IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 17, 2018, accepted July 1, 2018, date of publication July 9, 2018, date of current version July 30, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2853730

QoE-Aware Intelligent Vertical Handoff Scheme Over Heterogeneous Wireless Access Networks

JIAMEI CHEN[1], YAO WANG[2], YUFENG LI[1], AND ERSHEN WANG[1]

[1]College of Electrical and Information Engineering, Shenyang Aerospace University, Shenyang 110136, China
[2]Department of Air Defense Forces, Noncommissioned Officer Academy, Institute of Army Artillery and Air Defense Forces, Shenyang 110867, China

Corresponding author: Yao Wang (wangyaowh2005@126.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61571309, in part by the Doctoral Scientific Research Foundation of Liaoning Province under Grant 20170520228, in part by the Liaoning Provincial Education Department Foundation under Grant L201602, and in part by the School Doctoral Scientific Research Foundation under Grant16YB01.

ABSTRACT As a measurement, quality of service (QoS) has been commonly taken into account in the traditional vertical handoff schemes for the heterogeneous wireless access networks. However, the QoS is not sufficient to correlate well with the user satisfaction. In this paper, quality of experience (QoE) is introduced into the decision mechanism of the vertical handoff and a random neural network -based QoE estimation is proposed to determine the correlation between the QoE and the QoS in the heterogeneous networks. In addition, a Q-learning-based vertical handoff algorithm, designated as a QoE-Q algorithm, is presented in order to maximize the QoE utility for users. It can be observed from the simulation results that the proposed method not only outperforms the existing schemes with enhanced call blocking probability and handoff dropping probability property but also obtains better QoE performance in the service charges and the terminal power consumption than other schemes.

INDEX TERMS Heterogeneous wireless access networks, QoE, Q-learning, vertical handoff.

I. INTRODUCTION

The heterogeneous networks are widely regarded, by both industry and academia, as the solution to the problem of being able to cope with an ever increasing demand for ubiquitous coverage along with high data rate [1]–[3]. With the associated research carried out, the heterogeneous networks are becoming accessible. Exploiting the heterogeneous network resources effectively highly depends on the radio resource management (RRM) strategy, including handover. The handover scheme is crucial for guaranteeing the performance of the heterogeneous wireless networks [4], [5].

Although there are extensive researches on handover for heterogeneous wireless networks, most existing protocols put more emphasis on QoS [6], [7]. The reason why QoS often acts as the essential criterion is that it can be measured practically and help the engineers to improve the service quality. However, the traditional QoS solutions cannot guarantee QoE of users, since QoS merely reflects the network properties while it does not directly indicate the users' satisfaction [8]–[10]. Actually, the expectations of different users for the services and applications are different. For instance, some VIP users prefer stronger QoS with higher tariff, so as to achieve their desired level of QoE. Nevertheless, the higher QoE level of ordinary users is in view of a cheaper and general level QoS. Another case, some users experience a bad service due to crowd accessibility even though the signal strength is still good. Statistics on [11] show that about 90% of customers will not complain and they will simply leave (churn) once they become unsatisfied. This churn directly affects the profitability and the image of the operator, especially if it happens in the early stage of their induction. To deal with this new problem, novel handover techniques are required with a comprehensive consideration of both QoE and QoS provisioning. And the network stability and the resource utilization also need to be taken into account.

Compared with QoS, QoE concentrates more on the performance of the services which represents the subjective feelings of users. Actually, several works have been conducted in the recent literatures. Document [12] investigates an admission control mechanism for the macro or small cell networks. The decision is based on users' QoE. According to the Markov decision process model, the algorithm aims to

seek the optimal policy that maximizes users' QoE. Senouci proposes an optimal framework that minimizes the energy consumption while satisfying a desired level of a heterogeneous cellular network, considered as a QoE term [13]. In [14], Deng *et al.* propose a novel QoE prediction model. The rate allocation problem is formulated as a nonlinear optimization problem and a relaxing function called penalty function is exploited to transform the rate allocation problem into an unconstrained optimizing problem. The pattern search method is adopted to obtain the close-to-optimal solution in each rate allocation session. Furthermore, a general QoE model has been established by jointly characterizing the multimedia applications and heterogeneous networks [15]. The issue of the multimedia communications over the heterogeneous networks has been transformed as a stochastic optimization problem, and the optimal solution is provided by designing a QoE-aware multimedia scheduling scheme. More importantly, the proposed scheme is simple enough for online implementation. In addition, Wu *et al.* [16] tend to select networks that can maximize QoE of users. They formulate the network selection problem as a continuous-time multi-armed bandit (CT-MAB) problem. A traffic-aware online network selection (ONES) algorithm is designed to match typical traffic types of users with respective optimal networks in terms of QoE. Most of the existing methods get the score of QoE via the actual interview test to users. And the scoring system of Mean Opinion Score (MOS) is commonly adopted in these methods, which is generally divided into 5 grades, i.e., 1, 2, 3, 4, 5, according to the provisions of ITU [12]. Users that are requested to score their satisfaction levels may be given incentives for scoring the service. And the satisfaction levels can be collected from a randomly selected group of users, for instance, using Short Message Service (SMS), through a web portal or through a dedicated application [17]. Unfortunately, it is expensive, time consuming and the portability is not strong for this actual test way. Considering that the QoS parameters are easy to be measured, this paper uses a QoE evaluation to find the mapping relationship between MOS scores and QoS parameters, rather than testing the real users every time.

