QoS-Aware Energy-Efficient Power Control in Two-Tier Femtocell Networks Based on Q-Learning

Zhicai Zhang, Xiangming Wen, Zhengfu Li, Shenghua He, Wenpeng Jing, Jun Zhao
Beijing Key Laboratory of Network System Architecture and Convergence
School of Information and Communication Engineering
Beijing University of Posts and Telecommunications, Beijing
Email: zzcai@bupt.edu.cn

***Abstract*—Due to the time-varying nature of wireless channels, deterministic quality of service (QoS) is hard to guarantee in wireless networks. In this paper, by integrating information theory with the principle of effective capacity, we formulate an energy efficiency optimization problem with statistical QoS guarantee in the uplink of two-tier femtocell networks. To solve the problem, we introduce a Q-learning mechanism based on Stackelberg game framework, in which macro-user acts as a leader, and knows all femto-users' transmit power strategy; while femto-users are followers, and only communicate with macrocell base station (MBS) not with other femtocell base stations (FBS). In Stackelberg game studying procedure, macro-user selects transmit power level firstly based on the best responses of femto-users, femto-users interact with environment directly, and find their best responses. At last, a distributed Q-learning algorithm is proposed. Simulation results show the proposed algorithm has a better performance in terms of convergence speed while providing delay QoS provisioning.**

I. Introduction

With the exponential growth of mobile data traffics, wireless communication networks play a more and more important role in the global emissions of carbon dioxide [1]. Obviously, the growing energy cost will cause a significant operational expense (OPEX) for mobile operators. On the other hand, the limited battery resources can not meet the massive data rate requirement either. Based on this background, the concept of green communication is proposed to develop environment friendly and energy efficient technologies for future wireless communications. Therefore, adopting energy-aware communication technologies is a trend for the design of next generation wireless networks.

Energy efficiency was firstly proposed by D.J. Goodman, *etal*., which is defined as the number of error-free delivered bits for each energy-unit used in transmission and is measured in bit/joule [2]. In recent years, there have been many researches on energy-efficient resource management [3, 4]. A low complexity energy-efficient subchannel allocation scheme is proposed in [3], but the method does not consider interference caused by neighbors. In [4], joint subchannel allocation and power control are modeled as a potential game to maximize energy efficiency of multi-cell uplink OFDMA systems, but QoS guarantees are without consideration.

Besides energy-efficient radio resource management, femtocell network is another promising technology to save energy. Since this type of deployment strategy brings transmitters closer to receivers, and reduces the penetration loss and path loss. As we know, femtocell base station is installed by end-users, who have not enough professional skills to configure parameters of FBS. On this account, FBS should have self-learning ability to automatically configure and optimize the its operating information, e.g. transmit power assignment. In recent years, reinforcement learning mechanism, such as Q-learning, is widely used in radio resource allocation of wireless network [5–7], however, most of existing works are focusing in cognitive radio networks.

Furthermore, providing delay QoS guarantee while minimizing energy consuming is a key issue in green communication systems. For instance, in real-time service, such as in multimedia video conference and live broadcast of sporting events, *etc.*, latency time is a key QoS metric. Since the time-varying channel, deterministic delay QoS guarantee mechanisms used in wired networks can not take affect in wireless networks [8]. To address this issue, statistical QoS provisioning, in term of delay exponent and effective capacity has become an effective method to support real-time service in wireless networks [9–11].

In this paper, we will investigate energy efficient power control in the uplink two-tier femtocell networks with delay QoS guarantees. Based on the concept of effective capacity, we formulate an energy efficiency optimization problem with statistical QoS guarantee. To solve the problem, a transmit power learning mechanism based on Stackelberg game is proposed. In the learning procedure, macro-user behaves as a leader, and can communicate with femto-users; while femto-users act as followers, and only know leader's power strategy not other followers'. Besides, leader knows followers' best responses of transmit power and selects strategy firstly; followers move subsequently. At last, a distributed Q-learning procedure based on Stackelberg game is proposed. Simulation results show the proposed algorithm has a better performance in terms of convergence speed compared with a conjecture based multi-agent Q-learning (CMAQL) algorithm which involves no information exchange between each player [12].

