

Cell Selection in Two-Tier Femtocell Networks Using Q-Learning Algorithm

Xu Tan*, Xi Luan*, Yuxin Cheng*, Aimin Liu* and Jianjun Wu*†

*School of Electronics Engineering and Computer Science, Peking University, Beijing, P.R. China

Email: †just@pku.edu.cn

Abstract—Next-generation wireless networks will generate a heterogeneous network with micro base station (MBS) and femtocells where cell selection becomes crucial for balancing the utilization of the whole network. In this paper, we investigate cell selection problem in a two-tier femtocell network that contains a MBS and several femtocells with open/closed access methods and coverage areas. The selection process among groups of users in different service areas is formulated as a dynamic evolutionary game. In order to achieve an equilibrium, we present the Q-learning algorithm that can help distributed individual users adapt the situation and make cell selection decisions independently. With their own knowledge of the past, the users can learn to achieve the equilibrium without a centralized controller to gather other users information. Finally, simulation results present the convergence and effectiveness of the proposed algorithm.

Keywords—Cell selection, femtocell networks, evolutionary game, Q-learning, pure greedy algorithm.

I. INTRODUCTION

Next-generation wireless networks will generate a heterogeneous network and require high data rate. According to a survey in [1], more than 70% of data traffic originates indoors. Due to the penetration losses of the buildings, high data rates are hard to obtain indoors with the outdoor base stations. Recently, femtocells have gained great attention from academia and industry, because they can provide large indoor system capacity due to the short spread distance to users [2]. This leads to widespread deployment of femtocells. Thus, indoor users can switch to adjacent femtocells with better channel conditions to gain better service quality. Due to different access methods of femtocells, the switching process shall be considered with different standards [3]. According to the Third Generation Partnership Project (3GPP), there are three access methods defined as follows [4]:

- *Closed access*: Only a subset of users who have already registered can connect to the femtocell;
- *Open access*: All users are permitted to access the femtocell;
- *Hybrid access*: All users are allowed to access femtocells, but registered users are given higher priority.

The related problem of heterogeneous wireless network selection and cell selection was investigated in [5]. In [6] the author proposed an adaptive muting selection algorithm in the LTE-A network. Two algorithms were proposed for cell selection problems in OFDMA-based networks in [7].

However, these works ignored the congestions that are harmful to the network performance in cell selection process.

Evolutionary game was considered as a solution method of the cell selection problem in [8]. Although evolutionary game can be used to analyze the cell selection process, it requires a centralized controller to gather other users information which are hard to realize in many cases. Reinforcement learning (Q-learning) is an algorithm to solve problems with limited information exchange [9]. It was used to help distributed users make decisions based on their own information which is useful in the situation where complete information is not available [10]. A capacity performance under limited number of handovers was present in [11]. However, these works did not focus on overlapping femtocell networks and individual load balancing.

In this paper, we considered the equilibrium as the solution to cell selection competition. To obtain this solution, we present Q-learning algorithm. While the traditional evolutionary game needs information exchange in the coverage area, the reinforcement learning algorithm only utilizes local knowledge. Numerical results present the convergence and effectiveness of the proposed algorithm.

The rest of this paper is organized as follows. In Section II, the system model of a two-tier femtocell network with open and closed access is described. The Q-learning method to solve the cell selection problem is introduced in Section III. Finally, the simulation results are presented in Section IV and we draw the conclusion in Section V.

II. SYSTEM MODEL

We consider a two-tier femtocell network consisting of one MBS, M femtocells using open and closed access methods, as shown in Fig. 1. M femtocells are denoted by $\{F_1, F_2, \dots, F_M\}$ and N users are trying to obtain the service. In this network, the MBS has a large coverage area that can provide wireless service to a lot of users while M femtocells are small base stations that offer smaller coverage and support less users.

In this area shown in Fig. 1, femtocells are randomly deployed, the coverage may overlap and users can connect to different base stations. Users are divided into different groups according to their areas and are able to detect the presence of base stations and select the one which can offer them the highest channel capacities. MBS and femtocells will allocate their available channel capacities to the groups of users equally

and charge the users connected to them. In this scenario, we do not consider the mobility of the users.

