Copyright ⓒ 2013 KSII

Multi-agent Q-learning based Admission Control Mechanism in Heterogeneous Wireless Networks for Multiple Services

Jiamei Chen, Yubin Xu, Lin Ma, Yao Wang
Communication Research Center, Harbin Institute of Technology,
Harbin 150080, P.R. China
[e-mail: chenjiamei5870@hit.edu.cn, ybxu@hit.edu.cn, malin@hit.edu.cn, wangyaowh2005@hit.edu.cn]
*Correspondent author: Yubin Xu

Received April 3, 2013; revised June 7, 2013; revised August 4, 2013; accepted September 22, 2013; published October 29, 2013

Abstract

In order to ensure both of the whole system capacity and users QoS requirements in heterogeneous wireless networks, admission control mechanism should be well designed. In this paper, Multi-agent Q-learning based Admission Control Mechanism (MQACM) is proposed to handle new and handoff call access problems appropriately. MQACM obtains the optimal decision policy by using an improved form of single-agent Q-learning method, Multi-agent Q-learning (MQ) method. MQ method is creatively introduced to solve the admission control problem in heterogeneous wireless networks in this paper. In addition, different priorities are allocated to multiple services aiming to make MQACM perform even well in congested network scenarios. It can be observed from both analysis and simulation results that our proposed method not only outperforms existing schemes with enhanced call blocking probability and handoff dropping probability performance, but also has better network universality and stability than other schemes.

Keywords: Admission control, multi-agent Q-learning, reinforcement learning, heterogeneous wireless network, resource management

This research was supported by the National Natural Science Foundation of China (Grant No. 61071105). The authors will be appreciated to every team member for their help and suggestions.

http://dx.doi.org/10.3837/tiis.2013.10.003

1. Introduction

One of the most crucial challenges in the next generation wireless networks is the fact that no one single wireless network technology can simultaneously provide seamless coverage and continuous high level QoS services when users roam around several representative spaces, just as office, campus and airport [1]. It is expected to combine existing different wireless access network technologies and take the complementary advantages of them to satisfy the increasing requirements of users. Among all wireless access networks, WCDMA possesses wide coverage but low data rate while WLAN offers relatively high data rate in small area (hotspots), so WCDMA/WLAN heterogeneous networks become attractive because of the perfect complementary feature of the two subnets.

Admission control is always a key technology in any form of access networks, for it affects the performance of networks a lot [2][3][4]. Heterogeneous networks [5][6] aim to centralized control different access networks in just one unified system, thus it has to cope with multiplex mobility[7] patterns and more services types in a cooperative way. Consequently, admission control in heterogeneous networks faces more complicated environment. Furthermore, it intends to pursue maximum network profit and desired QoS quality of users simultaneously, which also brings difficulty for the design of admission control mechanism.

Although a number of admission control mechanisms for heterogeneous networks have been proposed in many documents, an overall optimization is still on the way. Many methods only consider QoS requirements of users, just as access fee, bandwidth and the power consumption [8], and how to improve the integrated system revenue has not been carefully addressed. Some other researches [9][10][11] develop access solution for new call and handoff requests of ongoing users separately, so joint method is scarce. Multiple attribute decision (MAD) method in literature [12] is a classic measurement scheme which could meet the needs of users and system. MAD uses cost function with several parameters and corresponding weights. However, the weights given by experts are not accurate and can not adapt to the dynamic states of the network. The introduction of a model based markov decision process (MDP) [13] algorithm by Fei Yu, particularly the transition probability function in it makes the self-adaptability of access control algorithm possible. Unfortunately, the state space dimensions and computational complexity will increase dramatically with the increase of users for real networks. Yung-han Chen [14] makes fine attempts in this field by proposing the fuzzy Q-learning admission control (FQAC) system. He focuses on the access mechanism for both new and ongoing users. Single-agent Q-learning [15][16] in this system is a model-free reinforcement learning method that is no longer restrained by state dimensions. It achieves the optimal policy by self-learning process without knowing the framework of transition probability. Nevertheless, FQAC only concentrates on the policy of accepting or rejecting users as long as one call request happens, but does not take care of the other resource managements. In fact, service priority scheme and resource reservation should be taken into account if network congestion appears. Furthermore, all methods above don't handle multiple services effectively which is especially a key issue in heterogeneous network. Although some of them [17] have mentioned it, or even given out simulation results.

Based on issues above, MQACM for WCDMA/WLAN heterogeneous networks is presented in this paper. WCDMA/WLAN heterogeneous networks environment in this paper includes one WCDMA subnet and one WLAN subnet. WLAN subnet overlaps on WCDMA and has smaller radius. The area that covered by both WCDMA subnet and WLAN subnet is

Copyright ⓒ 2013 KSII

called double-coverage area, and the area just covered by WCDMA subnet is called single-coverage area. Previous documents [18][19] demonstrate that admission control for heterogeneous networks become complicated only in double-coverage area no matter a new call happens or an ongoing user needs to handoff from single-coverage area to double-coverage area. As a result, here we just pay more attention to the double-coverage area, so-called target region below.

The main contributions and distinctions in our MQACM are four-fold. Firstly, we devise our MQACM primarily from the system perspective by employing load parameters as network states, and simultaneously satisfy users QoS requests by inversely inputting the QoS parameters as a feedback after the decision is carried out. This feedback is assistant information that could judge how correct the decision is. Secondly, we jointly cope with new and handoff users efficiently. A portion of resources are reserved for handoff users because ongoing call dropping is more intolerable than call blocking for users. Thirdly, in order to impose the universality of MQACM, different priority levels are assigned to multiple service types when network congestion happens. Last but not least, multi-agent Q-learning (MQ) is creatively adopted in this paper to solve the admission problem of heterogeneous networks. It ensures MQACM effective in any network situation.

The rest of this paper is organized as follows. Section 2 investigates the system states analysis of the heterogeneous network of WCDMA/WLAN, in which the states of the networks will be described using load parameters. The mathematical design of MQACM is discussed in section 3 and then the convergence proof of MQ is given. Section 4 provides the simulation results. And conclusions are drawn in Section 5.

2. Network States Analysis

The admission control algorithm in this paper is designed from the perspective of increasing the network capacity and network utilization, and therefore the network load characteristic is used as the network states. Once such a load measurement is found, we could make decisions that whether or not admit the users with different service requirements based on the load situation. This section gives the network load parameters and explains the parameter choosing reason of WCDMA subnet and WLAN subnet separately.

2.1 WCDMA System Load Parameter Analysis

For WCDMA subnet, the state measurement is based on the number of users in many literatures [20]. This method is easy to operate but not as accurate as interference power based methods. Different users require different services, bandwidth and have different distance from the base station (BS). They have different influence on the load of the networks. Hence, interference power is used as the description of WCDMA subnet state, which is the input parameters for MQACM in section 3.