The rest of paper is organized as follows. In Section II, we briefly discuss effective capacity and formulate an energy efficiency optimization problem with statistical delay provisioning. A Q-learning mechanism based on Stackelberg game framework and a distributed Q-learning algorithm are

978-1-4799-5141-3/14/$31.00 ©2014 IEEE

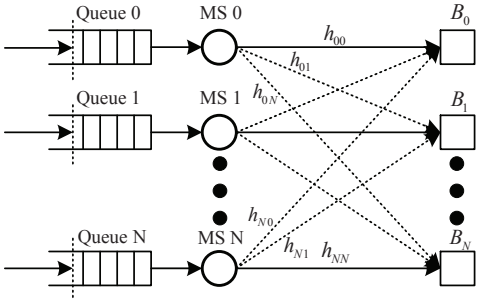

Fig. 1: System model of two-tier femtocell networks

proposed in Section III. Simulation results are shown in Section IV. In Section V, we conclude the paper.

II. System Model

A. System Description

The scenario considered in this paper is shown in Fig. 1, where N femtocells are overlaid in a macrocell, which constitutes a two-tier femtocell network. FBSs are in closed subscriber group (CSG) mode, i.e., mobile stations (MSs) that are not the members of the CSG, are not allowed to access the CSG FBSs.

Let B_i $(i \in \mathcal{N})$ denote the base station (BS), where $\mathcal{N} = \{0, 1, 2, \cdots, N\}$. B_0 denotes the MBS, and B_i $(i \in \mathcal{N}, i \neq 0)$ is FBS. We assume each MS will be allocated only one subchannel, and to avoid the intra-cell interference during each frame slot, the same frequency can be occupied by only one active MS in each cell. Let $i \in \mathcal{N}$ denote the index of scheduled user in B_i.

The received signal-to-interference-and-noise-ratio (SINR) of MS i in B_i can be expressed as

$$\gamma_i(p_i, \mathbf{p}_{-i}) = \frac{p_i h_{ii}}{\sum_{j \neq i} p_j h_{ij} + \sigma_i^2}, \forall i \in \mathcal{N}, \tag{1}$$

where, p_i denotes the transmit power of MS i, and $\mathbf{p}_{-i}, (-i \in \mathcal{N})$ denotes the transmit power of other MSs except MS i. h_{ii} and h_{ij} are the channel gains from MS i to BS B_i, B_j respectively. σ_i^2 is the variance of additive white Gaussian noise (AWGN) of MS i.

According to the Shannon's capacity formula, the ideal achievable data rate of MS i is

$$R_i(p_i, \mathbf{p}_{-i}) = w \log_2(1 + \gamma_i(p_i, \mathbf{p}_{-i})), \tag{2}$$

where w is the bandwidth of the occupied channel.

B. Effective Capacity

The concept of statistical delay guarantee has been extensively studied in the effective bandwidth theory [13]. Based on large deviation principle, author Chang in [13] has pointed out that with sufficient condition, for a dynamic queueing system with stationary ergodic arrival and service processes, the queue length process $Q(t)$ converges to a random variable $Q(\infty)$, such that,

$$\lim_{Q_{th} \to \infty} \frac{\log(\Pr\{Q(\infty) > Q_{th}\})}{Q_{th}} = -\theta \tag{3}$$

exists. Where Q_{th} is queue length bound, and $\theta > 0$ is the decay rate of the tail distribution of the queue length $Q(\infty)$. If $Q_{th} \to \infty$, we get the approximation of the buffer violation probability, i.e., $\Pr\{Q(\infty) > Q_{th}\} \approx e^{-\theta Q_{th}}$. We can find that a larger θ corresponds to a faster decay rate, which implies more strict QoS constraints, while a smaller θ leads to a slower decay rate, which means looser QoS requirement. Similarly, the delay-outage probability can be approximated by [8], i.e., $\Pr\{Delay > D_{th}\} \approx \xi e^{-\theta \delta D_{th}}$. In which, D_{th} is the maximum tolerable delay, ξ is the probability of a non-empty buffer, and δ is the maximum constant arrival rate.