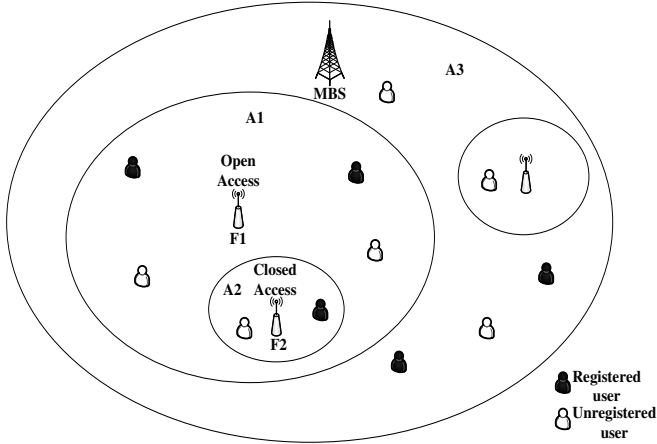


Fig. 1. System model of the two-tier femtocell network.

In this paper, the MBS covers a set of areas denoted as $\{A_1, A_2, A_3\}$, while F_1 and F_2 cover $\{A_1, A_2\}$ and $\{A_2\}$, respectively. Note that in the coverage area, there are registered users and unregistered users, who have different rights to access the closed access femtocell F_2 . We denote the type of a user as U and R to represent the unregistered and registered users, respectively. The MBS, open access femtocell and closed access femtocell can be denoted as $\{MB, OF, CF\}$, respectively. The number of users in area i with type j ($j = R, U$) is denoted as N_j^i . We use $x_{j,k}^i$ and $u_{j,k}^i$ to denote the proportion and utility of users in area i with type j who select base station k , respectively.

Each user's utility is measured by the capacity that the user obtains from the service provider and the price that the user pays to the base stations. We use a logarithmic revenue function $\mathbb{R}(c) = \log(1 + c)$ to denote the revenue of each user [12]. To express the payment of each user, we use a linear function $\mathbb{P}(n) = p_k n$ ($k \in \{MB, OF, CF\}$), where p_k is the coefficient of the pricing function [8]. The utility function of each user connected to base station k can be expressed as follows:

$$u_{j,k} = \mathbb{R} \left(\frac{C_{j,k}}{\sum_{i \in \Omega_k} N_j^i x_{j,k}^i} \right) - \mathbb{P} \left(\sum_{i \in \Omega_k} N_j^i x_{j,k}^i \right), \quad (1)$$

$$\mathbb{R} \left(\frac{C_{j,k}}{\sum_{i \in \Omega_k} N_j^i x_{j,k}^i} \right) = \log \left(1 + \frac{C_{j,k}}{\sum_{i \in \Omega_k} N_j^i x_{j,k}^i} \right), \quad (2)$$

$$\mathbb{P} \left(\sum_{i \in \Omega_k} N_j^i x_{j,k}^i \right) = p_k \sum_{i \in \Omega_k} N_j^i x_{j,k}^i, \quad (3)$$

where $C_{j,k}$ is the channel capacity provided by the base station k which is allocated to the users with the type j and Ω_k is the coverage area of base station k .

The selection process among groups of users in different service areas is formulated as a dynamic evolutionary game. Traditional evolutionary game use the replicator dynamics to model the rate of strategy change and reach the equilibrium, which can be expressed as follows:

$$\dot{x}_{j,k}^i = \lambda x_{j,k}^i (u_{j,k}^i - \bar{u}_j^i), \quad (4)$$

where \bar{u}_j^i is the average utility of the population and λ is the multiplier of the difference of the user's utility.

III. Q-LEARNING IN FEMTOCELL NETWORKS

In order to solve the cell selection problem with local information only, we introduce a Q-learning algorithm to obtain an equilibrium. A centralized controller can solve this problem by computing and analyzing the best result. However, a centralized controller is hard to achieve in the practical femtocell network. Q-learning can help users to learn based on their own past knowledge and take actions without the complete information of others.

A. Parameters

We define all parameters related to the Q-learning as follows:

- *Agent*: An agent is the decision maker that represents the user who wants to select the best base station that offers the best utility.
- *State*: A state represents the current serving cell of the users. The set of states is defined as $S = \{s | MB, OF, CF\}$, where MB , OF and CF represent MBS, open access femtocell (F_1) and closed access femtocell (F_2), respectively.
- *Action*: An action is the decision taken by the agent, which represents the target base station. The set of actions is defined as $A = \{a | MB, OF, CF\}$.
- *Reward*: Reward is the utility of the action. We define $R(s, a)$ as the expected utility for a user in state s who takes the action a .

