A Game-theoretic Approach for Relay Assignment over Distributed Wireless Networks

Xuedong Liang*, Min Chen†, Victor C.M. Leung*
*Dept. of Electrical and Computer Engineering, The University of British Columbia, Vancouver, Canada V6T 1Z4
Email: {xuedongl, vleung}@ece.ubc.ca
†School of Computer Science and Engineering, Seoul National University, Korea, 151-742, Email: minchen@ieee.org

***Abstract*—To achieve full cooperative diversity gains while still maintaining spectral and energy efficiency, relay assignment schemes for cooperative communications have been extensively studied in recent research. These schemes select only the best relay from multiple relaying candidates to cooperate with a communication link. In this paper, we formulate the problem of relay assignment as a non-cooperative, mixed strategy, repeated game, where relaying candidates are modeled as rational players. We then propose a Game Theory based Relay Assignment scheme *GTRA*, in which each player plays against all the other players, and determines whether to cooperate with a communication link on a packet-by-packet basis in a distributed manner. To adapt to dynamic environments, an adaptive learning algorithm is utilized by players to learn optimal strategies of relay assignment, as well as orienting the game to converge to a set of correlated equilibriums. We compare *GTRA* with *BR*, a fictitious game based approach. The simulation results show that *GTRA* outperforms *BR* in terms of network throughput, especially in environments where the channel fading becomes severe. It is also shown that *GTRA* can converge to a correlated equilibrium in a short period that enables it to work well in dynamic environments.**

***Keywords*—cooperative communications; relay assignment; game theory.**

I. Introduction

Relay assignment for cooperative communications [1], i.e., dynamically choosing the best relay from multiple relaying candidates to cooperate with a communication link, has emerged as an effective technique to improve the performance of wireless networks, by exploiting the spatial diversity of the wireless medium. Optimal relay assignment can enable a cooperative communication system to achieve full diversity gains while still obtaining high spectral and energy efficiency. However, finding the optimal relay in distributed wireless networks is challenging, as the link qualities vary over time.

To develop adaptive and distributed relay assignment schemes for wireless networks, game theory based approaches have received much research attention [2]. Various games, e.g., fictitious game [3], Stackelberg game [4], have been used to model nodes' interactions in the process of relay assignment. The game theory based schemes in the literature often assume that players have complete information of the game, e.g., all players' identities, strategies, payoffs, and/or utility functions. Furthermore, the game's history, e.g., actions taken in previous stages and the actions' corresponding outcomes, is also assumed to be known to all players [3]. However, these assumptions do not always hold in realistic scenarios, as nodes in wireless networks usually only have locally observed information and limited knowledge of others nodes' behavior. Therefore, adaptive learning, e.g., estimating payoffs that may obtain by taking certain actions, predicting the other players' possible behavior, should be involved in game designs.

In this paper, we propose a Game Theory based Relay Assignment scheme *GTRA*, in which the process of relay assignment is modeled as a non-cooperative, mixed-strategy, repeated game and the relaying candidates are modeled as players. Packet transmissions between a source and a destination are modeled as repeated stages in the multi-stage game, where each stage represents the transmission of one packet. At every stage of the proposed game, each player plays against all the other players, and determines whether to cooperate with the communication link to assist the packet transmission between the source and destination in a distributed manner. To adapt to dynamic environments, a modified-regret-matching (MRM) algorithm [5] is utilized by players to learn optimal strategies and to orient the game to converge to a set of correlated equilibriums (CEs) [6]. We compare *GTRA* with another game theory based approach *BR* [3], which models the process of relay assignment as a fictitious two player game. Simulation results show that *GTRA* outperforms *BR* in terms of network throughput, especially in environments where the channel fading becomes more severe. It is also shown that *GTRA* can converge to a set of CE(s) in a short period, which enables it to work well in dynamic environments.

The rest of the paper is organized as follows. We present the system model in Section II. Section III formulates the problem of relay assignment as a mixed-strategy game, and shows how to orient the game to converge to a set of CE(s) by using adaptive learning algorithms. The performance analysis is presented in Section IV. Finally, Section V concludes the paper and discusses the future research directions.