At BS, suppose the number of users is N . P_i is transmission signal power received from user i . I_{total} is the total received power including the background noise power P_n . I_{total} can be written as

$$I_{total} = P_n + \sum_{i=1}^{N} P_i \,. \tag{1}$$

For WCDMA system, the relation between the signal to interference plus noise ratio (SINR) [21] and P_i can be expressed as

$$\begin{aligned}(E_b / N_0)_i &= \left[W/(\mathrm{v}_i R_i)\right]\cdot\left[P_i/(I_{total}-P_i)\right] \\ \Rightarrow \quad P_i &= \{1/[1+W/(E_b/N_0)_i R_i v_i]\}\cdot I_{total}\ , \\ &= L_i I_{total}\end{aligned} \tag{2}$$

where W is the chip rate, and v_i is the activity factor. R_i is the data rate of user i, and L_i can be defined as the load factor of one link,

$$L_i = 1/[1+W/(E_b/N_0)_i R_i v_i]. \tag{3}$$

Then the sum of load factor of each link and the interference from other cells, total load factor η_{UL}, is expressed as

$$\begin{aligned}\eta_{UL} &= (1+f)\sum_{i=1}^{k} L_i \\ &= (1+f)\sum_{i=1}^{k} 1/[1+W/(E_b/N_0)_i R_i v_i]\end{aligned}\ , \tag{4}$$

where f is the ratio of the inter-cell interference power to the intra-cell interference power. According to the definition of I_{total} and P_n from(1) and the equation (2),(3), (4), it can be easily obtained that

$$\begin{aligned}&I_{total}/P_n = (1+f)/(f-\eta_{UL}) \\ &\Rightarrow I_{total} = (1+f)P_n/(f-\eta_{UL}) \\ &\Rightarrow dI_{total}/d\eta_{UL} = (1+f)P_n/(f-\eta_{UL})^2\end{aligned}\ . \tag{5}$$

The increment of interference power for uplink can be expressed using integration, and integral interval is from the old value of total load factor $\eta_{ULold} = \eta_{UL}$ to the new value $\eta_{ULnew} = \eta_{UL} + \Delta L$,

$$\begin{aligned}\Delta I &= \int_{\eta_{UL}}^{\eta_{UL}+\Delta L} dI_t = \int_{\eta_{UL}}^{\eta_{UL}+\Delta L} (1+f)\frac{P_n}{(f-\eta_{UL})^2} d\eta_{UL} \\ \Rightarrow \Delta I &= (1+f)P_n\left(\frac{1}{f-\eta_{UL}-\Delta L} - \frac{1}{f-\eta_{UL}}\right) \\ &= (1+f)P_n/(f-\eta_{UL})\cdot\frac{\Delta L}{f-\eta_{UL}-\Delta L} \\ &= \frac{I_{total}}{f-\eta_{UL}-\Delta L}\cdot\Delta L\end{aligned}\ , \tag{6}$$

where ΔL is the load factor of the new link, given as

$$\Delta L = \left(1+\frac{W}{vR(E_b/N_0)}\right)^{-1}, \tag{7}$$

where R is the data rate of the new user, and (E_b/N_0) is SINR of the new link. Thus, we can compute the increment of interference power when the load of the network changes.

The WCDMA subnet is divided into four states according to interference power, which are $I_{very\ low}$, I_{low}, I_{high}, $I_{very\ high}$. Accordingly, three interference thresholds I_1, I_2 and I_3 are set up. When the total interference power plus new interference power is $0 \le I + \Delta I < I_1$, the interference power belongs to $I_{verylow}$; when $I_1 \le I + \Delta I < I_2$, the interference power belongs

Copyright ⓒ 2013 KSII

to I_{low}; when $I_2 \le I + \Delta I < I_3$, the interference power belongs to I_{high}; when $I + \Delta I \ge I_3$, the interference power belongs to $I_{veryhigh}$.

2.2 WLAN System Load Parameter Analysis

Just as interference power is used to describe the load situation for WCDMA subnet, channel busyness ratio [22] is choosed for WLAN subnet. The channel busyness ratio is easy to be measured and it corresponds to the most important load performance index, i.e., throughput.

Suppose p is the probability that there is at least one transmission among the neighbors in the observed back-off time slot, which is given by

$$p = 1 - (1 - p_t)^{n-1}, \tag{8}$$

where n is the number of users, and p_t is the transmission probability for each node in any back-off time. Suppose $p_s = np_t(1-p_t)^{n-1}$ and $p_i = (1-p_t)^n$ are the probability that the data is successfully transmitted and the probability that the observed back-off time slot is idle, respectively. And $p_c = 1 - p_i - p_s$ is the collision probability that there are at least two concurrent transmissions at the same back-off time slot. Obviously, all these three probabilities are functions of p.

Let T_s be the average time period associated with one successful transmission and T_c be the average time period associated with collisions. When the request-to-send/clear-to-send (RTS/CTS) mechanism is used, T_s and T_c can be expressed as

$$\begin{aligned} T_s &= T_{RTS} + T_{CTS} + T_{\overline{DATA}} + T_{ACK} + 3T_{SIFS} + T_{DIFS} \\ T_c &= T_{RTS} + T_{CTS} + T_{SIFS} + T_{DIFS} \end{aligned}, \tag{9}$$

where T_{RTS}, T_{CTS}, $T_{\overline{DATA}}$, T_{ACK}, T_{SIFS}, and T_{DIFS} represent the average time of RTS message, CTS message, data-transmission period, acknowledgement message, short inter-frame space, and distributed inter-frame space, respectively. Thus, it can be easily obtained that

$$\begin{cases} R_b = 1 - p_i\sigma / (p_i\sigma + p_sT_s + p_cT_c) \\ R_s = p_sT_s / (p_i\sigma + p_sT_s + p_cT_c) \end{cases}, \tag{10}$$

where σ is the length of an empty back off time slot, and R_s is the channel utilization ratio, and R_b is the channel busyness ratio. Once R_s is obtained, the normalized throughput th can be expressed as

$$th = (R_s \times T_{\overline{DATA}}) / T_s. \tag{11}$$

Note that the normalized throughput is proportional to R_s, and the channel busyness ratio is injective. In fact, when $p \le 0.1$, R_b is almost the same as R_s. Thus, it can monitor the normalized throughput by simply measuring the channel busyness ratio, which can be easily done since IEEE 802.11 is a CSMA-based MAC protocol working on physical and virtual carrier sensing mechanisms [23].

Next, how to use R_b to describe the state of WLAN is investigated. The normalized throughput is also a function in terms of p. To obtain the maximum normalized throughput, we take the derivative of $th(p)$ with respect to p and let it equal 0, i.e.,

$$\frac{\mathrm{d}}{\mathrm{d}p} th(p) = \frac{\mathrm{d}}{\mathrm{d}p} R_S(p) = 0. \tag{12}$$

Moreover, for the specific number of user n, p has upper bound $\max(p)$. Suppose that

p_r is the root of (12). Let $p*$ denote the optimal value of collision probability

$$p* = \min\{p_r, \max(p)\}. \quad (13)$$

Then, the maximum throughput and the maximum R_s can be achieved. Finally, the maximum value $R_{\max} = R_b(p*)$ of R_b can be found.

The WLAN subnet is also divided into four states according to channel busyness ratio, which is $R_{bvery\ low}$, $R_{b\ low}$, R_{bhigh}, and $R_{bvery\ high}$. Accordingly, three channel busyness ratio thresholds R_{b1}, R_{b2}, and R_{b3} are set up. When the channel busyness ratio is $0 \le R_b < R_{b1}$, the channel busyness ratio belongs to $R_{bverylow}$; when $R_{b1} \le R_b < R_{b2}$, the channel busyness ratio belongs to R_{blow}; when $R_{b2} \le R_b < R_{b3}$, the channel busyness ratio belongs to R_{high}; when $R_{b3} \le R_b \le R_{\max}$, the channel busyness ratio belongs to $R_{veryhigh}$.

3. Design of MQACM

This section describes the MQACM system framework and then explores MQ algorithm for threshold estimation block and network selection block. MQ algorithm needs some special definitions for the system state, actions and payoffs. The system states have been discussed in section 2. This section focuses on the design of actions and payoffs and then seeks detailed MQ steps for the two blocks.

3.1 MQACM System Description

Fig. 1 shows the framework of MQ based MQACM system. Two Q-learning blocks are designed here, network states threshold estimation block and network selection block. Each block owns an agent that executes different optional actions so as to learn the optimal policy. Network states discussed in section 2 are the input parameters for the two blocks. Immediate payoff is the feedback that indicates the actions of the two agents correct or not. The two agents share the information of network states and immediate payoff.