B. Q-learning Algorithm

The Q-learning algorithm is an important algorithm of reinforcement learning. Thus, the Q-learning algorithm is used to solve the cell selection problem and obtain the equilibrium. Based on our scenario, we express it as follows:

With Q-learning algorithm, a user needs to find the strategy that can maximize the Q-value function. The expected reward of taking an action a in a current state s is given in this function. We express the Q-value function as:

$$Q(s, a) = u(s, a) + \gamma \sum_{v \in S} P_{s,v}(a) Q(v, b) \quad (5)$$

Where $u(s, a)$ is the expected utility of the action that we define in the previous subsection. γ is the discount factor ($0 \leq \gamma \leq 1$) and $P_{s,v}(a)$ is the probability from the initial state s to the new state v as a result of action a . $Q(v, b)$ is the Q-function

TABLE I. THE REINFORCEMENT LEARNING ALGORITHM

1. Initialize Q-values and states of all users in all groups
2. loop for the time from the beginning to the end
3. loop for all users in all groups
4. Select a base station using pure greedy algorithm
5. Computes the utility of each base station
6. Update Q-value through (7)
$Q_t(s, a), \forall s \in S, a \in A$
7. end loop for all users in all groups
8. end loop for the time from the beginning to the end

of the next state and action. For $\gamma = 1$, future utility can be as important as the current utility, while for $\gamma = 0$ future actions have no impact on the current state value.

Let $Q^*(s, a)$ be the maximum of the Q-function that determines the optimal action b for every possible next pair (v, b) .

$$Q^*(s, a) = u(s, a) + \gamma \sum_{v \in S} P_{s,v}(a) \max_{b \in A} Q^*(v, b) \quad (6)$$

Q-learning determines the optimal $Q^*(s, a)$ in an iterative process. The users use $Q_t(s, a)$ to estimate $Q(s, a)$ at time t . The Q-value function can be updated using the following equation at each step during the learning process:

$$Q_t(s, a) = (1 - \alpha)Q_{t-1}(s, a) + \alpha \left(u_t(s, a) + \gamma \max_b Q_{t-1}(s, a) \right) \quad (7)$$

Where α is the learning rate of this method. The knowledge of the expected utility shall be updated with the learning ability as follows:

$$\lim_{t \rightarrow \infty} Q_t(s, a) = Q(s, a). \quad (8)$$

We randomly choose the initial state of each user and set $Q_0(s, a) = 0, \forall s, a$ for all users. For each user, it is easy to make a choice by applying the greedy algorithm which means, given the current state s^* , user always selects the service provider with the largest utility in the previous step,

$$a_{t+1} = \arg \max_{s=s^*, a \in A} \{Q_t(s, a)\}. \quad (9)$$

Users make their decisions independently with their own knowledge. The Q-learning algorithm is described in Table I. We initialize their state and Q-values at time $t = 0$ for all users in all groups. For each user, given its current state, it can select the service provider with the largest Q-value. Then the user updates its Q-value in the new step.

IV. SIMULATION RESULTS

A. Parameters

We consider a two-tier femtocell network shown in Fig. 1. F_1 covers both the area A_1 and A_2 , F_2 provides service to a smaller area A_2 and to a certain small group of users who are registered. The MBS provides the largest service area A_1, A_2 and A_3 . We set the number of users in each area as $N^1=40$,

$N_R^2=10, N_U^2=10$ and $N^3=50$. We also assume that p_k in (3) is 0.01, 0.01 and 0.0075 for MBS, open access femtocell and closed access femtocell, respectively. The total capacities are $C_{MB} = 60, C_{OF} = 28$ and $C_{CF} = 6$. We assume that the multiplier of replicator dynamics as $\lambda = 1$. For the cell selection algorithm based on Q-learning, we assume that $\alpha = 0.2, \gamma = 0.2$. The initial proportion of users choosing each base station is varied.