II. System Model

As shown in Fig. 1, we consider a wireless network consisting of uniformly and randomly distributed nodes, which are functionally equivalent in terms of radio communications, signal processing, and power supply. Multiple pairs of source and destination nodes are randomly selected for data packet transmissions. For a pair of source s and destination d, we assume that there exits a set of common-neighboring nodes

978-1-4577-9538-2/11/$26.00 ©2011 IEEE

N which are connected with both of s and d, as nodes are usually densely deployed. Therefore, a node, e.g., node $n_i \in N$, may overhear the packet transmission between s and d due to the broadcast nature of the wireless medium. Node n_i may cooperate with the communication link between s and d by retransmitting the packet overheard from s.

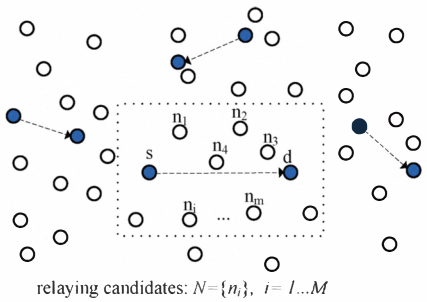

Fig. 1. Cooperative communications with relay assignment

We consider the use of decode-and-forward (DaF) [7] as the cooperative protocol, which operates in two phases, namely, direct transmission and relay transmission. In the direct transmission phase, the source transmits a packet to the destination and all the relaying candidates. In the relay transmission phase, a relay is chosen from the relaying candidates to retransmit the packet that overheard in the direct transmission phase to the destination. Then, the destination combines the signals received from both the source and the relay, and applies maximal ratio combining (MRC) [8] for optimal packet decoding. To reduce network operations and power consumption, the relay transmission will be invoked only if a packet transmission fails in the direct transmission. We assume that the destination can feedback the improvement on received signals' signal-to-noise-ratio (SNR) to the assigned relay by using acknowledgment (ACK) or negative-acknowledgment (NACK) packets with extended data fields.

III. Analysis of the Proposed Game

A. Description of The Proposed Game

The process of relay assignment is modeled as a non-cooperative, mixed strategy, repeated game G, which is denoted as $G = (N, A, U)$, where

- $N = \{n_1, \cdots, n_m\}$ denotes the set of players;
- $A = \{A_1, \cdots, A_m\}$ denotes the set of actions;
- $U = \{u_1, \cdots, u_m\}$ denotes the set of utility functions.

In the proposed game, the nodes are modeled as rational players, which means that the nodes are expected to follow a common set of strategies and choose actions from the strategies to maximize their utilities. In a distributed system, the nodes behaves selfishly, i.e., a node always chooses actions to maximize its own utility, without considering the utilities of other nodes. For a player, e.g., n_i, the strategy A_i consists of two actions, i.e., $A_i = \{cc, ncc\}$, where cc denotes that the player n_i decides to cooperates with the communication link between s and d by retransmitting the overheard packet, and ncc denotes that n_i chooses to remain silent. As a mixed-strategy game, each player takes its actions in accordance with a probability distribution. That is, the player n_i chooses the actions of cc and ncc with the probabilities of $p_i(cc)$ and $p_i(ncc)$, respectively, such that $p_i(cc) + p_i(ncc) = 1$.

Since the number of packets transmitted between the source and destination is often assumed to be large, we model the process of relay assignment as a multi-stage game, and the process of relay assignment iterates in each stage of the game. Each candidate simultaneously plays the game against the other candidates in a distributed manner. Whether a candidate cooperates with a communication link or not depends on the payoff it may obtain. As the environment is assumed to be dynamic, each player, e.g., n_i, updates its strategy by adjusting the probability distribution over the actions of cc and ncc, according to the link qualities between $s \rightarrow n_i$ (source-relay) and $n_i \rightarrow d$ (relay-destination).

Each packet transmission is modeled as a stage in the multi-stage game and consists of the following steps:

1) The source transmits a packet to the destination and all the players.
2) Each player chooses an action of cc or ncc autonomously and simultaneously, based on the probability distribution of $p_i(cc)$ and $p_i(ncc)$, respectively.
3) Each player evaluates the quality of the selected action by computing the obtained payoff.
4) Each relay updates its strategy by adjusting the probability distribution of $p_i(cc)$ and $p_i(ncc)$.