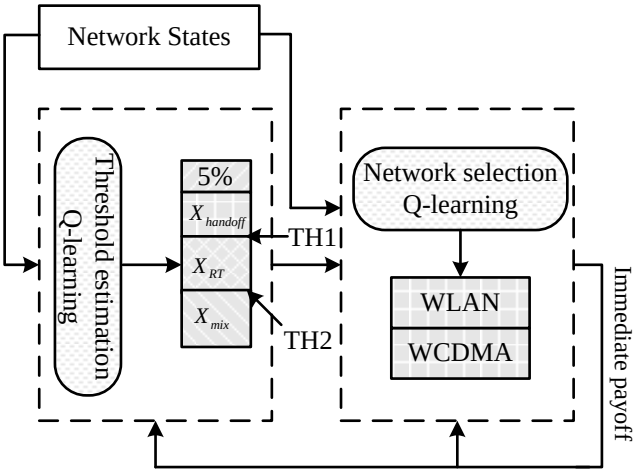


Fig. 1. Multi-agent Q-learning based MSACM framework.

In heterogeneous networks, there are more complicated service types that often have different QoS requirements. For simplicity, two service types are considered here: Real Time

Copyright ⓒ 2013 KSII

(RT) service type and Not Real Time (NRT) service type. NRT is more tolerant to delay, so we have reasons to believe that RT is more important than NRT. The threshold estimation block reserves 5% of the total network resources to avoid extortionate system saturation. By considering that handoff dropping is even more unbearable than call blocking for users, $X_{handoff}$ percent resources are reserved for handoff services. The rest resources are divided in to two parts: the first part X_{mix} percent is shared by both RT and NRT service; the second part X_{RT} percent is kept for RT service to allocate higher priority to RT service. As shown in **Fig. 1**, the threshold between X_{RT} and $X_{handoff}$ is TH_1. The threshold between X_{mix} and X_{RT} is TH_2. The agent in this block tries to learn the optimal dynamic distribution of TH_1 and TH_2.

Coming out of the threshold estimation block, TH_1 and TH_2 are two important constraints to the next network selection block. The agent in network selection block aims to optimize the decisions of connecting users to WCDMA subnet or WLAN subnet. Suppose "*resource*" is the percentage of the total resources that has already been occupied by the connected users. And the admission control mechanism could be:

If $0 \le resource < X_{mix}$, accept all RT/NRT call requests;

If $X_{mix} \le resource < X_{mix} + X_{RT}$, accept RT call requests, and reject NRT call requests;

If $X_{mix} + X_{RT} \le resource < X_{mix} + X_{RT} + X_{handoff}$, reject all new RT/NRT call requests, and accept handoff requests;

If $X_{mix} + X_{RT} + X_{handoff} \le resource < 1 - 5\%$, reject all call requests.

3.2 Single-agent Q-learning Method

Q-learning is a self-learning method to optimize the kind of decisions that depend both on the current and the history state-action-pair. Agent in Q-learning system learns how to improve its decisions during the learning process according to its experience [24]. In a decision epoch t, the agent observes the network state e and implements an action a. When it arrives at the next epoch $t+1$, the network environment indicates the action correct or not by giving the agent an immediate payoff $p_t(e,a)$, and then switches to a new state. In a single-agent Q-learning system, the agent updates a Q-table $Q_{t+1}(e,a)$ with the immediate payoff and a previous Q-value,

$$Q_{t+1}(e,a) = (1-\alpha)Q_t(e,a) + \alpha\{p_{t+1}(e,a) + \beta V_t(e)\}. \tag{14}$$

Formula (14) shows the classic single-agent Q-learning method. Where $0 \le \alpha < 1$ is the learning rate parameter. It is an important factor for the performance of the algorithm and need to be set reasonably. $0 \le \beta \le 1$ is the system discount factor. The value function $V_t(e)$ in (14) is defined as

$$V_t(e) = \max_b \{Q_t(e,b)\}, \tag{15}$$

where b is the optional action in epoch t. The more correct the action a is, the larger $p_t(e,a)$ is. The agent finds the optimal policy $\pi^*(e) \in A$ by maximizing the total expected discounted payoffs, which can be denoted as

$$P^{\pi}(e) = \max_{\pi} \mathrm{E}\left\{\sum_{t=0}^{\infty} \beta^t p_t(e,a)\right\}, \tag{16}$$

where E is an operator that stands for expectation.

3.3 Multi-agent Q-learning Method

In this paper, MQACM obtains the optimal decisions of threshold estimation block and network selection block through MQ method. Suppose i indicates i-th agent and i=1,2,…n. Here n=2, that is to say, the system has two agents, one threshold estimation agent and one network selection agent. Thus, the single-agent Q-learning can be extended to MQ, in which the two agents share the same immediate payoff $p_t^i(e_i,a_i)$ and $T_{e_i e_i'}^i$ information. Similarly, the system updates MQ-value table as follow,

$$\begin{aligned} MQ_{t+1}^i(e_i,a_i) &= (1-\alpha_{t+1}^i)MQ_t^i(e_i,a_i) \\ &+ \alpha_{t+1}^i\left\{p_{t+1}^i(e_i,a_i)+\beta\sum_{e_i'}T_{e_i e_i'}^i[a_i]V_t^i(e_i')\right\}, \end{aligned} \tag{17}$$

where $MQ_t^i(e_i,a_i)$ is MQ-value; $T_{e_i e_i'}^i$ is the state transition probability between the current state e_i and the next state e_i'. The MQ-value $MQ_{t+1}^i(e_i,a_i)$ holds previous MQ-value $MQ_t^i(e_i,a_i)$ with probability $(1-\alpha_{t+1}^i)$, and takes two parts with probability α_{t+1}^i: an immediate payoff $p_{t+1}^i(e_i,a_i)$, and a discounted long-term value $V_t^i(e_i')$ for next state. $V_t^i(e_i')$ is controlled by an equilibrium strategy [25],

$$V_t^i(e_i') \equiv EQ_t^i[\prod_{j=1}^{n} a_j \times \max_b MQ_t^i(e_i',b)]. \tag{18}$$

The equilibrium functions EQ_t^i evaluate utilities for possible sets of next actions, and choose optimal actions which bring maximum MQ-values.

Similar to the single-agent Q-learning algorithm, the total expected discounted payoff in MQ can be denoted as

$$P^{\pi}(e) = \max_{\pi} \mathrm{E}\left\{\sum_{i=1}^{n}\sum_{t=0}^{\infty}\beta p_t^i(e_i,a_i)\right\}. \tag{19}$$

3.4 Action Definition

Action definition for threshold estimation agent and network selection agent are addressed separately in this section. In threshold estimation block, the agent changes TH_1 and TH_2 according to the system states and QoS feedback, so the definition of the action is increasing TH_1 / TH_2 by θ, keeping TH_1 / TH_2 unchanged or decreasing TH_1 / TH_2 by θ. The threshold action set A_1 is

$$\begin{aligned} A_1 = \{&[(TH_1-\theta),(TH_2-\theta)],[(TH_1-\theta),(TH_2+\theta)],[(TH_1-\theta),(TH_2+0)], \\ &[(TH_1+0),(TH_2-\theta)],[(TH_1+0),(TH_2+\theta)],[(TH_1+0),(TH_2+0)], \\ &[(TH_1+\theta),(TH_2-\theta)],[(TH_1+\theta),(TH_2+\theta)],[(TH_1+\theta),(TH_2+0)]\} \end{aligned} \tag{20}$$

where θ is adjustment quantity for TH_1 and TH_2. It should be set reasonably to optimize the threshold estimation action speedily.