Based on issues above, we mainly focus on the user QoE and present QoE-Q handover mechanism from the perspective of the user experience for heterogeneous networks. First of all, a RNN based QoE estimation algorithm is developed as a foundation of the following handover algorithm. The advantage of RNN is that it can obtain the output associated with any input, which is conducive to study the relationship between QoE and any objective QoS parameters. A well trained RNN can get output with quite high quality even the input parameter values extend beyond the normal range. This is due to the marvelous inference ability of RNN. Afterwards, a Q-learning algorithm based vertical handoff mechanism is proposed. Specifically, we first design the primary elements of Q-learning algorithm, namely status space, actions and immediate reward, then, the implementation steps of Q-learning algorithm are described to maximize users' QoE and ultimately increase the revenue. It is worth noting that the heterogeneous networks here refer to that have been introduced in the LTE-Advanced standardization [18]. A heterogeneous network uses a mixture of macrocells and small cells such as microcells, picocells, and femtocells.

The objective of this article is to improve the overall QoE level of users in the hetoregeneous networks. The main contributions and distinctions in this paper are Four-fold. Firstly, a RNN based QoE estimation algorithm is proposed to avoid the complicated MOS test work by training the relationship between the MOS scores and the QoS values. Secondly, a Q-learning algorithm is developed for the heterogeneous network handover, and the network agent can obtain the optimal handover strategy through the continuous interaction with the environment according to its own online learning characteristics. Thirdly, thanks to RNN for the established link between the objective QoS and the subjective MOS, the MOS scores out from RNN can be used as the system state of Q-learning algorithm, and the other subjective factors such as the network charges and the power consumption rate of terminals are integrated into the immediate reward of Q-learning. All these subjective and objective factors are all taken into account and combined well together to achieve the purpose of improving users' QoE. Last but not least, we conduct some simulations to confirm the effectiveness of the proposed QoE-Q algorithm compared with the traditional schemes.

The rest of this paper is organized as follows. Section II investigates the Random Neural Network theoretic, and then, the RNN based QoE estimation is described. The key elements of Q-learning are designed in section III and then mathematical design of QoE-Q is discussed. Section IV provides the simulation results. And conclusions are drawn in section V.

II. RNN BASED QoE EVALUATION

In addition to the objective QoS parameters, the elements that affect QoE also include the subjective factors, such as the power consumption rate of the mobile terminals and the difference of the charges for different subnets. Consequently, we will seek the maximization of users' QoE taking into account both of the two elements above. In this section, the impact of QoS on QoE is invesgated by means of QoE evaluation. The subjective factors affecting QoE will be considered in section III.

A. RANDOM NEURAL NETWORK THEORETIC

RNN imitating the biological neuron is initially proposed by Gelenbe [19], [20]. The random nature of RNN is embodied in the phenomenon that the neuron (node) reaction on the same stimulus is different at different time. RNN is a kind of dynamic stochastic system which is constituted of many interconnected neurons, and the calculation units in this system are in charge of the calculation. Hereinafter we focus mainly on the feed-forward RNN which consists of input neurons Γ with I elements, output neurons Ω with O elements and hidden

neurons Ψ with H elements, just like other feed-forward neural networks. Here, hidden neurons can be one layer or more layers. The calculation units enter the system from the input neurons; arrive at the output neurons through the hidden neurons; and finally leave RNN. For notational convenience, we make the following hypothesis: the movement speed of one calculation unit in the node i is v_i, and there are n ($n > 0$) calculation units are stored in node i at time t_1, and then the n calculation units leave node i at time t_2. Assume $t_2 - t_1$ obeys the exponential distribution with density v_i. When the calculation unit moves to the output neurons, node i belongs to output neurons ($i \in \Omega$). It indicates that the calculation units have left RNN. The probability of a calculation unit transferred from node i to the next node j is $p^+_{i,j}$ or $p^-_{i,j}$, in which $i \in \Gamma$, $j \in \Psi$, or $i \in \Psi$, $j \in \Omega$. Additionally, $p^+_{i,j}$ and $p^-_{i,j}$ represent the probability of the passing positive and negative signals, respectively. And the relationship between them can be given by

$$\sum\nolimits_j (p^+_{i,j} + p^-_{i,j}) = 1 \tag{1}$$

p_{ii} and p_{jj} refer to the probability that the node transmits signal to itself, and the equation (2) represents node itself does not transmits signal to itself,

$$p_{ii} = p_{jj} = 0 \ (1 \le i, j \le N) \tag{2}$$

Where N is the total number of the nodes. The calculation units arrive at input neurons Γ from the external environment according to the independent Poisson distribution. The calculation units with a positive and negative signal arrive at input neurons Γ under the Poisson distribution with speed λ^+_i or λ^-_i, respectively. Assume that for any $i \in \Gamma$ there is some $h \in \Psi$ and some $o \in \Omega$ such that $\left(p^+_{i,h} + p^-_{i,h}\right)\left(p^+_{h,o} + p^-_{h,o}\right) > 0$. In other words, every node in Γ and in Ψ can send its units outside.

B. FUZZY NORMALIZATION OF QoS PARAMETERS

The range of every QoS parameter is different, which makes the evaluation difficult. Therefore, it is necessary to normalize the QoS parameters to the interval [0, 1]. Several QoS metrics commonly used are the transmission rate, delay, jitter and bit error rate. Next, the normalization of these metrics is elaborated as follows.