B. Payoff Calculation

In this game theory based approach, payoff is used to describe the difference between the benefit, i.e., improvement on link quality achieved by relay retransmission, and the cost associated, i.e., the channel occupancy time of a relay, and the energy consumed by the relay. The utility function designed to compute the payoff is defined as

$$u_i = \omega_1 \frac{SNR_{s,n_i,d} - SNR_{s,d}}{SNR_{s,d}} - (\omega_2 \frac{T_{n_i,t_x} - T_{n_i,r_x}}{T_{avr}} + \omega_3 \frac{P_{n_i}}{P_m}), \quad (1)$$

where $SNR_{s,d}$ denotes the SNR of the received signal at the destination d that is transmitted by the source s in the direct transmission phase, and $SNR_{s,n_i,d}$ denotes the SNR of the signal that combines the signals received from the source s and the selected relay n_i, respectively. T_{n_i,t_x} and T_{n_i,r_x} represent the packet retransmitting time and the packet receiving time at the relay n_i, respectively. The value difference between T_{n_i,t_x} and T_{n_i,r_x} reflects the packet processing, queuing, and channel access contention delays at node n_i. T_{avr} is the average amount of time needed for preparing a packet retransmission without considering processing, queuing, and channel access contention delays. T_{avr} is calculated as

$$T_{avr} = T_{TA} + T_{BO}, \quad (2)$$

where T_{TA} denotes the transceiver's receiving to transmitting turnover time which is a constant value for a specific radio hardware, and T_{BO} is the average backoff time at n_i without any channel contention, and the value is determined by the underlying medium access control (MAC) layer protocol.

P_{n_i} is the transmission power level of player n_i. P_m is the medium power level between P_{min} and P_{max}, where P_{min} and P_{max} are the minimum and maximum available transmission power levels of player n_i.

ω_1, ω_2, and ω_3 are the weighting factors for the metrics of SNR, delay, and energy consumption, respectively. The values of the weighting factors can be adjusted to adapt to the QoS requirements of the communication link.

In (1), the first term represents the improvement (in percentage) on the link quality in terms of SNR by employing the relay transmission. The second term denotes the relative delay for preparing a packet retransmission, including processing, queuing, and channel access contention delays. The third term represents the relative energy efficiency, compared with using a fixed, medium transmission power level. If a player takes the action of cc, the achieved payoff computed by using (1) reflects both the benefit and cost; otherwise if a player chooses the action of ncc, the payoff is zero, as there is neither benefit nor cost caused by the action.

C. Correlated Equilibrium

In a game, if players can receive a signal containing strategy recommendations from a central coordinator, and all the players follow the recommendations on how to play the game, the outcome of the game can converge to a set of CE(s), which is often the most system efficient state of the game. In a CE, it is the best interest of a player to follow the recommendations, i.e., a player will not get a higher payoff by taking any other actions, provided all other players follow the recommendations. In the context of relay assignment, a CE can be interpreted as a *steady* state, as none of the nodes has incentive to unilaterally deviate from the recommendation profile to increase its payoff.

For a game, a probability distribution p is a CE of the game, if and only if $\forall\ n_i \in N$, $a_i \in A_i$, $\forall a_i^{'} \in A_i$, it holds that

$$\sum_{a_{-i} \in A_{-i}} p(a_i, a_{-i}) \left[u_i(a_i^{'}, a_{-i}) - u_i(a_i, a_{-i})\right] \leq 0, \qquad (3)$$

or equivalently,

$$\sum_{a_{-i} \in A_{-i}} p(a_{-i}|a_i) \left[u_i(a_i^{'}, a_{-i}) - u_i(a_i, a_{-i})\right] \leq 0, \qquad (4)$$

where A_i denotes n_i's action space, and A_{-i} denotes the action space of n_i's opponents, i.e., all player except n_i. a_i is the action chosen by n_i from A_i, and a_{-i} is the action combination of n_i's opponents. $u_i(a_i, a_{-i})$ is n_i's obtained payoff by taking the action of a_i, and its opponents taking the action combination of a_{-i}. $p(a_i, a_{-i})$ is the joint probability distribution over actions for all players. Eq. (4) can be interpreted as that when the player n_i is recommended to choose action a_i, then choosing $a_i^{'}$ ($a_i^{'} \in A_i, a_i^{'} \neq a_i$) will not lead to a higher payoff.

