number of packets transmitted $packetNum$	1000
The path loss coefficient α	3.0
The minimum data rate R	1.0
The outage probability constraint β	0.01
The learning rate δ	0.1
The weights ω_1 and ω_2	1.0

Firstly, we uniformly and randomly generate 50 nodes distributed in a region of 300m×300m, and also randomly generate 5 transmission pairs. We analyze the unilateral deviation of the strategies learned by these players. In Fig. 1, it can be found that lower benefit is obtained by unilateral deviation experiments, which indicates that each player cannot gain additional benefit by unilaterally changing their strategies, suggesting that a Nash equilibrium state is reached by the proposed algorithm.

Next, we compare the performance of the proposed learning-based algorithm ('Proposed') with two other schemes, one of which is the minimum transmission power based scheme [1] ('Compared#1'), and the other is based on the maintenance of an adaptive relay candidate set [4] ('Compared#2'). We set $|\mathcal{S}| \in \{5, 10, 15, \ldots, 40\}$ and the total node number is $2|\mathcal{S}| + 50$. For comparison, we evaluate the average normalized residual energy of the network, the well-known Jain's fairness index (JFI) of energy residuals as well as the average reward of the sources against different $|\mathcal{S}|$. Monte Carlo simulations are conducted for performance comparison. All the Monte Carlo simulations have been performed with 100 replications per simulation point, and the results are shown with the standard deviations (See Fig. 2, Fig. 3 and Fig. 4). We can find that the Nash-equilibrium energy-efficient cooperative network achieved by our proposed method has comprehensive advantages over the other two schemes, since it can benefit the cooperative network more in terms of energy benefit and fairness performance metrics.

V. Conclusion

In this letter, we have investigated the issue of energy-efficient cooperative transmission in wireless ad-hoc network. This problem is formulated as a multi-player game automata model, and a decentralized learning-based relay selection algorithm has been proposed to achieve a self-organized cooperative network. The convergence of the proposed algorithm to a pure strategy Nash equilibrium is confirmed through simulations. Monte Carlo experiment results reveal the comprehensive strength of this algorithm in terms of average reward and energy consumption balance.

References

[1] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, 2004.

[2] A. Scaglione, D. L. Goeckel, and J. N. Laneman, "Cooperative communications in mobile ad hoc networks," *IEEE Signal Processing Magazine*, vol. 23, no. 5, pp. 18–29, 2006.

[3] W. Su and X. Liu, "On optimum selection relaying protocols in cooperative wireless networks," *IEEE Transactions on Communications*, vol. 58, no. 1, pp. 52–57, 2010.

[4] Z. Sheng, J. Fan, C. H. Liu, V. C. M. Leung, X. Liu, and K. K. Leung, "Energy-efficient relay selection for cooperative relaying in wireless multimedia networks," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 3, pp. 1156–1170, 2015.

[5] N. Nomikos, T. Charalambous, I. Krikidis, D. N. Skoutas, D. Vouyioukas, M. Johansson, and C. Skianis, "A survey on buffer-aided relay selection," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1073–1097, 2016.

[6] D. Monderer and L. S. Shapley, "Potential games," *Games and Economic Behavior*, vol. 14, no. 1, pp. 124–143, 1996.

[7] P. S. Sastry, V. V. Phansalkar, and M. A. L. Thathachar, "Decentralized learning of Nash equilibria in multi-person stochastic games with incomplete information," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 24, no. 5, pp. 769–777, 1994.