The concept of effective capacity is proposed by Wu, *etal.*, in [8], which is defined as the maximum constant arrival rate that the time-varying channel can support while guaranteeing a statistical delay requirement specified by the QoS exponent θ. The effective capacity is formulated as

$$E^c(\theta) = -\lim_{K \to \infty} \frac{1}{K\theta} \ln(\mathbb{E}\{e^{-\theta \sum_{k=1}^{K} S[k]}\}), \tag{4}$$

where, $\{S[k] | k = 1, 2, \cdots, K\}$ denotes the discrete-time, stationary and ergodic stochastic service process. $\mathbb{E}\{\cdot\}$ is the expectation over the channel state.

In this paper, we assume that the channel fading coefficients stay constant over the frame duration T, and vary independently for each frame and each MS. According to (2), $S_i[k] = TR_i[k]$ is obtained. Based on above analysis, the effective capacity of MS i can be simplified as

$$E_i^c(\theta_i) = -\frac{1}{\theta_i T} \ln(\mathbb{E}\{e^{-\theta_i T R_i(p_i, \mathbf{p}_{-i})}\}). \tag{5}$$

C. Problem formulation

The energy efficiency under statistical delay guarantees of MS i is defined as the ratio of the effective capacity to the totally consumed energy as following

$$\eta_i(p_i, \mathbf{p}_{-i}) = \frac{E_i^c(\theta_i)}{p_i + p_c}. \tag{6}$$

In (6), p_c represents the average energy consumption of device electronics, including mixers, filters, and digital to-analog converters, and excludes that of the power amplifier.

Our target is to maximize the energy-efficiency of each MS, while satisfying delay-QoS guarantee. Therefore, the corresponding problem is

$$\max \frac{-\ln(\mathbb{E}\{e^{-\theta_i T R_i(p_i, \mathbf{p}_{-i})}\})}{\theta_i T(p_i + p_c)} \tag{7a}$$

$$p_i \geq p_{\min}, \forall i \in \mathcal{N}, \tag{7b}$$

$$p_i \leq p_{\max}, \forall i \in \mathcal{N}, \tag{7c}$$

$$\theta_i > 0, \forall i \in \mathcal{N}, \tag{7d}$$

where, $p_{\min}$ and $p_{\max}$ are the lower and upper bounds of each MS's transmit power respectively.

III. Q-LEARNING BASED ON STACKELBERG GAME FRAMEWORK

As we know, FBS is installed by end-users, who have not enough professional skill to configure parameters of FBS. On this account, FBS should have self-learning ability to automatically configure and optimize the FBS's operating information.

In this section, we will employ a reinforcement learning mechanism based on Stackelberg game framework to implement the energy efficient transmit power allocation while guaranteeing delay-QoS requirement.

To be compatible with reinforcement learning mechanism [14], the transmit power of MS i is discretized as $\mathcal{P}_i = (p_{i,v_i}|v_i = 1, 2, \cdots, V_i)$. The probability of MS i choosing transmit power p_{i,v_i} at time slot t is π^t_{i,v_i} ($\pi^t_{i,v_i} \in \pi^t_i$), and $\pi^t_i = (\pi^t_{i,v_i}|v_i = 1, 2, \cdots, V_i)$, which satisfies $\sum_{v_i=1}^{V_i} \pi^t_{i,v_i} = 1$.