In network selection block, if a new user triggers a call in the double-coverage area or an ongoing user tries to handover from single-coverage area to double-coverage area, the agent could connect the user to WCDMA subnet or WLAN subnet. On the contrary, if the new/handoff call happens in the single-coverage area, it only can be connected into WCDMA subnet. If there are not enough resources in both of the two subnets, the call will be rejected.

Copyright ⓒ 2013 KSII

The chosen subnet is expected to make MQACM system receive the best payoffs. So the decision action that which subnet the user can be connected into is defined as a , which can be written as

$$a=\begin{cases}1 & \text{connect into WCDMA}\\ 2 & \text{connect into WLAN}\\ 3 & \text{keep in original subnet}\\ 4 & \text{reject}\end{cases}. \tag{21}$$

According to the action definition, the network selection action set A_2 can be defined as

$$\begin{aligned}A_2=\{[&a_{s\_n\_RT},a_{s\_n\_NRT},a_{s\_h\_RT},a_{s\_h\_NRT},a_{d\_n\_RT},\\ &a_{d\_n\_NRT},a_{d\_h\_RT},a_{d\_h\_NRT}],\ a_{s\_n\_RT/NRT}\in\{1,4\},\\ &a_{s\_h\_RT/NRT}\in\{1,3,4\},a_{d\_n\_RT/NRT}\in\{1,2,4\},\\ &a_{d\_n\_RT/NRT}\in\{2,3,4\}\}\end{aligned}, \tag{22}$$

where $a_{s\_n\_RT/NRT}$ / $a_{s\_h\_RT/NRT}$ is the action for new/handoff user in single-coverage area; $a_{d\_n\_RT/NRT}$ / $a_{d\_n\_RT/NRT}$ is the action for new/handoff users in double-coverage area.

3.5 Immediate Payoff Definition

As mentioned in section 3.3, the two agents in MQACM share the same immediate payoff $p_t^i(e_i,a_i)$ information, so we obtain $p_t^1(e_1,a_1)=p_t^2(e_2,a_2)=p_t(e,a)$. The immediate payoff $p_t(e,a)$ with two parts is designed here. The first part is the traditional form that often appears in the classic Q-learning algorithm, that is

$$p_t(e,a)\_access=\begin{cases}2 & n_{s\_RT}+1/\,n_{d\_NRT}+1\\ 1 & n_{s\_NRT}+1/\,n_{d\_RT}+1\\ 2 & n_{d\_RT}-1,\,n_{s\_RT}+1/\,n_{s\_NRT}-1,\,n_{d\_NRT}+1\\ 1 & n_{d\_NRT}-1,\,n_{s\_NRT}+1/\,n_{s\_RT}-1,\,n_{d\_RT}+1\\ -1 & \text{new user is rejected/ handoff user drops}\\ 0 & \text{otherwise}\end{cases}, \tag{23}$$

where n_{s_RT} and n_{d_RT} is the number of RT users in single-coverage area and double-coverage area respectively; n_{s_NRT} and n_{d_NRT} is the number of NRT users in single-coverage area and double-coverage respectively. RT users prefer to be connected to WCDMA network whereas delay-tolerant NRT users prefer WLAN network. If a new/handoff RT user is connected into WCDMA subnet, the immediate payoff is bigger; if a new/handoff NRT user is connected into WLAN subnet, the immediate payoff is bigger; if the handoff call drops or there are not enough resources for new user in the both of the two subnets, the new call will be blocked and $p(e,a)\_access=-1$.

Considering that the resulting performance of the implemented action should be not only related to the network-perceived service quality but also related to the QoS satisfaction degree of users, we propose an improved payoff function as the second part of the immediate payoff, which is defined as

$$p_t(e,a)\_QoS = -\left\{\left[\frac{\left|R^* - R_t(e,a)\right|}{R^*}\right]^2 + \left[\frac{\left|D_t(e,a) - D^*\right|}{D^*}\right]^2 + \left[\frac{\left|E_t(e,a) - E^*\right|}{E^*}\right]^2\right\}, \quad (24)$$

where $R_t(e,a)$, $D_t(e,a)$ and $E_t(e,a)$ are the data rate, transmission delay and BER of users measured after the action is implemented at state e_t. R^*, D^* and E^* are the expected data rate, transmission delay and BER of users. Thus, the total immediate payoff function can be improved, and the total immediate payoff can be written as

$$p_t(e,a)=p_t(e,a)\_access+p_t(e,a)\_QoS. \quad (25)$$

The bigger $p_t(e,a)$ is, the better effect the action has. The design of the immediate payoff function considers the benefit of both the network and users, so it can keep balance between network profit and users QoS requirements.

3.6 Multi-agent Q-learning Steps for MQACM

In this section, the detailed steps of MQ for MQACM are addressed and then MQ proposed in this paper is proved convergent with probability 1. First, define row vector $e_c = [I_{very\ low}, I_{low}, I_{high}, I_{very\ high}]$. Where $I_{very\ low}$, I_{low}, I_{high}, and $I_{very\ high}$ denote the four states in WCDMA subnet respectively, which have been discussed in section 2. Define row vector $e_w = [R_{bvery\ low}, R_{b\ low}, R_{b\ high}, R_{bvery\ high}]$, where $R_{bvery\ low}$, $R_{b\ low}$, R_{bhigh} and $R_{bvery\ high}$ denote the four states in WLAN subnet, respectively. The state vector of the system is given by

$$\begin{aligned} E = [&R_{bvery\ low}I_{very\ low}, R_{bvery\ low}I_{low}, R_{bvery\ low}I_{high}, R_{bvery\ low}I_{very\ high}, R_{b\ low}I_{very\ low}, \\ &R_{b\ low}I_{low}, R_{b\ low}I_{high}, R_{b\ low}I_{very\ high}, R_{b\ high}I_{very\ low}, R_{b\ high}I_{low}, R_{b\ high}I_{high}, \\ &R_{b\ high}I_{very\ high}, R_{bvery\ high}I_{very\ low}, R_{bvery\ high}I_{low}, R_{bvery\ high}I_{high}, R_{bvery\ high}I_{very\ high}] \end{aligned} \quad (26)$$

Action set A_1 and A_2 discussed in section 3.4 are available at each state e_t. The two agents implement an optional action respectively and the environment gives an immediate payoff $p^i_{t+1}(e_i,a_i)$. Depending on the payoff and corresponding action, the system updates the MQ-value according to (17). MQ steps are explained in detail as below:

1) Initialize MQ-value table for each agent.

2) At each user arrival, according to the network states:

a. Each agent selects an optional action in their action sets and memorizes it,

b. Memorize the current network state e^i_t at the arrival time and the next network state e^i_{t+1} after the two agents take their actions.

3) When the state of the heterogeneous network changes, calculate the immediate payoff $p^i_{t+1}(e_i,a_i)$.

4) Update the MQ-value matrix $MQ^i_{t+1}(e_i,a_i)$ according to equation (17).

5) If $\Delta MQ_{t+1}(e_i,a_i) < \varepsilon, \forall e_i \in E, a_i \in A_1 / A_2$, convergence has occurred and stop learning. Otherwise continue learning by repeating steps 2-4.