D. Convergence to Correlated Equilibrium

Orienting the relay assignment game to converge to a set of CE(s) is not trivial, as a central coordinator broadcasting recommendations on how to play the game is often not available in distributed wireless networks. A feasible approach [9] of orienting a game to a set of CE(s) is using the common history of the game as a coordinator. That is, a game can converge to a set of CE(s) if each player adjusts its strategy by tracking a series of *regret* values, which are quantitative measures for not taking certain actions in previous stages.

1) Calculation of Regret Values: Regret-matching (RM) [9], also called no-regret learning, can be used by players to calculate the regret values.

Assuming player n_i has taken action a_i at each of the past M stages, the difference of average payoff $D_i^M(a_i^{'}, a_i)$ between the player that has actually obtained and the player that would have obtained if it had taken the action $a_i^{'}$ instead of action a_i is defined as

$$D_i^M(a_i^{'}, a_i) = \frac{1}{M} \sum_{m \leq M} (u_i^m(a_i^{'}, a_{-i}) - u_i^m(a_i, a_{-i})), \quad (5)$$

where $u_i^m(a_i^{'}, a_{-i})$ is the payoff the player would have obtained if the player had taken action $a_i^{'}$ at stage m, and $u_i^m(a_i, a_{-i})$ is the payoff the player has actually obtained by taking action a_i at stage m.

For any two actions $a_i^{'}$ and a_i, the regret value for not taking action $a_i^{'}$ at the previous M stages is defined as:

$$R_i^M(a_i^{'}, a_i) = \max\{D_i^M(a_i^{'}, a_{-i}), 0\}. \qquad (6)$$

The regret value is proportional to the difference of the average payoffs, and is lower bounded by zero to ensure that the probabilities of taking any actions are positive. If the regret value is zero, it means that the player has obtained a higher payoff by taking action a_i than taking $a_i^{'}$, and thus there is no regret. Otherwise, if the regret value is greater than zero, it means that the player would have obtained a higher payoff if the player had taken action $a_i^{'}$.

As observed in (5) and (6), the implementation of RM requires that the player should know the payoffs it would have obtained if its actions in previous stages had been different from the actions that the player has actually taken. However, it is difficult for a player to compute the payoff that the corresponding action has not been taken. A common approach [10] in the literature is that using a coordinator broadcasting references to all the players at each stage of the game, indicating the potential payoffs that players would have obtained if the players had taken certain actions in the previous stage. By observing the references, the players can compute the payoff differences and thus calculate the regret values. However, this approach is often not feasible in distributed wireless networks wherein central coordinating is often not available.

Modified-regret-matching (MRM) [5] has been proposed for players to estimate the payoffs that the corresponding actions have not been actually taken, by only using the available historical information. That is, a player only need to know the probability distribution over its actions, and the payoffs it has obtained in the previous stages, to determine the probabilities of actions from the actually realizations only. MRM can be interpreted as a *reinforcement* or *stimulus-response* mechanism, as in the procedure of MRM, a relative high payoff at stage m will tend to increase the probability of playing the same action at stage $m+1$.

In MRM, the difference of the average payoffs of player n_i would have obtained if it had taken the action $a_i^{'}$ every time in the previous stages instead of taking a_i is defined as

$$C_i^M(a_i^{'}, a_i) = \frac{1}{M} \sum_{m \leq M} \frac{p_i^m(a_i^{'})}{p_i^m(a_i)} u_i^m(a_i, a_{-i}) - \frac{1}{M} \sum_{m \leq M} u_i^m(a_i, a_{-i}), \quad (7)$$

where $p_i^m(a_i^{'})$ and $p_i^m(a_i)$ denote the probabilities of taking $a_i^{'}$ and a_i at stage m, respectively. The first term in (7) is an estimation of the average payoff the player would have obtained if player n_i had taken the action of $a_i^{'}$ in the previous M stages; and the second term is the average payoff that player n_i has actually obtained by taking action a_i at every stage in the past M stages.