Then, the expected utility of MS i is given by

$$u_i(\pi^t_i, \boldsymbol{\pi}^t_{-i}) = \mathbb{E}\{\eta_i(\mathbf{p})|\pi^t_i, \boldsymbol{\pi}^t_{-i}\} = \sum_{\mathbf{p}\in\mathbf{P}} \eta_i(\mathbf{p}) \prod_{j\in\mathcal{N}} \pi^t_{j,v_j}, \tag{8}$$

where, $\mathbf{p} = (p_{0,v_0}, \cdots p_{i,v_i} \cdots, p_{N,v_N}) \in \mathbf{P}$ is the power level of all MSs at time slot t, and $\mathbf{P} = \times_{i\in\mathcal{N}}\mathcal{P}_i$. $\boldsymbol{\pi}^t_{-i}$ ($-i \in \mathcal{N}$) denotes the strategies of all MSs except MS i.

A. Stackelberg Game Framework

The Stackelberg game model [15] is very suitable for two-tier femtocell networks, where MS 0 is formulated as a leader, and MSs $\{i|i \in \mathcal{N}, i \neq 0\}$ are modeled as followers. In Stackelberg game framework, leader knows strategy information of all followers and selects action firstly; followers can receive the policy of leader and move subsequently.

Based on above analysis, it is easy to find that the MS 0's objective is to maximize its revenue as

$$\underline{\textbf{Problem 3.1}} \quad \max u_0(\pi_0, \boldsymbol{\pi}_{-0}), \tag{9}$$

and the objective of MS $i, (i \in \mathcal{N}, i \neq 0)$ is

$$\underline{\textbf{Problem 3.2}} \quad \max u_i(\pi_i, \boldsymbol{\pi}_{-i}). \tag{10}$$

Due to the fact, FBSs are deployed by end-users randomly, and there is no communications or coordination among femtocells, who are selfishly pursuing their own profits. The problem 3.2 can be modeled as a non-cooperative power allocation subgame $G = [\{i\}, \{\mathcal{P}_i\}, \{u_i\}]$ $(i \in \mathcal{N}, i \neq 0)$.

Theorem 1: Given MS 0's strategy π_0, there exists a mixed strategy $\{\pi^*_i, \boldsymbol{\pi}^*_{-i}\}$ satisfies $u_i(\pi^*_i, \boldsymbol{\pi}^*_{-i}) \geq u_i(\pi_i, \boldsymbol{\pi}^*_{-i})$, which is a Nash equilibrium (NE) point.

proof: As it has been shown in [15], every finite strategic-game has a mixed strategy equilibrium, i.e., there exists $NE(\pi_0)$ for given π_0.

lemma 1: The joint **problem 3.1** and **problem 3.2** exists a Stackelberg equilibrium (SE) point $\{\pi^*_0, \pi^*_i, \boldsymbol{\pi}^*_{-i}\}$ $(\forall i \in \mathcal{N}, i \neq 0)$, which is a mixed strategy.

Based on **Theorem 1**, it is easy to verify the existence of SE point, and the process of the proof is omitted here for brevity. In Section III-B, we will employ reinforcement learning mechanism, called Q-learning, to find SE point.

B. Q-learning

Q-learning is a common reinforcement learning method that is used widely in self-organized femtocell networks. Each BS acts as an intelligent agent to maximize their profits by directly interacting with the environment.

We define $p_{i,v_i} \in \mathcal{P}_i$ $(\forall i \in \mathcal{N})$ as actions of Q-learning model, and $\boldsymbol{\pi}^t_{-i}$ $(-i \in \mathcal{N})$ are environment states. Q-learning represents the knowledge by means of a Q-function, whose Q-value is defined as $Q^{t+1}_i(p_{i,v_i}, \boldsymbol{\pi}^{t+1}_{-i})$, and is updated according to