Apparently, when the transmission rate is normalized to the interval [0, 1], the upper limit 1 is best. Here, the half rising trapezoid fuzzy membership function is adopted in the normalization

$$f_D(r_m, k) = \begin{cases} 0 & r_m \le r_{k\,\min} \\ (r_m - r_{k\,\min})/(r_{k\,\max} - r_{k\,\min}) & \\ & r_{k\,\min} < r_m < r_{k\,\max} \\ 1 & r_m \ge r_{k\,\max} \end{cases} \tag{3}$$

Where $f_D(r_m, k)$ represents the normalized data rate for user m with service k; r_m is the actual data rate; $r_{k\,\max}$ is the ideal data rate; $r_{k\,\min}$ is the minimum data rate.

Different from the transmission rate, the other three QoS parameters are expected as small as possible. In other words, the lower limit 0 is best. Now suppose there are β kinds of QoS parameters, and the semi-descending trapezoidal fuzzy membership function is adopted to describe the normalization of these parameters, which can be written as:

$$f_D(q^\beta_m, k) = \begin{cases} 1 & q^\beta_m \le q^\beta_{k\,\min} \\ \dfrac{q^\beta_{k\,\max} - q^\beta_m}{q^\beta_{k\,\max} - q^\beta_{k\,\min}} & q^\beta_{k\,\min} < q^\beta_m < q^\beta_{k\,\max} \\ 0 & q^\beta_m \ge q^\beta_{k\,\max} \end{cases}$$
$$\text{s.t. } q^\beta_m \ne r_m, \quad \beta = 3 \tag{4}$$

$f_D(q^\beta_m, k)$ is the normalization of the βth parameter for user m with service k. q^β_m is the actual value of the βth parameter for user$m$. $q^\beta_{k\,\max}$ and $q^\beta_{k\,\min}$ are the upper limit value and the lower limit value of the βth parameter for users under service k, respectively. For instance, the delay of the terminals can be normalized as

$$f_D(o_m, k) = \begin{cases} 1 & o_m \le o_{k\,\min} \\ \dfrac{o_{k\,\max} - o_m}{o_{k\,\max} - o_{k\,\min}} & o_{k\,\min} < o_m < o_{k\,\max} \\ 0 & o_m \ge o_{k\,\max} \end{cases} \tag{5}$$

Where $f_D(o_m, k)$ is the normalized delay for user m. o_m is the actual delay. $o_{k\,\max}$ and $o_{k\,\min}$ are the upper and lower limit of the delay, respectively. The other parameters can also be normalized through the similar derivation process, so we do not reiterate them for simplicity.

C. STEPS OF QoE EVALUATION

The design of the RNN evaluation algorithm is capable of avoiding the tedious and complex work of MOS score testing, while ensuring the degree of accuracy. In other words, the MOS scores can be obtained by inputting the QoS parameters into the RNN estimator, without having to dial and test the users every time. And the MOS scores output from the RNN evaluation are able to consistent with the actual scores to a certain extent.

Let vector $\vec{s} = (s_1, s_2, \ldots, s_n)$ denote the QoS parameters, and g denote the MOS value of QoE. Suppose that there are Z QoS parameters and MOS pairs, which can be defined as $(\vec{s}^{(z)}, g^{(z)})_{z=1..Z}$. Now a RNN with $I = n$ input nodes and $O = 1$ output nodes is considered. Then, we set the negative units arriving from outside to 0. Correspondingly, let $\lambda^+_i = s_i$, $i = 1, 2 \ldots n$, and $\rho_o = g$. For all $i \in \Omega$, we have $\lambda^+_i = s^{(z)}_i$ and $\rho^{(z)}_o \approx g^{(z)}$. For notational convenience, all the QoS parameters are normalized to [0, 1] by means of the method in part *B* of section II. And here the value of v_i can be set. For example, let $H = 2n$, $v_i = 1$ for all $i \in \Gamma$ and let $v_i = n$ for the other nodes. In order to achieve RNN with the ideal performance, a function can be denoted as

$$C = \frac{1}{2}\sum_{z=1}^{Z} (\rho^{(z)}_o - g^{(z)})^2 \tag{6}$$

Where for a given RNN, $\rho_o^{(z)}$ is the probability that neuron $o(o \in \Omega)$ is excited under the condition of $\lambda_i^+ = s_i^{(z)}$ for all $i \in \Theta$, i.e., the probability that there are calculation units out from the output nodes.

The purpose of RNN learning is to minimize the function C. Firstly, the variable substitution is formulated as

$$w_{i,j}^+ = v_i p_{i,j}^+ \tag{7}$$

$$w_{i,j}^- = v_i p_{i,j}^- \tag{8}$$

Just like other neural networks, we represent this new variable substitution $\vec{\boldsymbol{w}}$ as weight, and the function C can be written as an implicit function in $\vec{\boldsymbol{w}}$, i.e.,$C = C(\vec{\boldsymbol{w}})$, which is derived as

$$\rho_o^{(z)} = \frac{\sum_{h\in\Psi} \frac{\sum_{i\in\Gamma}(s_i^{(z)}/v_i)w_{i,h}^+}{v_h+\sum_{i\in\Gamma}(s_i^{(z)}/v_i)w_{i,h}^-} w_{h,o}^+}{\mu_o + \sum_{h\in\Psi} \frac{\sum_{i\in\Gamma}(s_i^{(z)}/v_i)w_{i,h}^+}{v_h+\sum_{i\in\Gamma}(s_i^{(z)}/v_i)w_{i,h}^-} w_{h,o}^-} \tag{9}$$