Now we end our design with the proof of the convergence for MQ. MQ is an extended version of single-agent Q-learning method. For every network state-action pair (e_i,a_i), the value of $\Delta MQ_{t+1}(e_i,a_i)$ is the change of MQ-value between before and after the action is implemented at every iteration. To examine the convergence of MQ method in MQACM, let $\max_{e\in E, a_i\in A_i} \Delta MQ_{t+1}(e_i,a_i)$ be the convergence performance index. We prove that this

Copyright ⓒ 2013 KSII

convergence performance index is bounded with a small value as follow:

$$\begin{aligned}
&\max_{e\in E, a_i\in A_i} \Delta MQ_{t+1}(e_i,a_i) \\
&= \max_{e\in E, a_i\in A_i, t\in(0,\infty)} \left| MQ_{t+1}(e_i,a_i) - MQ_t(e_i,a_i) \right| \\
&\le \max_{e\in E, a_i\in A_i, t\in(0,\infty)} \left\{ \left| p^i_{t+1}(e_i,a_i) - p^i_t(e_i,a_i) \right| + \beta^i \left| \sum_{e_i'} T^i_{e_i e_i'}[a_i] V^i_{t+1}(e_i') - \sum_{e_i'} T^i_{e_i e_i'}[a_i] V^i_t(e_i') \right| \right\}. \quad (27) \\
&= \max_{e\in E, a_i\in A_i, t\in(0,\infty)} \left\{ \left| p^i_{t+1}(e_i,a_i) - p^i_t(e_i,a_i) \right| + \beta^i \left[\sum_{e_i'} T^i_{e_i e_i'}[a_i] \right] \cdot \left| V^i_{t+1}(e_i') - V^i_t(e_i') \right| \right\} \\
&\le \max_{e\in E, a_i\in A_i, t\in(0,\infty)} \left\{ \left| p^i_{t+1}(e_i,a_i) - p^i_t(e_i,a_i) \right| + \beta^i \left[\sum_{e_i'} T^i_{e_i e_i'}[a_i] \right] \cdot \max_{i\in(1,n)} \left| V_{t+1}(e') - V_t(e') \right| \right\}
\end{aligned}$$

From [25], we know a conclusion that the value of immediate payoff $p^i_t(e_i,a_i)$ is bounded with η^i, which is obvious in our immediate payoff definition. Where η^i is a defined probability,

$$\left| p^i_{t+1}(e_i,a_i) - p^i_t(e_i,a_i) \right| < \eta^i . \quad (28)$$

By [26], we know a theorem in single-agent Q-learning: given bounded immediate payoffs $\left| p_t(e,a) \right| \le \Re$, learning rate parameter $0 \le \alpha < 1$ and

$$\sum_{j=1}^{\infty} \alpha^j(e,a) = \infty, \sum_{j=1}^{\infty} (\alpha^j(e,a))^2 < \infty, \forall e, a . \quad (29)$$

Then $Q_t(e,a) \to Q^*(e,a)$ as $n \to \infty$, $\forall e, a$, with probability 1. Where, $Q_t{}^*(e,a)$ is optimal value of $Q_t(e,a)$, and $\Re$ is the largest $p_t(e,a)$. And it is simple to prove that all the convergence conditions above are satisfied in the last part of (27). As $n \to \infty$,

$$\left.\begin{aligned}
&V_{t+1}(e) = \max_b \{Q_{t+1}(e,b)\} \to Q^*(e,b) \\
&V_t(e) = \max_b \{Q_t(e,b)\} \to Q^*(e,b) \\
&\Rightarrow \left| V_{t+1}(e) - V_t(e) \right| \to 0
\end{aligned}\right\} . \quad (30)$$

Using (27), (28) and (30),

$$\max_{e\in E, a_i\in A_i} \Delta MQ_{t+1}(e_i,a_i) \le \max_{e\in E, a_i\in A_i, t\in(0,\infty)} \left| \eta^i \right| = \eta^i . \quad (31)$$

Thus we prove that $\max_{e\in E, a_i\in A_i} \Delta MQ_{t+1}(e_i,a_i)$ is bounded with η^i. Now (30) and (31) show that MQ-value converges to a fixed value with probability of 1, where epoch t increases to infinity.

4. Simulation and Discussion

4.1 Simulation Environment

Our simulation is conducted with the following settings in the integrated WCDMA/WLAN system. For WCDMA subnet, the coverage radius of the BS is 1 km. The channel suffers AWGN noise, log-normal shadowing, and multipath fading. Perfect power control is used in the system. For WLAN subnet, the coverage radius of the AP is 100 meters. IEEE 802.11b is used and the average bit rate is assumed to be 11Mbps. Rayleigh channel model is considered.

The other parameters are given in **Table 1**.

Table 1. Network parameters used in the simulation

Parameter	Notion	Value
Path-loss exponent	L	4.35
Spread spectrum factor	F	4-256
Time required to transmit a request-to-send (RTS)	T_{RTS}	15μs
Time required to transmit a clear-to-send (CTS)	T_{CTS}	10μs
Time required to transmit a ACK	T_{ACK}	10μs
Short inter-frame space (SIFS)	T_{SIFS}	10μs
Distributed inter-frame space (DIFS)	T_{DIFS}	50μs
Learning parameter	α	0.1
Discount factor	γ	0.95

In simulation, the arrival rate of new (handoff) calls in WCDMA/WLAN obeys the Poisson process [27] and the mean arrival rate is λ_n (λ_h), and $\lambda = \lambda_n + \lambda_h$. A call request could be a RT call or a NRT call with the possibility of 60% and 40%, respectively. For simplicity, the simulation uses voice call and data call to represent RT and NRT service, respectively. Users are assumed to be uniformly distributed in WCDMA/WLAN and the mobility model is random-walk model [28]. All access transmissions of users are always on. The system QoS requirements are listed in **Table 2**.

Table 2. QoS requirement for different service types

Traffic Type	Transmission Delay	Data Rate	BER
Voice call	<150ms	32kbps	10^{-3}
Data call	<1000ms	128kbps	10^{-6}

4.2 Simulation Results

The simulation evaluates the relative performance of MQACM proposed in this paper by comparing it with an existing model based MDP Algorithm in literature [13] and a single-agent Q-learning algorithm (FQAC) in literature [14].

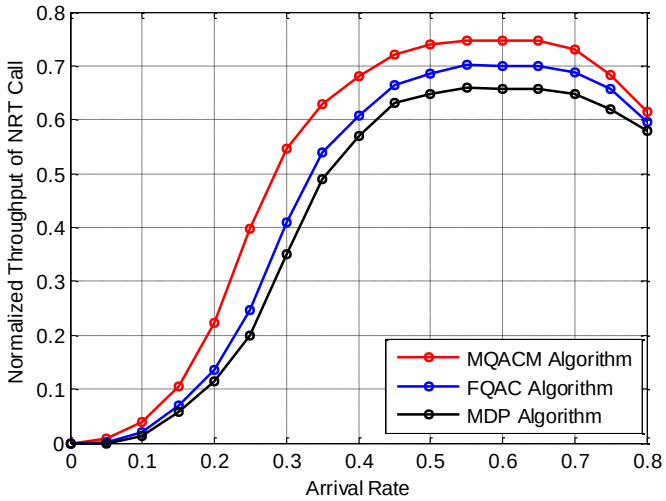


Fig. 2. Normalized throughput of new NRT call under different arrival rate.

Fig. 2 demonstrates the normalized throughput of new NRT call at the corresponding heterogeneous network arrival rate (λ) on the X axis. The definition of the normalized

Copyright ⓒ 2013 KSII

throughput is shown in equation (11). From the figure, we can observe that the normalized throughput of all the three algorithms increases to their maximum value around $\lambda = 0.6$, and then reduces as λ increases until the network becomes saturated. Actually, it is not hard to understand. The largest throughput should not be achieved when the network arrives at saturation condition. NRT services are firstly considered to be attached to WLAN subnet. When λ is extremely high, the collision probability in WLAN becomes greater, so the throughput of NRT call reduces. It also can be seen from **Fig. 2** that MQACM always performs better than the other two algorithms. The reasons are that MQACM adopts MQ which can adapt to system states dynamics with its self-learning capability and can always appropriately make decisions to admit or reject the new (handoff) call by the intelligent method.