$$\begin{aligned} &Q^{t+1}_i(p_{i,v_i}, \boldsymbol{\pi}^{t+1}_{-i}) = \\ &Q^t_i(p_{i,v_i}, \boldsymbol{\pi}^{t+1}_{-i}) + \alpha^t(r_i(p_{i,v_i}, \boldsymbol{\pi}^{t+1}_{-i}) - Q^t_i(p_{i,v_i}, \boldsymbol{\pi}^{t+1}_{-i})), \end{aligned} \tag{11}$$

where, $\alpha^t \in [0, 1)$ is the learning rate. In (11), $r_i(p_{i,v_i}, \boldsymbol{\pi}^{t+1}_{-i})$ is the reward function of MS i when selecting p_{i,v_i} and other MSs' strategies are $\boldsymbol{\pi}^{t+1}_{-i}$. The relationship between reward and utility function of MS i is

$$u_i(\pi^t_i, \boldsymbol{\pi}^t_{-i}) = \sum_{v_i=1}^{V_i} \pi_{i,v_i} r_i(p_{i,v_i}, \boldsymbol{\pi}^t_{-i}).$$

Each BS updates its strategy based on Boltzmann distribution [14], which is formally described as

$$\pi^t_{i,v_i} = \frac{\exp(Q^t_i(p_{i,v_i}, \boldsymbol{\pi}^{t+1}_{-i})/\tau)}{\sum_{v_i=1}^{V_i} \exp(Q^t_i(p_{i,v_i}, \boldsymbol{\pi}^{t+1}_{-i})/\tau)}, \tag{12}$$

where, τ $(\tau > 0)$ is temperature parameter. Higher value of τ causes the probabilities of all actions of MS i to be nearly equal; lower value of τ leads to the probability of actions bigger difference with respect to their Q-values.

C. Q-learning procedure

In this section, we will investigate QoS-aware energy-efficient power allocation in sparsely deployed and densely deployed femtocell networks respectively, by employing Q-learning mechanism based on Stackelberg game framework.

1) *Sparsely deployed scenario*

In sparsely deployed femtocell networks, e.g. in rural area, due to the path loss and penetration loss, the interference between FBSs can be ignored. As we have assumed before, MBS knows completely strategies of all FBSs, and updates its Q-value by (11). The reward function of MS 0 is following

$$r_0(p_{0,v_0}, \boldsymbol{\pi}^{t+1}_{-0}) = \sum_{\mathbf{p}\in\mathbf{P}} \{\eta_0(\mathbf{p})\delta^{t+1}_{-(0,v_0)}\}, \tag{13}$$

where, $\delta^{t+1}_{-(0,v_0)} = \prod_{j\in\mathcal{N}, j\neq 0} \pi^{t+1}_{j,v_j}$ denotes the probability of actions vector $\mathbf{p}_{-(0,v_0)} = (p_{1,v_1}, \cdots p_{i,v_i} \cdots, p_{N,v_N})$.

For MS i $(\forall i \in \mathcal{N}, i \neq 0)$, due to the fact that, FBSs can receive MBS's transmit power strategy, and there is no

interference between FBSs, the reward function of MS i is

$$r_i(p_{i,v_i}, \pi_0^{t+1}) = \sum_{v_0=1}^{V_0} \delta_{-(i,v_i)}^{t+1} \eta_i(p_{i,v_i}, p_{0,v_0}), \qquad (14)$$

where $\delta_{-(i,v_i)}^{t+1} = \pi_{0,v_0}^{t+1}$.

2) *Densely Deployed Scenario*

In densely deployed femtocell networks, e.g., in urban areas, the FBSs are deployed closely to each other, the mutual interference between FBSs can not be ignored.