It is worth noting that, $\vec{\boldsymbol{w}}$ is actually a set of new variables; for instance, $w_{h,o}^+$ in Eq.(9) represents the probability that the calculation unit is transmitted from node h ($h \in \Psi$) to node $o(o \in \Omega)$. Under the appropriate mathematical constraints, minimizing $C(\vec{\boldsymbol{w}})$ by the specific minimization process can be achieved. Meanwhile, $\vec{\boldsymbol{w}}$ should conform to the following set

$$\{\vec{\boldsymbol{w}} \geq \vec{0} : \ \forall i \notin \Theta, \ \sum_j (w_{i,j}^+ + w_{i,j}^-) = v_i \} \tag{10}$$

According to the theoretical analyses above, the specific steps and processes of the QoE evaluation are given as follows.

Step 1: Extract M users randomly in the heterogeneous networks. Investigate D time intervals for each kind of service k the user carrying on. The MOS scores $g_m^{k,d}$ of every kind of service k in various time segments D is obtained by calling each extracted user m. Thus, the average value of the MOS scores can be derived as g_m^k:

$$g_m^k = \frac{1}{D}\sum_{d=1}^{D} g_m^{k,d} \tag{11}$$

Step 2: Choose l kinds of QoS parameters, e.g., transmission rate, packet loss rate, delay and jitter. Then collect and note the QoS sample values $\vec{\boldsymbol{q}}_m^{k,d} = \left(q_{m,1}^{k,d}, q_{m,2}^{k,d}, \ldots, q_{m,l}^{k,d}\right)$ of every kind of service k in various time segments D. Accordingly, the average QoS value is $\vec{\boldsymbol{q}}_m^k = \left(q_{m,1}^k, q_{m,2}^k, \ldots, q_{m,l}^k\right)$. After normalization, this average value of the interval [0,1] can be expressed by $\vec{\boldsymbol{f}}_D(q_m, k) = f_D(q_{m,1}, k),\ f_D(q_{m,2}, k), \ldots, f_D(q_{m,l}, k)$. Then, the variable substitution can be conducted as below. Let $\vec{\boldsymbol{s}}_m^k = \vec{\boldsymbol{f}}_D(q_m, k)$, i.e., $\vec{\boldsymbol{s}}_m^k = \left(s_{m,1}^k,\ s_{m,2}^k, \ldots, s_{m,l}^k\right)$.

Step 3: Divide Z pairs $(\vec{\boldsymbol{s}}_z^k,\ g_z^k)_{z=1..Z}$ into training samples $(\vec{\boldsymbol{s}}_{tr}^k,\ g_{tr}^k)_{tr=1..TR}$ and test samples $(\vec{\boldsymbol{s}}_{te}^k,\ g_{te}^k)_{te=1..TE}$, in which $TR + TE = M$ is satisfied. Then, the samples of the two parts are put into RNN.

Step 4: Train RNN with the training samples $(\vec{\boldsymbol{v}}_{tr}^k,\ g_{tr}^k)_{tr=1..TR}$. The mapping relationship of the MOS scores and the QoS values is obtained by adjusting the weights continuously. When the convergence condition is satisfied, the mapping relation $f(\cdot)$ between the MOS scores and the QoS values can be achieved. It is worth mentioning that this process is an off-line training process and does not occupy the computing resources of the handover algorithm.

Step 5: Input the QoS values of the test samples to RNN, and then RNN will give out the MOS scores according to $f(\cdot)$. Compare the obtained MOS scores with the real MOS values. If the two values mentioned above are close enough, the QoE evaluation is successful. In other words, the mapping between QoS and QoE has been established through the RNN.

It is worth noting that, considering VIP and the ordinary users may have different expectations for the same QoS, we separate them and train different RNN models to get different corresponding relationship between the MOS scores and the QoS values. In the two models, VIP users and the ordinary users give different MOS scores for the same QoS, resulting in different maps of the RNN training. Actually, these two maps bear no difference in essence, so in the simulation part we only do the statistics for the VIP users.

III. Q-LEARNING BASED VERTICAL HANDOFF

The optimum decisions of handover are obtained through embedded Q-learning. Q-learning system uses an agent to learn how to improve its decisions during the time of learning, according to their historical experiments.

A. STATUS SPACE AND ACTIONS

In this paper, the vertical handoff algorithm is designed based on the user QoE. Therefore, when a user initiates a handoff call request, the handoff decision criteria is no longer according to the network status of the subnets, but according to the QoE level state of the users that have been accessed to the subnets, so as to ensure that the user can get the best quality of service experience. Suppose the heterogeneous network environment has been already established, which is an integrated system containing a single cellular subnet and a single WLAN subnet. And the WLAN subnet overlaid on the cellular subnet. Assume e is the QoE level state, and the definition of the state space E of the system can be given as following:

$$E = \{k, \text{MOS}\} \tag{12}$$

Where k is the service type, which includes four kinds: conversation class, streaming class, interactive class and background class, i.e., $k = 4$. MOS is the satisfaction level for users. Moreover, MOS is divided into five grades: 1, 2, 3, 4 and 5. Thus, the state space dimension for users is $4 \times 5 = 20$.