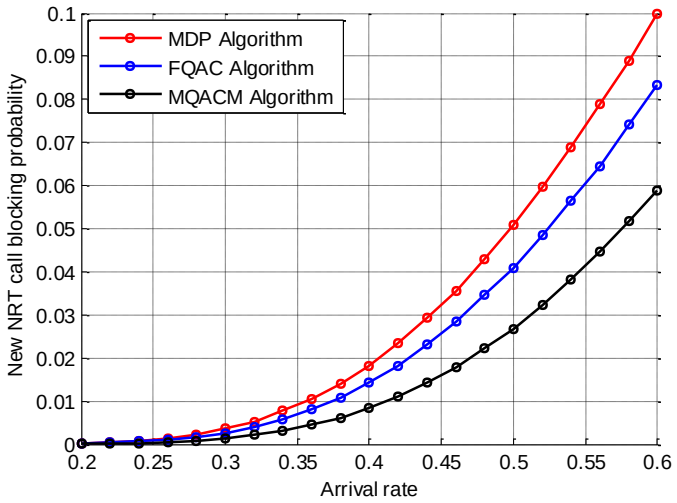


Fig. 3. New NRT call blocking probability under different arrival rate

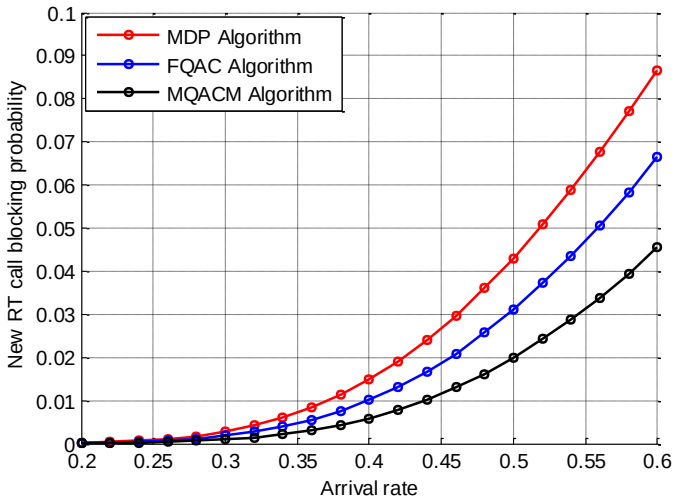


Fig. 4. New RT call blocking probability under different arrival rate.

Fig. 3 and **Fig. 4** show the new NRT and RT call blocking probability under different arrival rate respectively. The blocking probability using the MQACM is much lower than that using the other two algorithms for both NRT and RT call. It is because MQACM uses load parameters to describe the network states, and receives accurate feedback of users QoS

immediately after the action is executed. That guarantees tradeoff between the network capacity and the QoS requirements of users. Additionally, we distribute higher priority to RT call in threshold estimation block. That is why the blocking probability of RT call is lower than that of NRT call in **Fig. 3** and **Fig. 4**.

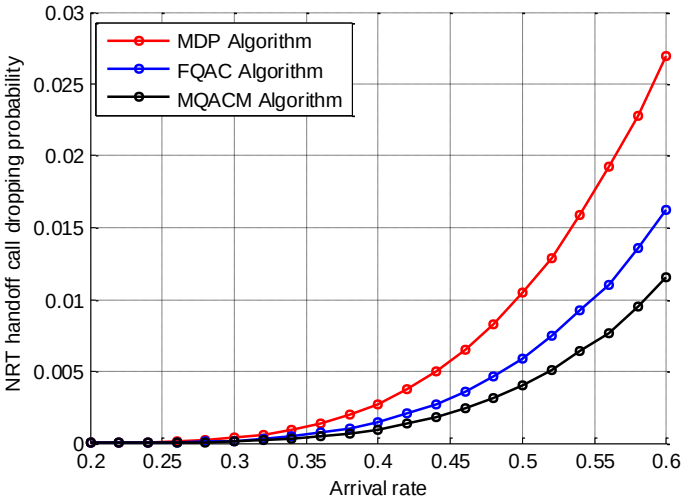


Fig. 5. NRT handoff call dropping probability under different arrival rate.

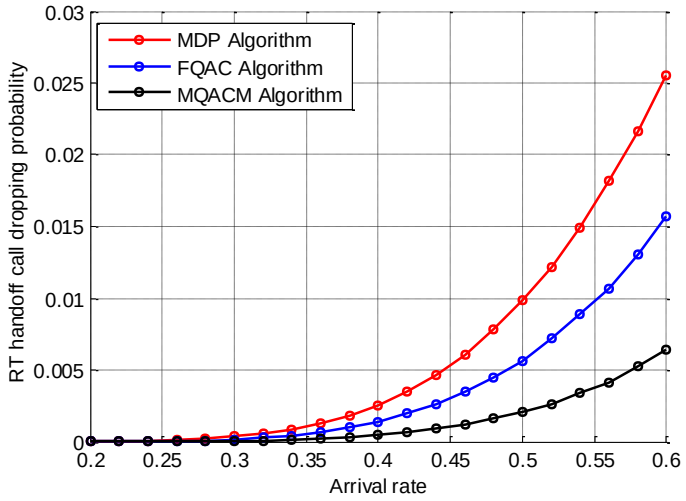


Fig. 6. RT handoff call dropping probability under different arrival rate.

Fig. 5 and **Fig. 6** illustrate how RT and NRT handoff call dropping probabilities behaves under different arrival rate. In MDP, the dimension and the computation complexity will increase dramatically with the increase of states. And FQAC does not reserve resources for more important handoff service requests. As a result, MQACM obtains better performance of handoff call dropping probability for both RT and NRT services than MDP and FQAC. Moreover, RT and NRT handoff call dropping probabilities are almost the same in MDP and FQAC without allocating higher priority for RT service.

In **Fig. 7**, the handoff numbers increase with the growth of the arrival rate. We can see that MQACM cuts down the meaningless handoff numbers and avoids ping-pong effect substantially. It is due to the fact that long time learning process in our MQACM makes the agent of the network selection system know what handoff is unnecessary. For example, if a

Copyright ⓒ 2013 KSII

user with high speed in single-coverage area moves into double-coverage area, the agent may not change his access subnet by considering that this user would move out of double-coverage area and go back to the single-coverage area quickly. Furthermore, the handoff number performance is more stable in the two Q-learning algorithms (FQAC and MQACM). It is because Q-learning is an online method and is able to optimize the state transition probability until it approximates to the actual network state transition. On the contrary, MDP defines the transition probability offline according to certain rules, so it absolutely will not be that accurate. The reason why MQACM is more stable than FQAC is that the threshold block improves the convergence rate of MQ.

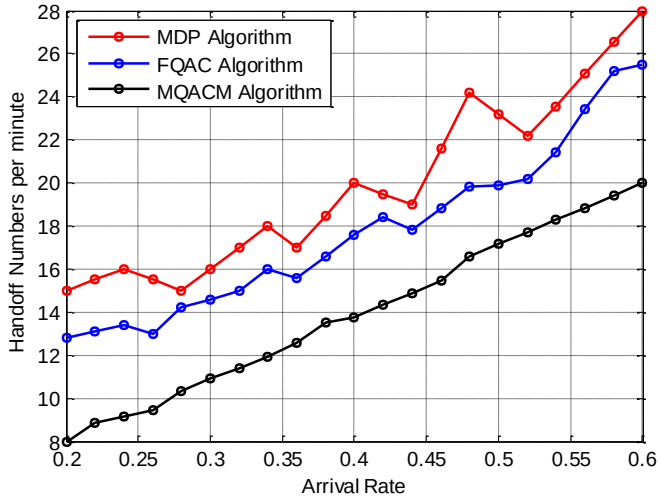


Fig. 7. Handoff numbers per minute under different arrival rate.