The modified regret value of not playing $a_i^{'}$ is defined as

$$Q_i^M(a_i^{'}, a_i) = \max(C_i^M(a_i^{'}, a_i), 0) \quad (8)$$

Eq. (7) and (8) show that a player can estimate its regret value of not taking a certain action by only using locally available information, i.e, the probabilities of actions, and the payoffs actually obtained in the previous stages.

2) Strategy update: adjusting probability distribution of actions: To maximize payoffs in a multi-stage game, a player updates its strategy by adjusting the probability distribution over different actions throughout the game. In MRM, a player, e.g., n_i, updates its probabilities of actions at stage $m+1$ as:

$$p_i^{m+1}(a_i^{'}) = (1 - \frac{\delta}{m^\gamma}) \min(\tfrac{1}{\mu} Q_i^m(a_i^{'}, a_i), \frac{1}{K_i - 1}) + \frac{\delta}{m^\gamma} \frac{1}{K_i};$$
$$p_i^{m+1}(a_i) = 1 - \sum_{a_i^{'} \neq a_i} p_i^{m+1}(a_i^{'}), \quad (9)$$

where $a_i^{'}$ and $a_i \in A_i$ and $a_i^{'} \neq a_i$, K_i is the number of actions available for player n_i, μ is a sufficiently large number which controls the probabilities of action switching and convergence speed. To ensure the game converge to a limit CE, the term $\frac{\delta}{m^\gamma}$ decreases to zero as m increases, where $0 < \delta < 1$ and $0 < \gamma < 0.25$. More details on parameter settings can be found in [5].

Eq. (9) can be interpreted as follows. The first term, with weight $(1 - \frac{\delta}{m^\gamma})$, denotes the modified regret value of not taking $a_i^{'}$, and shows how strongly the player intends to switch from a_i to $a_i^{'}$ at the next stage $m+1$. The second term, with weight $\frac{\delta}{m^\gamma}$, denotes the uniform distribution over the available actions of n_i. This uniform distribution guarantees that all possible actions at stage $m+1$ can be taken with the probabilities of $\frac{\delta}{m^\gamma}$ at least. The first term is upper bounded to ensure that the sum of the probabilities does not exceed one.

The mathematic property of (9) shows that if an action can get a relative high payoff at stages m, then the belief of taking the same action at $m+1$ is *reinforced*. For an action, a higher payoff will generate a greater reinforcement. All the effects, e.g., belief reinforcement, action switching, decrease with the evolution of the game, as m increases over time.

E. GTRA *Algorithm based on MRM Learning*

In the relay selection game, for player n_i, the probabilities of taking actions cc and ncc at stage m are denoted as $p_i^m(cc)$ and $p_i^m(ncc)$, respectively. The pseudo code of the GTRA algorithm implemented at n_i is listed in Algorithm 1.

Algorithm 1 The GTRA algorithm based on MRM learning

begin
initialization
Generate an arbitrarily probability distribution of $p_i^0(cc)$ and $p_i^0(ncc)$, and $p_i^0(cc)$ + $p_i^0(ncc)$ =1.
for m=1,2,3, $\cdots$
1. Compute the difference of average payoff $C_i^m(a_i^{'}, a_i)$ using (7).
2. Compute the regret value $Q_i^m(a_i^{'}, a_i)$ using (8).
3. Update the probability distribution $p_i^{m+1}(cc)$ and $p_i^{m+1}(ncc)$ at stage $m+1$ using (9).
if cc is the action chosen at each of the m stages,
then adjust the probabilities of taken actions of ncc and cc as:
$p_i^{m+1}(ncc) = (1 - \frac{\delta}{m^\gamma}) \min(\frac{1}{u} Q_i^m(ncc, cc), 1) + \frac{\delta}{m^\gamma}$;
$p_i^{m+1}(cc) = 1 - p_i^{m+1}(ncc)$.
else adjust the probabilities of taken actions of cc and ncc as:
$p_i^{m+1}(cc) = (1 - \frac{\delta}{m^\gamma}) \min(\frac{1}{u} Q_i^m(cc, ncc), 1) + \frac{\delta}{2m^\gamma}$;
$p_i^{m+1}(ncc) = 1 - p_i^{m+1}(cc)$.
end for