In this scenario, the reward function of MS 0 is according to (13). Similar to (14), the reward function of MS i ($\forall i \in \mathcal{N}$, $i \neq 0$) in this scenario is

$$r_i(p_{i,v_i}, \pi_0^{t+1}) = \sum_{v_0=1}^{V_0} \delta_{-(i,v_i)}^{t+1} \hat{\eta}_i(p_{i,v_i}, p_{0,v_0}). \qquad (15)$$

Since there is no communication or cooperation between FBSs, if the selected power level at time shot $t+1$ satisfies $p_{i,v_i}^{t+1} = p_{i,v_i}$, then $\hat{\eta}_i^{t+1}(p_{i,v_i}, p_{0,v_0})$ is estimated by (16), else $\hat{\eta}_i^{t+1}(p_{i,v_i}, p_{0,v_0}) = \hat{\eta}_i^t(p_{i,v_i}, p_{0,v_0})$.

$$\hat{\eta}_i^{t+1}(p_{i,v_i}, p_{0,v_0}) = \frac{\eta_i(p_{i,v_i}, \mathbf{p}_{-i}) - \hat{\eta}_i^t(p_{i,v_i}, p_{0,v_0})}{\rho^t(p_{i,v_i}, p_{0,v_0}) + 1} + \hat{\eta}_i^t(p_{i,v_i}, p_{0,v_0}). \qquad (16)$$

In (16), $\eta_i(p_{i,v_i}, \mathbf{p}_{-i})$ is the real value when $p_{i,v_i}^{t+1} = p_{i,v_i}$, which can be calculated by the feedback information from FBS B_i. $\rho^t(p_{i,v_i}, p_{0,v_0})$ is the times number when the MS 0's transmit power is p_{0,v_0}, and MS i selects power level p_{i,v_i} until time shot t [14].

3) *Distributed Q-learning algorithm*

Theorem 2: the proposed algorithm can discover a SE mixed strategy.

Due to the limited space, the convergence of the proposed algorithm can be found in [16].

Algorithm 1: Distributed Q-learning algorithm

Step 1: Initialization: for $t = 0$, $Q_i^t(p_{i,v_i}, \boldsymbol{\pi}_{-i}^t), \forall i \in \mathcal{N}$; power discretization: $\mathbf{p}_i = (p_{i,1}, \cdots, p_{i,v_i}, \cdots, p_{i,V_i})$;
Learning:
Step 2: Update $t = t + 1$;
Step 3: Update π_i^t according to (12);
Step 4: Update MS 0's transmit power according to $p_{0,v_0^*} = \arg\max(Q_0^t(p_{0,v_0}, \boldsymbol{\pi}_{-0}^t))$, and send the value of π_0^t to FBS.
Step 5: Update MS i's ($i \neq 0$) transmit power according to $p_{i,v_i^*} = \arg\max(Q_i^t(p_{i,v_i}, \boldsymbol{\pi}_{-i}^t))$, and send the value of π_i^t to MBS.
Step 6: Calculate MS 0's reward according to (13), calculate MS i's ($i \neq 0$) reward by (14) or (15).
Step 7: Update MS i's Q-value by (11).
Step 8: Back to **Step 2**.
End learning

IV. Simulation and Numerical Analysis

In this section, we present the simulation for the proposed algorithm. A conjecture based multi-agent Q-learning (CMAQL) algorithm is also simulated for comparing with the proposed algorithm [12].

We will use Monte Carlo method in the simulation process. Macro-user and femto-users are distributed randomly in the two-tier femtocell networks and share the same spectrum with $w = 200kHz$. The channel-fading is modeled as Rayleigh block-fading channels, the fading-block duration $T = 1ms$. Noise spectral density is $N_0 = -174dBm/Hz$. The channel gain for macro-user and femto-users are λL^{-3} and λL^{-4} respectively, where L is the transmit-receiver separation in meters, and λ=2×10^{-4} [17].

The additional circuit power p_c is $10dBm$ for all users, the lower bound of transmit power for each user is $p_{\min} = 10dBm$, and upper bounds for femto-users and macro-user are $p_{\max} = 20dBm$ and $p_{\max} = 30dBm$ respectively. In the Q-learning procedure, the transmit power region $[p_{\min}, p_{\max}]$ is divided into d parts equally, and we consider $d = 3, 10, 20$ respectively in the simulation.