In the switch process, if a new user initiates an access request in the double coverage area, the agent in the Q-learning will access the user to the cellular or WLAN subnet according to the QoE level state of each subnet; if a

new user initiates an access request in the single coverage area, the agent will access the user to the corresponding cellular subnet. If there are not enough resources in the corresponding subnet, the user request will be rejected. For handoff users, if the handover request is initiated from a single to double coverage area, the agent will access the user to the cellular or WLAN subnet. On the contrary, if the engaged user initiates a handoff request from a double coverage area to a single coverage area, he/she can only be accessed to the cellular subnet. Similarly, if there are not enough resources in the corresponding subnet, the handoff request will be rejected. Thus, the action a is defined as follows:

$$a = \begin{cases} 1 & \text{access to cellular subnet} \\ 2 & \text{access to WLAN subnet} \\ 3 & \text{handoff to cellular subnet} \\ 4 & \text{handoff to WLAN subnet} \\ -1 & \text{dwell in the original subnet} \\ 0 & \text{reject} \end{cases} \tag{13}$$

The optional action set for network switching A can be defined as

$$\begin{aligned} A = \{&[a_{s\_n\_RT}, a_{s\_n\_NRT}, a_{s_h\_RT}, a_{s_h\_NRT}, a_{d_n\_RT}, \\ &a_{d_n\_NRT}, a_{d_h\_RT}, a_{d\_h\_NRT}],\ a_{s\_n\_RT/NRT} \in \{1, 0\}, \\ &a_{s_h\_RT/NRT} \in \{4, -1, 0\}, a_{d_n\_RT/NRT} \in \{1, 2, 0\}, \\ &a_{d_h\_RT/NRT} \in \{3, -1, 0\}\} \end{aligned} \tag{14}$$

Where $a_{s\_n\_RT/NRT}$ and $a_{s_h\_RT/NRT}$ are the actions of a new and handoff user in the single overlay area, respectively. $a_{d_n\_RT/NRT}$ and $a_{d_n\_RT/NRT}$ are the actions of a new and handoff user in the double overlay area, respectively.

B. IMMEDIATE REWARD

Denote $p_t(e, a)$ as the immediate reward, which is the enhanced signal of the Q-learning agent. Through the feedback of $p_t(e, a)$, the agent can gradually approximate to the optimum strategy. Since different service k will produce different feedback, the specific form of $p_t(e, a)$ is written as

$$p_t^k(e, a) = p(Q(k),\ C(k),\ U(k)) \tag{15}$$

$Q(k) = Q^k(rat, del, jit, ber)$ indicates the condition of QoS, which contains the value of the data rate, the delay, the jitter and the BER for users. $C(k)$ is the network charge and can be set according to the actual tariff standard of network operators. $U(k)$ represents the power consumption rate of terminals, and its unit is mA. The instantaneous current intensity is adopted to stand for the power consumption rate. Generally, diverse network cards will bring different power consumption rate for terminals. And the power consumption rate using cellular network cards is much smaller than that using WLAN network cards.

As mentioned above, immediate rewards need to be counted for different service types. The reason is that different service types require different QoS parameter values. Here a threshold is defined for each QoS parameter in terms of service type k, which is given by

$$th = \{th_k^{rat}, th_k^{del}, th_k^{jit}, th_k^{ber}\} \tag{16}$$

Where th_k^{rat} is the minimum threshold of data rate; th_k^{del} is the maximum threshold of delay; th_k^{jit} is the maximum threshold of jitter; th_k^{ber} is the maximum threshold of BER. In particular, we select one concrete service in each type of services, i.e., interactive voice service in the session class service, one-way video streaming service in the browse business, web browsing interactive in services and email in background services. Table 1 lists the threshold requirements for each QoS parameter.

TABLE 1. Threshold requirements of QoS parameter under different services.

Service type	Data rate	Delay	Jitter	BER
Interactive voice	4kbit/s	400ms	1ms	3%
One-way video browsing	16kbit/s	1000ms	1ms	3%
Web browsing	80kbit/s	4000ms	——	0
Email	80kbit/s	A few minutes or longer	——	10-6

Thus, the immediate rewards can be formulated as

$$p = \text{flag} \cdot [\alpha_1 Q + \alpha_2 \frac{1}{1 + e^{C_{ne} - C_{cu}}} + \alpha_3 \frac{1}{1 + e^{U_{ne} - U_{cu}}}] \tag{17}$$

Where C_{cu} is the charge in the current subnet of the user. C_{ne} is the charge in the target subnet that the user will switch into. U_{cu} is the current power consumption rate of the terminal. U_{ne} is the power consumption rate after the corresponding action is executed. α_1, α_2 and α_3 are three weight parameters that can be adjusted. Furthermore, these three parameters can be set according to the preferences of users, and satisfy the following constraints

$$\{\alpha_1, \alpha_2, \alpha_3 \in [0, 1],\ \alpha_1 + \alpha_2 + \alpha_3 = 1\} \tag{18}$$

In addition, flag $\in \{0, 1\}$. Where, flag $= 0$ indicates that one or more QoS parameters cannot meet the request of the minimum thresholds. flag $= 1$ means that all the QoS parameters are able to reach the minimum thresholds. Q represents the QoS of users, which can be obtained as

$$\begin{aligned} Q = &\frac{1}{1 + e^{-(rat_{ne} - rat_{cu})}} + \frac{1}{1 + e^{(del_{ne} - del_{cu})}} \\ &+ \frac{1}{1 + e^{(jit_{ne} - jit_{cu})}} + \frac{1}{1 + e^{(ber_{ne} - ber_{cu})}} \end{aligned} \tag{19}$$

Where rat_{cu}, del_{cu}, jit_{cu} and ber_{cu} are the QoS of users in current subnets. rat_{ne}, del_{ne}, jit_{ne} and ber_{ne} are the QoS of users in target subnets.