F. Convergence of the GTRA Algorithm

For the number of stages M, the relative frequency of players' action a played until M stages is defined as

$$z_M(a) = \frac{1}{M} \left|\{m \leq M \ : \ a_m = a\}\right|, \quad (10)$$

where a_m denotes all the users' action at stage m. It has been proved in [5] that z_M is guaranteed to converge almost surly at $m \to \infty$ to the set of CE(s) of the game G, if each player plays according to the adaptive MRM learning procedure, and adjusts its probability distribution over actions as defined in (9). As the number of packets transmitted from the source to the destination is often assumed to be large, i.e., the number of stages is sufficiently large, the relay assignment game will converge to a set of CE(s).

IV. Performance Evaluation

We simulate a WSN where 20 sensor nodes are randomly distributed in a 50m × 50m area. A constant bit rate (CBR) traffic with 5 packets per second is used as the communication pattern, and the source and destination nodes are chosen randomly in each simulation run. We define the network

TABLE I
SIMULATION PARAMETERS

Parameters	**Value**
Wireless channel model	log shadowing path loss channel
Path loss exponent	2.4
Channel deviation (in dB)	1
Collision model	Additive interference model
Physical and MAC layer	IEEE 802.15.4 standard
Packet length	50 bytes
Transmitting power level	[-25, -15, -10, -7, -5, -3, -1, 0] dbm
Node's initial energy	12 J
Data transmission rate	250 kbps
Simulation time	100 s
ω_1	0.5
ω_2	0.3
ω_3	0.2

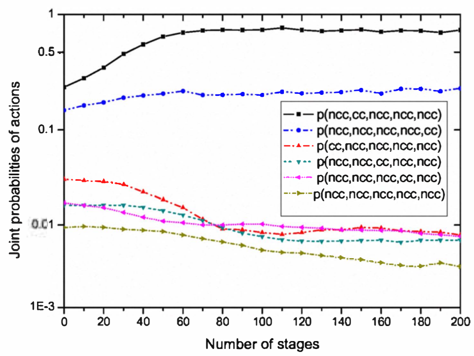

Fig. 2. Evolution of the *GTRA* game

lifetime is the time when the first node exhausts its battery's energy. Table I lists the detailed simulation parameters.

The Castalia [11] wireless sensor network simulator, which is based on the OMNeT++ [12] discrete event simulation platform, is used as the simulation environment. The data link layer in Castalia is modified to facilitate *MRC* combining and decoding, and we also extend the ACK and NACK signals with new fields to feedback the received signals' SNR information.

Fig. 2 illustrates the evolution of the *GTRA* game in which 5 players cooperating with a source-destination link, and δ and γ are set to 0.5 and 0.1, respectively. In this game, the total number of joint action space is 32, as each player either chooses cc or ncc in a distributed manner.