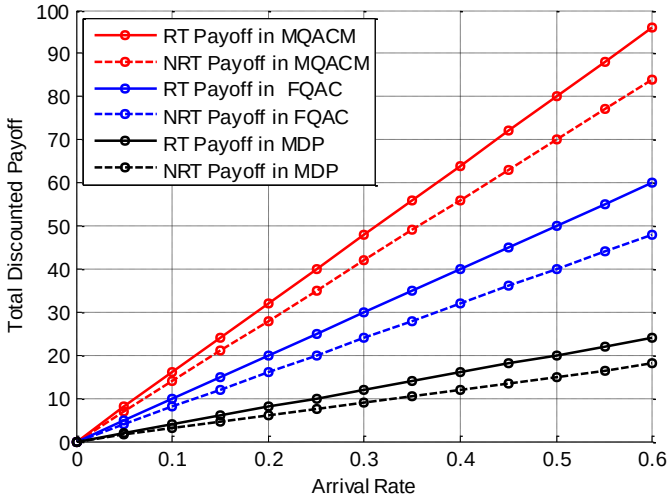


Fig. 8. The total discounted payoff under different arrival rate.

Fig. 8 shows the total discounted payoff of different service types, which is defined by (21). It can be obviously found that our MQACM receives a much better payoff performance than the other two methods for both RT and NRT call. The total payoff of RT and NRT call in MQACM is about 4 and 2 times than that in MDP and FQAC method respectively. That is owing to the self-learning capacity of Q-learning that could adapt to the system variation. Therefore, systems with Q-learning algorithm (FQAC and MQACM) are able to

accommodate more users and make better admission decisions while maintaining QoS guarantee. And they obtain much payoff than MDP. Besides, MQACM uses MQ for both network selection and threshold estimation block, which enhances the network selection process even no sufficient resources left in the system. It means MQACM can handle the network jam situation better than FQAC which uses single-agent Q-learning.

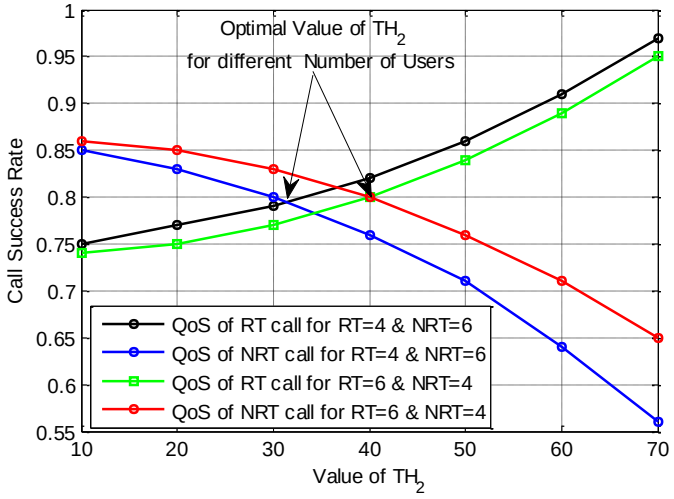


Fig. 9. Call success rate under different value of TH_2.

Fig. 9 vividly describes the call success rate under different value of TH_2 for different number of RT and NRT users. From the figure we can see that call success rate of RT call increases and call success rate of NRT call reduces with the increase of TH_2. The reason is that the bigger the value of TH_2 becomes, the more resources is reserved for RT users. And that is also why the effect of TH_2 for the call success rate performance is more obvious in the situation that the number of RT users is smaller (RT=4). The intersection of the black and the blue curves, and the intersection of the green and the red curves are the optimal values of TH_2 for different number of users, which guarantee the call success rate for both RT and NRT call.

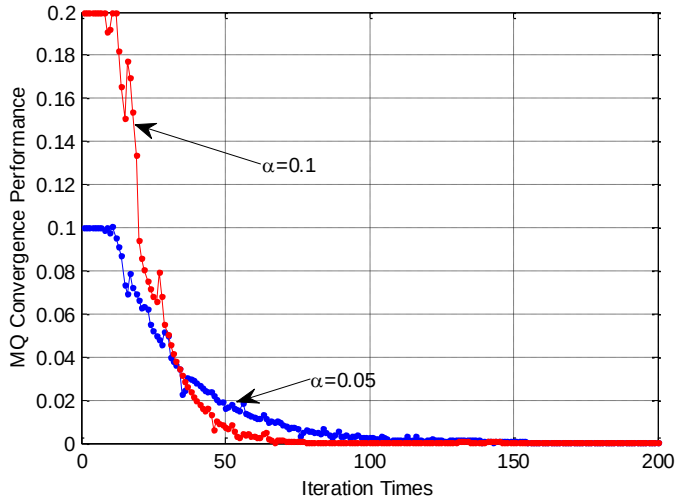


Fig. 10. MQ convergence performance under different iteration times.

Copyright ⓒ 2013 KSII

Fig. 10 shows that $\max_{e\in E, a_i\in A_i} \Delta MQ_{t+1}(e_i, a_i)$ reduces quickly with the increase of iteration times. In other words, simulations illustrate that MQ method can receive good convergence performance. Thus, heterogeneous network system can be always in a stable state after MQ converges. $\max_{e\in E, a_i\in A_i} \Delta MQ_{t+1}(e_i, a_i)$ stays unchanged at the beginning of the iteration because the initial value of MQ-value is set to be 0, while it has no performance effect in the algorithm design. It also shows the impact of learning parameter α^i on $\max_{e\in E, a_i\in A_i} \Delta MQ_{t+1}(e_i, a_i)$ .The convergence rate is more quick when $\alpha^i = 0.1$, i.e., the greater the α^i is, the stronger the learning ability of MQ is. However, the initial value of $\max_{e\in E, a_i\in A_i} \Delta MQ_{t+1}(e_i, a_i)$ when $\alpha^i = 0.1$ is two times as big as that when $\alpha^i = 0.05$. When α^i becomes greater, the change span of $\max_{e\in E, a_i\in A_i} \Delta MQ_{t+1}(e_i, a_i)$ becomes much bigger, which means the algorithm stability is lower.

5. Conclusion and Future Work

This paper originally proposed a multi-agent Q-learning to investigate the problem of admission control in WCDMA/WLAN heterogeneous networks. It provided efficient admission control policy for both new and handoff users thanks to the self-learning feature of multi-agent Q-learning method. In addition, the agent in threshold estimation block reserved parts of resources for handoff and RT call. This mechanism ensured the effectiveness of MQACM in the network jam situation, which resulted in a better universality than the other methods. Our MQACM could access more users while guaranteeing the QoS of users. Finally, the convergence proof of MQ was given.

MQACM in this paper is developed for WCDMA/WLAN heterogeneous networks. Actually, besides WCDMA and WLAN sub-networks, MQACM can be extended to heterogeneous networks including many types of networks in future.