Fig. 2 shows expected utilities with respect to the QoS exponent. When the value of θ is small, i.e. $\theta \leq 10^{-4}$, there is no significant expected utility change. This is because the smaller of QoS exponent, the looser requirement of delay is, and effective capacity approaches to Shannon capacity, which is independent of arrival rate and delay requirement. Instead, when the value of θ is larger, and the delay requirement is tighter, effective capacity and expected utility decrease correspondingly. On the other hand, the transmit power discretization induces the best transmit power error, and the smaller of d results in a higher loss of expected utility.

Fig. 3 and Fig. 4 show the convergence of the proposed algorithm. From these figures, we can find that, the proposed algorithm has a faster convergence speed than CMAQL algorithm. The reason is that femto-users in the proposed Q-learning mechanism can share transmit power strategy with macro-user, while the value of $\delta_{-(i,v_i)}^{t+1}$ is estimated by only the past experiences in CMAQL algorithm.

V. Conclusion

In this paper, we investigate the energy efficient power control in two-tier femtocell networks with considering delay-QoS guarantee. To enhance the self-configuring and self-optimizing abilities of FBSs, we propose a Q-learning mechanism based on Stackelberg game framework. In the learning procedure, macro-user is a leader, who knows transmit power strategies of all femto-users and chooses power level firstly; while femto-users acting as followers can communicate with only leader and move subsequently. At last, a distributed Q-learning algorithm based on Stackelberg game is proposed. Simulation results show the proposed algorithm has a faster convergence speed than CMAQL algorithm.

VI. Acknowledgement

This work was supported by the National Natural Science Foundation of China (61101109, 61271179), and the Beijing Municipal Science and Technology Project (No. D121100002112002).