The result shows that a player takes its actions of cc and ncc with arbitrary probabilities in the beginning of the game. Then, each player adjusts the probabilities of different actions by computing a series of regret values. After about 60 iterations, the *GTRA* game converges, i.e., the joint probability distribution of players converge to a set of CEs. We can observe that the 2 joint actions, i.e., (ncc, cc, ncc, ncc, ncc) and (ncc, ncc, ncc, ncc, cc) are chosen with probabilities of about 0.74 and 0.24, respectively. The other joint probability distributions are all of small values, i.e., less than 0.01 (we only plot the curves of 6 joint probabilities because of limited space). The result can be interpreted as follows. A strategy recommendation signal, generated at each player by using the MRM algorithm based on historical information, recommends the players on how to play the relay assignment game. That is, with the probabilities of about 0.74 and 0.24, the signal recommends the player n_2 and n_5 to act as relays, respectively. At the same time, the signal recommends the other players to remain silent. The joint probabilities except $p(ncc, cc, ncc, ncc, ncc)$ and $p(ncc, ncc, ncc, ncc, cc)$ are all of very small values, which means that the possibilities of joint actions except the two joint actions (ncc, cc, ncc, ncc, ncc) and (ncc, ncc, ncc, ncc, cc) can be neglected. It can be also observed in Fig. 2 that there are some deviations from the recommended strategies even after the game converges. The are two reasons for the strategy deviation. First, as a mixed strategy game, a player takes its actions with a probability distribution, which means that the player's strategy is of a *probabilistic* nature throughout the game. The other reason is the result of applying the MRM based algorithm. That is, a player takes each of its actions with the probability of $\frac{\delta}{m^{\gamma}}$ at least at every stage, as shown in (9), to ensure that all actions have chances to be evaluated. By doing so, a player explores the dynamic environment continuously.

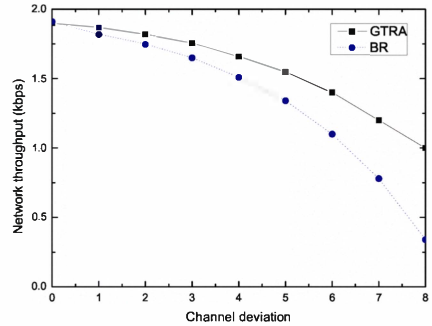

Fig. 3. Network throughput versus channel deviation σ

To investigate the performance of *GTRA* in a wireless channel with different fading, the average network throughput versus the channel deviation is shown in Fig. 3.

The simulation results show that *GTRA* outperforms *BR*, especially when the channel deviation σ becomes higher. We explain this as follows. The parameter of X_σ with standard deviation σ reflects the signal attenuation caused by the channel fading. That is, the higher the channel deviation σ, the more variation of the instantaneous strength of the received signals. In a wireless channel with higher variations, packets transmitted between a source and destination are more likely to be corrupted. Therefore, it is more critical to choose the best relay to cooperate with the communication link to help the packet delivery. In *BR*, a player, e.g., n_i, assumes that its opponent n_{-i}'s strategy is stationary, and estimates n_{-i}'s possible behavior by simply tracking the frequency of actions that has been taken by n_{-i} in previous stages. This approach works well in static environments but does not fit in dynamic environments. In contrast, *GTRA* is more adaptive in relay assignment, as players continuously evaluate the qualities of the actions that have been taken in previous stages, as well as evaluating the actions that have not been taken by using

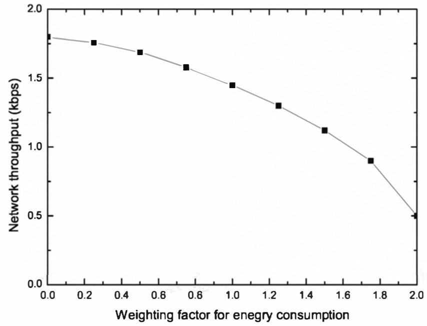

Fig. 4. Network throughput versus weighting factor for energy consumption

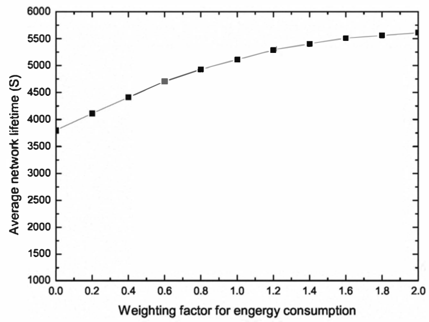

Fig. 5. Network lifetime versus weighting factor for energy consumption

the adaptive learning algorithm MRM. Then, players update their strategies by adjusting the probability distribution of the actions based on the payoff differences. Therefore, the flexible nature of the learning algorithm allows *GTRA* to adapt to dynamic environments, especially in networks with varying link qualities.

Fig. 4 and Fig. 5 show the impacts of the weighting factor ω_3 for energy consumption on the performance of network throughput and lifetime, respectively.