B. Numerical Results

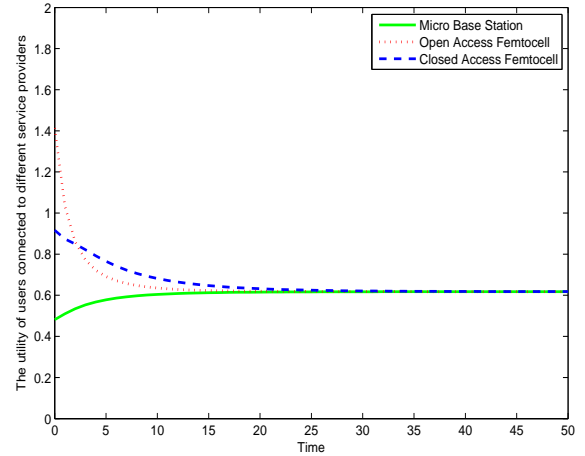


Fig. 2. Dynamics of utilities in the cell selection process.

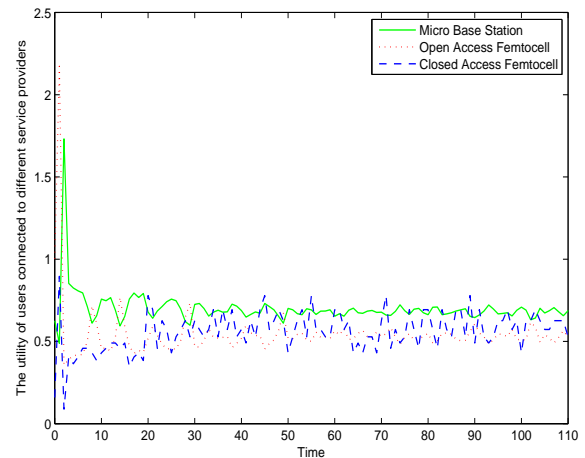


Fig. 3. Dynamics of utilities in the cell selection process using Q-learning algorithm.

The convergence properties of the cell selection algorithms based on replicator dynamics and Q-learning are presented in Fig. 2 and Fig. 3, respectively. We can see the dynamics of utilities in the cell selection process and the convergence

of the replicator dynamics. For Q-learning algorithm, utilities obtained by the users from all service providers reach a stable value at the equilibrium, so there is no incentive for any user to change its current base station. As a result, this equilibrium which can be achieved by local information, is a stable solution for the cell selection problem.

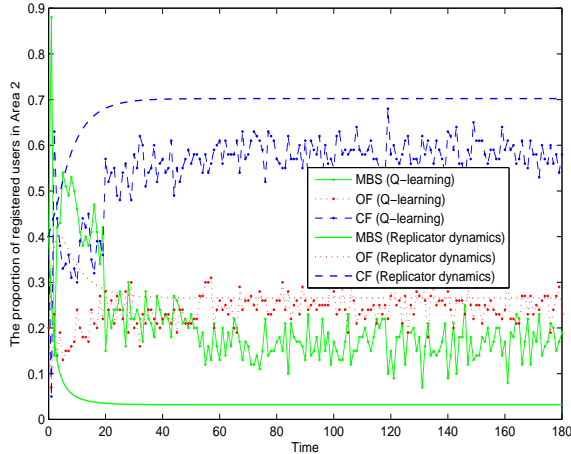


Fig. 4. Dynamics of proportions in the cell selection process.

Fig. 4 shows the proportion changes of registered users in A_2 over time. They can select their service providers from MBS, open access femtocell (F_1) and closed access femtocell (F_2) according to the utilities that the users obtain. Given the initial proportions of users in A_2 as $x_{R,MB}^2 = 0.4$, $x_{R,OF}^2 = 0.2$ and $x_{R,CF}^2 = 0.4$.

According to the rules of Q-learning, more users in A_2 tend to select femtocells rather than the MBS as shown in Fig. 4. We also observe that the proportions of registered users in A_2 become stable over time by using Q-learning algorithm. It means that users achieve a stable equilibrium that no user will change its service provider.

While the cell selection algorithm based on replicator dynamics can reach the evolutionary equilibrium faster, the Q-learning algorithm is still an attractive method from a practical viewpoint. This is due to the fact that, a centralized controller may not be available in actual system. In addition, the users do not want their information to be shared.

Note that an equilibrium can be found such that the payoffs of all users choosing different cells are approximately identical. In this paper, with the assumed system parameters, we can find the equilibrium that no user will change its selection. With this scenario, the Q-learning algorithm can approach the equilibrium well.