C. STEPS FOR Q-LEARNING BASED HANDOVER PROCESS

In the Q-learning [21], [22] based handover process, the agent observes the state e of the environment and implements an action a from the action set A. The environment manifests

the consequence of that actions correct or not by delivering the agent the immediate rewards, and then steps to a new state. The purpose of the agent is to find an optimal policy, i.e., a sequence of actions for each state, to maximize the total expected discounted rewards. The Q-learning algorithm exploits the optimal policy in an iterative manner by updating the Q-values as follows

$$Q_{t+1}(e,a) = (1-\alpha)Q_t(e,a) + \alpha\left\{p_t(e,a) + \gamma \max_{b\in A}[Q_{t+1}(e,b)]\right\} \quad (20)$$

Where α stands for the learning parameter, and $0 < \gamma < 1$ is the discount factor. The Q-learning agent implements an action at a particular state, and then evaluates the consequences of the action through the sum of the immediate rewards and the rewards obtained in future. Thus the agent learns the best action that maximizes the long-term discounted rewards by implementing each action. The main steps are given as follows.

Step 1: Initialize the Q-value. A Q-value matrix is generated in the form of random numbers.

Step 2: When a user launches a handoff requirement, the agent implements an action a by Eq.(13) according to the current state e_t; records the current state and the next state after implementing the action a.

Step 3: When the network changes to another state, the agent calculates the immediate reward $p_t(e,a)$ and records the next state e_{t+1} after implementing the action a.

Step 4: Update the matrix $Q_{t+1}(e,a)$ according to Eq.(20).

Step 5: When $\Delta Q(e,a) < \varepsilon$, $\forall e \in E$, $a \in A$, the algorithm arrives at its convergence. If it does not arrive at the convergence, continue to repeat step 2 to step 4.

By [23], we know a theorem in the single-agent Q-learning: given bounded immediate rewards $|p_t(e,a)| \leq \Re$, learning rate parameter $0 \leq \alpha < 1$ and

$$\sum_{j=1}^{\infty}\alpha^j = \infty, \quad \sum_{j=1}^{\infty}(\alpha^j)^2 < \infty, \ \forall e, \ a \quad (21)$$

Then $Q_t(e,a) \to Q^*(e,a)$ as $t \to \infty$, $\forall e$, a, with probability 1. Where, $Q^*(e,a)$ is optimal value of $Q_t(e,a)$ of state e, and $\Re$ is the largest $p_t(e,a)$. Furthermore, it is simple to prove that all the convergence conditions above are satisfied in our algorithm.

IV. SIMULATION RESULTS AND ANALYSIS

The simulation evaluates the relative performance of the proposed QoE-Q algorithm by comparing with two other classic algorithms, Fuzzy Q-learning Admission Control (FQAC) [24] and Markov Decision Processes (MDP) [25]. And the aim is to observe the trends in the performance metrics with varying arrival rate. The simulation is conducted with the following settings: for the cellular subnet, the channel experiences additive Gauss white noise; the path loss model is based on the log-normal shadowing model and multipath fading; the power control is ideal. For WLAN subnet, IEEE 802.11b acts as the standard and the average transmission rate is set to be 11Mbps. Rayleigh channel model is adopted. The users are uniformly distributed in the heterogeneous network environment. New users and handoff users both obey Poisson distribution with arrive rate λ_n and λ_h, respectively, and $\lambda = \lambda_n + \lambda_h$ is satisfied. The proportions of the four types for services are: conversation class is 40%; streaming class is 30%; interactive class is 20%; background class is 10%. Simulation parameters are shown in Table 2.

TABLE 2. Simulation parameters.

Parameter	Value
Learning factor(α)	0.1
Discount factor(γ)	0.95
Initial battery power	1900mAh
Power consumption rate of WLAN card	280mA
Power consumption rate of cell card	150mA

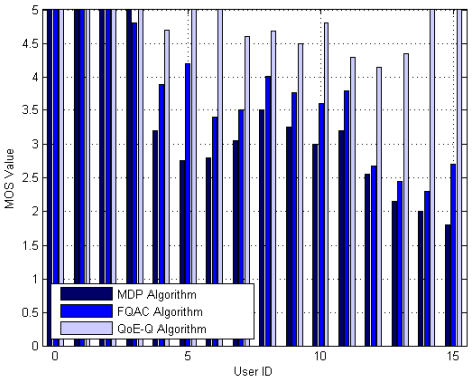


FIGURE 1. MOS values of different user ID.

Fig. 1 shows the MOS values versus different users, i.e., the real user experience of service quality. It can be obviously found that the MOS value in our QoE-Q algorithm is higher than that in the other two algorithms for the same user. The reason is that QoE-Q algorithm considers more about QoE. For instance, QoE is used as the state of the system in Q-learning. Some subjective QoE factors, such as the business tariff and the terminal power, are also embedded in the immediate rewards. Therefore, the handoff decision of QoE-Q algorithm could maximize the user QoE. In other words, QoE-Q algorithm not only meets the QoS requirements of the users, but also optimizes the tariff and the terminal power consumption. The reason why the MOS value of FQAC algorithm is a little bit worse is that FQAC algorithm just optimizes the objective QoS parameters. Therefore, it cannot give users the best service experience value in many cases.