References

[1] O.E. Falowo, "Joint call admission control algorithm for reducing call blocking/dropping probability in heterogeneous wireless networks supporting multihoming," *in Proc. of GLOBECOM Workshops*, pp. 611-615, 2010. Article (CrossRef Link)

[2] Y.S. Zhao, X. Li and H. Ji, "Radio admission control scheme for high-speed railway communication with MIMO antennas," *in Proc. of IEEE International Conference on Communications (ICC)*, pp. 5005-5009, 2012. Article (CrossRef Link)

[3] S.A. AlQahtani, A.S. Mahmoud, "Performance analysis of two throughput-based call admission control schemes for 3G WCDMA wireless networks supporting multiservices," *Computer Communications*, vol. 31, pp. 49-57, 2008. Article (CrossRef Link)

[4] S. Kim, Y.J. Cho, Y.K. Kim, "Computer networks admission control scheme based on priority access for wireless LANs," *Computer Networks*, vol. 54, pp. 3-12, 2010. Article (CrossRef Link)

[5] D. Fooladivanda and C. Rosenberg, "Joint resource allocation and user association for heterogeneous," *IEEE Transactions on Wireless Communications*, vol. 12, pp. 248-257, 2012. Digital Object Identifier: Article (CrossRef Link)

[6] D. Niyato and E. Hossain, "Dynamics of network selection in heterogeneous wireless networks: an evolutionary game approach," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 4, pp. 2008-2017, May, 2009. Article (CrossRef Link)

[7] C. Makaya, and S. Pierre, "An architecture for seamless mobility support in IP-based

next-generation wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 2, March, 2008. Article (CrossRef Link)

[8] H.T. Cheng and W.H. Zhuang, "QoS-driven MAC-layer resource allocation for wireless mesh networks with non-altruistic node cooperation and service differentiation," *IEEE Transactions on Wireless Communications*, vol. 8, no. 12, pp. 6089-6103, December, 2009. Article (CrossRef Link)

[9] Y. Choi, H. Kim, S.W. Han, et al, "Joint resource allocation for parallel multi-radio access in heterogeneous wireless networks," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3324-3329, November, 2010. Article (CrossRef Link)

[10] S.K. Lee, K. Sriram, K. Kim, et al, "Vertical handoff decision algorithms for providing optimized performance in heterogeneous wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 2, pp. 865-881, 2009. Article (CrossRef Link)

[11] K. Samdanis, and A.H. Aghvami, "Scalable inter-area handovers for hierarchical wireless networks", *IEEE Wireless Communications*, vol.16, no. 6, pp. 62-68, 2009. Article (CrossRef Link)

[12] W. Shen and Q.A. Zeng, "Cost-function-based network selection strategy in integrated wireless and mobile networks," *IEEE Transactions Vehicular Technology*, vol. 57, no. 6, pp. 3778-3788, November, 2008. Article (CrossRef Link)

[13] F. Yu and V. Krishnamurthy, "Optimal joint session admission control in integrated WLAN and CDMA cellular networks with vertical handoff," *IEEE Transactions Mobile Computing*, vol. 6, no. 1, pp. 126-139, April, 2007. Article (CrossRef Link)

[14] Y.H. Chen, C.J. Chang, and C.Y. Huang, "Fuzzy Q-learning admission control for WCDMA/WLAN heterogeneous networks with multimedia traffic," *IEEE Transactions on Mobile Computing*, vol. 8, no. 11, pp. 1469-1479, November, 2009. Article (CrossRef Link)

[15] Y. Hosoya and M. Umano, "Australia dynamic fuzzy Q-learning with facility of tuning and removing fuzzy rules," *in Proc. of IEEE World Congress on Computational Intelligence*, pp. 1-8. June, 2012. Article (CrossRef Link)

[16] D.Y. Ye, M.J. Zhang and D. Sutanto, "A hybrid multiagent framework with Q-learning for power grid systems restoration," *IEEE Transactions on Power Systems*, vol.26, pp. 2434-2441, 2011. Article (CrossRef Link)

[17] J.S. Lin and K.T. Feng, "QoS-based adaptive contention/reservation medium access control protocols for wireless local area networks," *IEEE Transactions on Mobile Computing*, vol. 10, pp. 1785-1803, 2011. Article (CrossRef Link)

[18] S. Zhao, W.X. Shi and S.S. Fan, et al, "A GRA-based network selection mechanism in heterogeneous wireless networks," *in Proc. of International Conference on Computer, Mechatronics, Control and Electronic Engineering (CMCE)*, pp. 215-218, 2010. Article (CrossRef Link)

[19] R. Ben Ali and S. Pierre, "On the impact of soft vertical handoff on optimal voice admission control in PCF-based WLANs loosely coupled to 3G networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 3, pp. 1356-1365, March, 2009. Article (CrossRef Link)

[20] K.S. Munasinghe and A. Jamalipour, "Interworked WiMAX-3G cellular data networks: an architecture for mobility management and performance evaluation," *Transactions on Wireless Communications*, vol. 8, no. 4, pp. 1847-1853, April, 2009. Article (CrossRef Link)

[21] J.Y.K. Aulin and D. Jeremic, "Compressive sensing aided determination of WCDMA constrained capacity," *in Proc. of IEEE International Conference on Communications (ICC)*, pp. 4072-4077, 2012. Article (CrossRef Link)

[22] H.Q. Zhai, X. Chen and Y.G. Fang, "How well can the IEEE 802.11 wireless LAN support quality of service?" *IEEE Transactions on Wireless Communications*, vol. 4, no. 6, pp. 3084-3094, 2005. Article (CrossRef Link)

[23] C. Brouzioutis, V. Vitsas, P. Chatzimisios, "Studying the impact of data traffic on voice capacity in IEEE 802.11 WLANs," *in Proc. of IEEE International Conference on Communications (ICC)*, pp. 1-6, 2010. Article (CrossRef Link)

[24] T. Venkatesh, Y.V. Kiran and C.S.R. Murthy, "Joint path and wavelength selection using

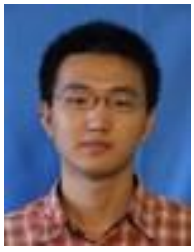
Copyright ⓒ 2013 KSII

Q-learning in optical burst switching networks," *in Proc. of IEEE International Conference on Communications*, pp. 1-5, August, 2009. Article (CrossRef Link)
[25] H.E. Kim and H.S. Ahn, "Convergence of multiagent Q-learning: multi action replay process approach," *in Proc. of IEEE International Symposium on Intelligent Control, Part of IEEE Multi-Conference on Systems and Control*, pp. 789-794, September, 2010. Article (CrossRef Link)
[26] C.J.C.H. Watkins and P. Dayan, "Technical note Q-learning," *Machine Learning*, vol. 8, pp. 279-292, 1992. Article (CrossRef Link)
[27] C. Makaya, and S. Pierre, "An analytical framework for performance evaluation of ipv6-based mobility management protocols," *IEEE Transactions on Wireless Communications*, vol. 7, no. 3, March, 2008. Article (CrossRef Link)
[28] Q.S. Guan, F.R. Yu, S.M. Jiang, et al, "Prediction-based topology control and routing in cognitive radio mobile Ad Hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 9, pp. 4443-4452, November, 2010. Article (CrossRef Link)

Dr. Jiamei Chen received the B.S. degree in 2008 and the M.S. degree in 2010, both in Information and Communication Engineering from Harbin Institute of Technology (HIT), Harbin, China. She studied in Purdue University as a visiting scholar from 2011 to 2012. Now, She is working toward the ph.D. degree in communication research center, Harbin institute of technology. Her research interests are heterogeneous wireless networks, artificial intelligence, cognitive radio and WLAN indoor location.

Yubin Xu was born in 1954. He received the B.S. degree in radio measurement, in 1986, the M.S. degree in Electronics and Communication System, in 1993, and Ph.D. degree in Electronics and Communication System, in 2005, all from HIT, Harbin, China. He is currently a Professor in Hit, Harbin, China. His research interests include wireless positioning, radio-wave propagation, collective and mobile communications.

Lin Ma was born in 1980. He received the B.S. degree in 2003, the M.S. degree in 2005 and PhD degree in 2009 all in Information and Communication Engineering from Harbin Institute of Technology (HIT), Harbin, China. Now, he is a lecturer in Information and Communication Engineering, HIT, China. His current research interests include WLAN indoor location, artificial intelligence and wireless communications.



Yao Wang received the B.S., M.S. degrees in communications engineering from Harbin institute of technology (HIT) in 2007, 2010, respectively. Now, he is currently working toward the Ph.D degree in communication research center, Harbin institute of technology. His major research interests are spectrum resource management and power control in cognitive radio and spread spectrum communication technology.