V. CONCLUSION

In this paper, we have analyzed the competition and cell selection process using the Q-learning algorithm. The cell selection process has been mathematically modeled by traditional evolutionary game. Equilibrium is considered to be the solution

of the cell selection problem. At the equilibrium, no one will change its selection because the utilities are stable. To achieve the equilibrium, we present Q-learning algorithm that each user can learn the performances and make decisions independently. According to the simulation results and analysis, the Q-learning algorithm can approach the equilibrium well.

ACKNOWLEDGMENT

This research was partly supported by the National Science Foundation of China (Grant No. NFSC # 61071083, # 61371073) and the National High Technology Research and Development Program of China (863 program, # 2012AA01A506). Correspondence should be addressed to Jianjun Wu, just@pku.edu.cn.

REFERENCES

- [1] V. Chandrasekhar, J. Andrews, and A. Gatherer, "Femtocell Networks: A Survey," *IEEE Communications Magazine*, vol. 46, no. 9, pp. 59-67, Sep. 2008.
- [2] S. P. Yeh, S. Talwar, S. C. Lee, and H. Kim, "WiMAX femtocells: a perspective on network architecture, capacity, and coverage," *IEEE Communications Magazine*, vol. 46, no. 10, pp. 58-65, Oct. 2008.
- [3] L. Song and J. Shen, *Evolved Cellular Network Planning and Optimization for UMTS and LTE*, CRC Press, 2010.
- [4] G. de la Roche, A. Valcarce, D. Lopez-Perez, and J. Zhang, "Access Control Mechanisms for Femtocells," *IEEE Communications Magazine*, vol. 48, no. 1, pp. 33-39, Jan. 2010.
- [5] G. T. Karetos, S. A. Kyriazakos, E. Groustiotis, F. D. Giandomenico, and I. Mura, "A hierarchical radio resource management framework for integrating WLANs in cellular networking environments," *IEEE Wireless Communications*, vol. 12, no. 6, pp. 11-17, Dec. 2005.
- [6] M. Feng, X. She, L. Chen, and Y. Kishiyama, "Enhanced Dynamic Cell Selection with Muting Scheme for DL CoMP in LTE-A," in *Proc. IEEE VTC Spring*, Taipei, May 2010.
- [7] D. Amzallag, R. Bar-Yehuda, D. Raz, and G. Scalosub, "Cell Selection in 4G Cellular Networks," in *Proc. IEEE INFOCOM*, Phoenix, AZ, Apr. 2008.
- [8] Z. Feng, L. Song, Z. Han, D. Niyato, and X. Zhao, "Cell Selection in Two-Tier Femtocell Networks with Open/Closed Access Using Evolutionary Game," in *Proc. IEEE WCNC*, Shanghai, Apr. 2013.
- [9] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction (Adaptive Computation and Machine Learning)*, MIT Press, 1998.
- [10] A. Galindo-Serrano, L. Giupponi, "Distributed Q-Learning for Aggregated Interference Control in Cognitive Radio Networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 4, pp. 1823-1834, May 2010.
- [11] C. Dhahri, T. Ohtsuki, "Q-learning cell selection for femtocell networks: Single- and multi-user case," in *Proc. IEEE Global Communications Conference(Globecom)*, Anaheim, CA, Dec. 2012.
- [12] Z. Jiang, Y. Ge, and Y. Li, "Max-utility wireless resource management for best-effort traffic," *IEEE Trans Wireless Communications*, vol. 4, no. 1, pp. 100-111, Jan. 2005.



Xu Tan received his bachelor degree in electronics from Peking University, Beijing, P.R. China, in 2012. Since 2012, he has been a postgraduate student in Institution of Advanced Communications, Peking University, China. His research interests are in the areas of wireless communications. Email: tanxu@pku.edu.cn.



Jianjun Wu received his B.S., M.S. and Ph.D. degree from Peking University, Beijing, P. R. China, in 1989, 1992 and 2006, respectively. Since 1992, he has joined the School of Electronics Engineering and Computer Science, Peking University, and has been appointed as an associate professor since 2002. His research interests are in the areas of satellite communications, wireless communications, and communications signal processing. *The corresponding author. Email: just@pku.edu.cn.