Fig. 2 and Fig. 3 vividly depict the call blocking probability of new NRT/RT services for the three algorithms, respectively. From the two figures, we can see that the call blocking performance of the proposed QoE-Q algorithm is better

than the other two algorithms. It is because QoE-Q algorithm, which is designed from the perspective of QoE, can access more users with lower QoS requirement when the long period of Q-learning process converges. In addition, the sharp increase of the dimension and computational complexity with the growth of the state space brings great difficulty to MDP algorithm, which leads to a high call blocking probability.

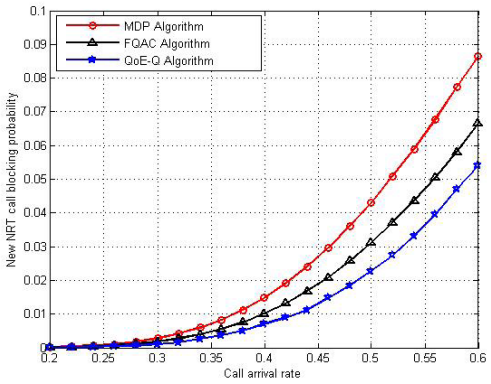


FIGURE 2. New NRT call blocking probability underunder different arrival rate.

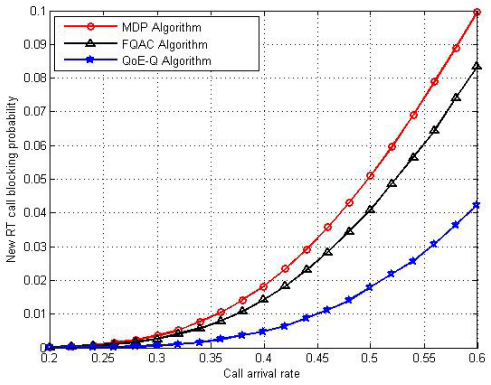


FIGURE 3. New RT call blocking probability under different arrival rate.

Fig. 4 and Fig. 5 show the handoff dropping probability of engaged NRT/RT services for the three algorithms, respectively. According to the two figures, it can be observed that the proposed QoE-Q algorithm obtains better handoff probability performance than the other two algorithms. It is due to the fact that the rejected users, whose QoS cannot be satisfied in FQAC algorithm and MDP algorithm, can be accessed to the system for QoE-Q algorithm. In a sense, it actually improves the relative resources of the network through QoE-Q algorithm, thereby reducing the dropping probability for users.

As shown from Fig. 2 to Fig. 5, it also can be obtained that the call blocking and handoff dropping rate of RT service is higher than NRT service for all the three algorithms under the same conditions. The reason is that RT users are more impatient than NRT users. In other words, the users with RT service will leave the network in a short period of time if the service is not available, which results in a higher call blocking

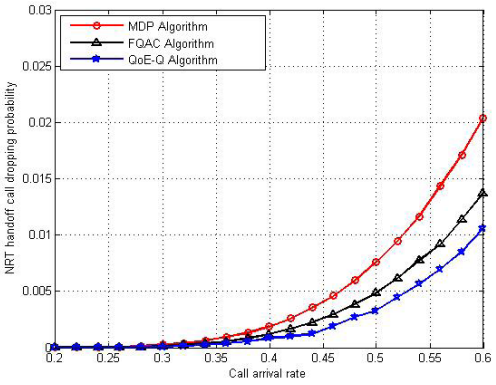


FIGURE 4. NRT handoff call dropping probability under different arrival rate.

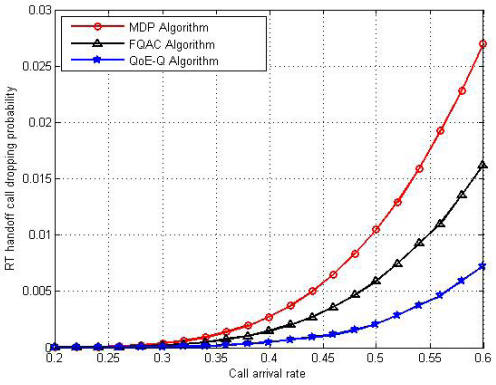


FIGURE 5. RT handoff call dropping probability under different arrival rate.

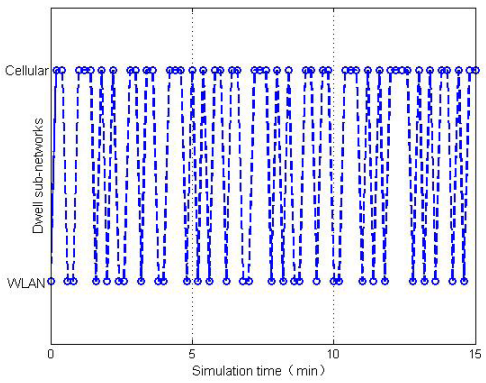


FIGURE 6. Dwell subnets for MDP algorithm.

and handoff dropping rate. However, the corresponding call blocking and handoff dropping rate of NRT service is higher than that of RT service in QoE-Q algorithm under the same conditions, which is because of that QoE-Q algorithm puts more emphasis on RT users and distributes a higher priority for RT users. This is also a manifestation of the pursuit of high QoE by the proposed algorithm.