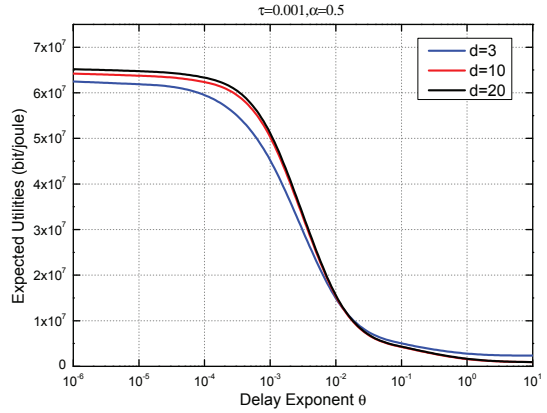

Fig. 2: Expected utilities versus different QoS exponent

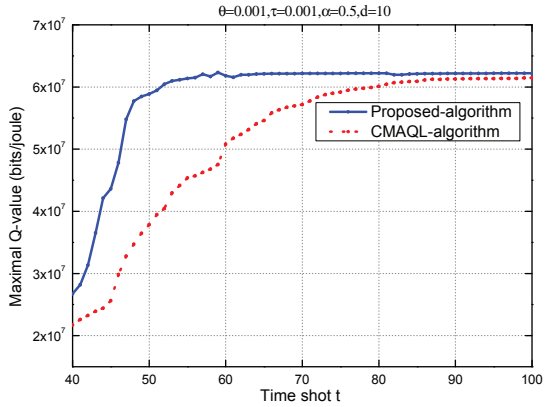

Fig. 3: The convergence of Q-learning mechanism

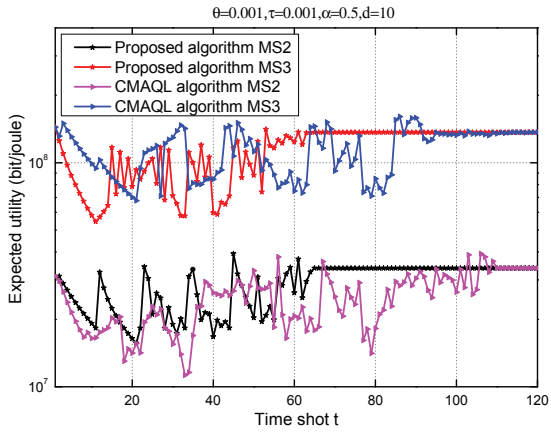

Fig. 4: The convergence of expected utilities

References

[1] G. Li, Z. Xu, C. Xiong, C. Yang, S. Zhang, Y. Chen, and S. Xu, "Energy-efficient wireless communications: tutorial, survey, and open issues," *IEEE Wireless Commun.*, vol. 18, no. 6, pp. 28–35, 2011.

[2] C. Saraydar, N. B. Mandayam, and D. Goodman, "Pareto efficiency of pricing-based power control in wireless data networks," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 1999, pp. 231–235 vol.1.

[3] G. Miao, N. Himayat, G. Li, and S. Talwar, "Low-complexity energy-efficient scheduling for uplink ofdma," *IEEE Trans. on Commun.*, vol. 60, no. 1, pp. 112–120, 2012.

[4] A. Zappone, G. Alfano, S. Buzzi, and M. Meo, "Energy-efficient non-cooperative resource allocation in multi-cell ofdma systems with multiple base station antennas," in *IEEE GreenCom*, 2011, pp. 82–87.

[5] L. Wang, X. Chen, Z. Zhao, and H. Zhang, "Exploration vs exploitation for distributed channel access in cognitive radio networks: A multi-user case study," in *11th International Symposium on Communications and Information Technologies (ISCIT)*, 2011, pp. 360–365.

[6] O. van den Biggelaar, J. Dricot, P. De Doncker, and F. Horlin, "A new distributed algorithm for the allocation of cognitive radio sensing times," in *IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, 2012, pp. 1208–1213.

[7] F. Panahi and T. Ohtsuki, "Optimal channel-sensing policy based on fuzzy q-learning process over cognitive radio systems," in *IEEE International Conference on Communications (ICC)*, 2013, pp. 2677–2682.

[8] D. Wu and R. Negi, "Effective capacity: a wireless link model for support of quality of service," *IEEE Transactions on Wireless Communications*, vol. 2, no. 4, pp. 630–643, 2003.

[9] D. Qiao, M. Gursoy, and S. Velipasalar, "Energy efficiency in multiaccess fading channels under qos constraints," in *IEEE International Conference on Communications (ICC)*, 2012, pp. 2318–2322.

[10] L. Musavian and T. Le-Ngoc, "Energy-efficient power allocation for delay-constrained systems," in *IEEE Global Communications Conference (GLOBECOM)*, 2012, pp. 3554–3559.

[11] C. Xiong, G. Li, Y. Liu, and S. Xu, "Qos driven energy-efficient design for downlink ofdma networks," in *IEEE Global Communications Conference (GLOBECOM)*, 2012, pp. 4320–4325.

[12] X. Cheng, Z. Zhao, H. Zhang, and T. Chen, "Conjectural variations in multi-agent reinforcement learning for energy-efficient cognitive wireless mesh networks," in *IEEE Wireless Communication and Networking Conference (WCNC)*, 2012, pp. 820–825.

[13] C.-S. Chang, "Stability, queue length, and delay of deterministic and stochastic queueing networks," *IEEE Transactions on Automatic Control*, vol. 39, no. 5, pp. 913–931, 1994.

[14] C. Long, Q. Zhang, B. Li, H. Yang, and X. Guan, "Non-cooperative power control for wireless ad hoc networks with repeated games," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 6, pp. 1101–1112, 2007.

[15] D.Fudenburg and J.Tirole, "Game theory," in *The MIT press*, 1991.

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

[17] V. Chandrasekhar, J. Andrews, T. Muharemovic, Z. Shen, and A. Gatherer, "Power control in two-tier femtocell networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4316–4328, Aug. 2009.