Fig. 4 shows that the network throughput decreases with the increment of ω_3. The reason is that when ω_3 is small, the benefit of acting as a relay (SNR improvement) outweighs the cost (energy consumption), thus players tend to retransmit packets to obtain higher payoffs which lead to a better performance on network throughput. However, when the ω_3 increases over a certain value, the cost becomes a dominating factor in computing the payoffs. In order to obtain higher payoffs, players tend to remain silent instead of retransmitting packets, which results in a lower network throughput.

Fig. 5 illustrates that the network lifetime alway increases with the increment of ω_3. The reason is that when ω_3 becomes sufficient large, the cost is an important factor in payoff computing. Thus, all players tend to remain silent instead of retransmitting packets, which leads to a longer network lifetime. However, as observed in Fig. 4 and Fig. 5, the longer lifetime is achieved by sacrificing the performance on network throughput. Therefore, a tradeoff must be considered when choosing the value of the weighting factor ω_3.

V. CONCLUSIONS AND FUTURE RESEARCH

In this paper, we have studied the problem of relay assignment for cooperative communications, and have formulated the problem as a non-cooperative, mixed strategy, and repeated game, in which each player plays against all the other players, and determines whether to cooperate with a communication link on a packet-by-packet basis in a distributed manner. To learn optimal cooperating strategies in dynamic environments, the MRM adaptive learning algorithm has been implemented at each player to adjust the probability distribution over actions, as well as orienting the game to converge to a set of CE(s), Simulation results have shown that *GTRA* outperforms *BR* in terms of network throughput, and can converge in a short period that enables it to work well in dynamic environments.

In future research, we will examine the issue of system fairness to ensure that each node to achieve an effort-balance, and to receive a fair share of the channel access in both retransmitting packets for other nodes and sending its own packets. Furthermore, we will also consider employing a power allocation scheme to prolong the network lifetime, as well as reducing concurrent transmission interferences to improve the network performance.

ACKNOWLEDGMENT

This research was supported by the Canadian Natural Sciences and Engineering Research Council under grant RGPIN 44286-09.

REFERENCES

[1] A. Nosratinia, T. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Communications Magazine*, vol. 42, no. 10, pp. 74–80, Oct. 2004.

[2] Y. Chen and S. Kishore, "A game-theoretic analysis of decode-and-forward user cooperation," *IEEE Transactions on Wireless Communications*, vol. 7, no. 2, May 2008.

[3] S. Sergiilia, F. Pancaldi, and G. M. Vitetta, "A game theory approach to selection diversity in wireless ad-hoc networks," in *Proc. The 8th 2009 IEEE international conference on Communications (ICC'09)*, Dresden, Germany, Jun. 2009, pp. 4184–4189.

[4] B. Wang, Z. Han, and K. R. Liu, "Distributed relay selection and power control for multiuser cooperative communication networks using stackelberg game," *IEEE Transactions on Mobile Computing*, vol. 8, no. 7, pp. 975–990, Jul. 2009.

[5] S. Hart and A. Mas-Colell, "A reinforcement procedure leading to correlated equilibrium," *Economic Essays*, pp. 181–200, Sep. 2001.

[6] B. Wang, Y. Wu, and K. R. Liu, "Game theory for cognitive radio networks: An overview," *Computer Networks*, vol. 54, no. 14, pp. 2537–2561, Oct. 2010.

[7] 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, Dec. 2004.

[8] D. G. Brennan, "Linear diversity combining techniques," *Proceedings of the IEEE*, vol. 91, no. 2, pp. 331–356, Feb. 2003.

[9] S. Hart and A. Mas-Colell, "A simple adaptive procedure leading to correlated equilibrium," *Econometrica*, vol. 68, no. 5, pp. 1127–1150, Sep. 2000.

[10] L. Chen, "A distributed access point selection algorithm based on no-regret learning for wireless access networks," in *Proc. The IEEE 71st Vehicular Technology Conference (VTC'10)*, Taiwan, May 2010, pp. 1–5.

[11] (2010) The Castalia website. [Online]. Available: http://castalia.npc.nicta.com.au/

[12] (2010) The OMNeT++ website. [Online]. Available: http://www.omnet++.org/