Fig. 6, Fig. 7 and Fig. 8 demonstrate the dwell time of the mobile terminals in the subnets for the three handoff algorithms. Furthermore, it can observed that, the mobile terminals guided by QoE-Q algorithm resides in the WLAN subnet longer than in the cellular sub-networks compared with the

other two algorithms. It is because the tariff is reflected by the designing of the immediate reward in QoE-Q algorithm. As a result, more users are considered to be accessed to the WLAN subnet by the agent in QoE-Q algorithm, which can avoid the high service charges.

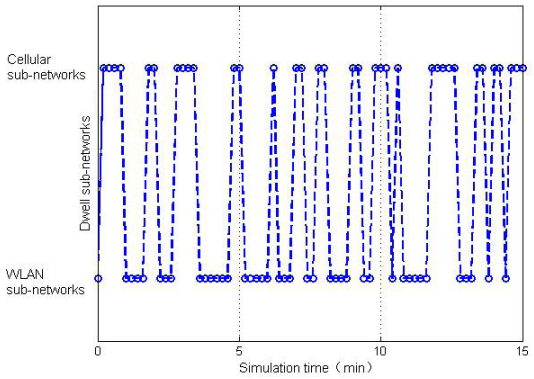


FIGURE 7. Dwell subnets for FQAC algorithm.

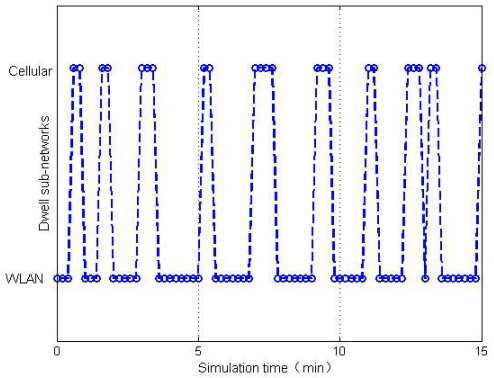


FIGURE 8. Dwell subnets for QoE-Q algorithm.

TABLE 3. Comparison of handoff numbers.

Algorithm	Handover Times
QoE-Q Algorithm	18
FQAC Algorithm	27
MDP Algorithm	49

Additionally, it can be concluded from Fig. 6 to Fig. 8 that the three algorithms have different switching times. The specific number of switches is shown in Table 3. The switching times for QoE-Q algorithm are significantly lower than the other two algorithms. Moreover, lower switching times by QoE-Q algorithm lead to lower power consumption to some extent. The reason is that QoE-Q algorithm considers more objective QoS parameters in RNN. Meanwhile, it takes some subjective QoE ingredients into account, just like terminal power consumption and network tariff, which provides more optimized handoff decision guidance for users. QoE-Q algorithm learns the optimal switching strategy from the perspective of long-term benefits for users. However, the ultimate target is to increase the whole network benefits, and the decrease of unnecessary switching times is a good proof.

V. CONCLUSION

In this paper, QoE as an essential criterion is taken into the consideration in order to represent the user satisfaction in the heterogeneous network environment. First of all, a QoE evaluation mechanism on the basis of RNN is proposed to seek the mapping relationship between the QoS values and the MOS values. Moreover, QoE-Q algorithm is proposed in view of the Q-learning theory, which utilizes the MOS value as the state of the system and designs the immediate rewards by considering the terminal power consumption and service charges. Simulation results confirm that our QoE-Q algorithm outperforms the other representative algorithms in terms of the call blocking probability and the handoff dropping probability. Especially, QoE-Q algorithm shows its excellent UE performance in the service charges and the terminal power consumption than other algorithms.

APPENDIX

CONVERGENCE ANALYSIS OF THE HANDOFF ALGORITHM

The convergence speed is an important factor influencing the performance of QoE-Q algorithm. If the convergence speed is too slow or it does not converge, the algorithm is hard to be applied in practice. The convergence property and the convergence speed of QoE-Q algorithm are directly related to the behavior exploration strategy. Some researches show that the Glie strategy can make Q-learning exhibit better convergence characteristics. Boltzmann, as a common exploration strategy, can be expressed as:

$$P[A_t = a|r] = \frac{e^{r(a)/T_t}}{\sum_{u\in A} e^{r(u)/T_t}} \tag{22}$$

T_t is the temperature parameter that is used to control the attenuation rate of Boltzmann policy [26]. Assuming $\Gamma = (N, (A_k), (r_k))$ is a non weak periodic game for N person, if

$$K \le \frac{m}{L(\Gamma)+2} \tag{23}$$

Moreover, if the agent uses Glie strategy to search, it can obtain an optimal action selection strategy with probability one. Formula (23) satisfies $1 \le K \le m$, and $L(\Gamma)$ is defined as:

$$L(\Gamma) = \max_{a\in A} L(a) \tag{24}$$

$L(a)$ represents the shortest path to the Nash equilibrium, and its detailed proof process can be seen in literature [27] and [28].

REFERENCES

[1] X. Xu and M. Wang, "Inferring disease associated phosphorylation sites via random walk on multi-layer heterogeneous network," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 13, no. 5, pp. 836–844, Sep./Oct. 2016, doi: 10.1109/TCBB.2015.2498548.

[2] M. Mahdian and E. M. Yeh, "Throughput and delay scaling of content-centric ad HOC and heterogeneous wireless networks," *IEEE/ACM Trans. Netw.*, vol. 25, no. 5, pp. 3030–3043, Oct. 2017, doi: 10.1109/TNET.2017.2718021.

[3] N. Kato *et al.*, "The deep learning vision for heterogeneous network traffic control: Proposal, challenges, and future perspective," *IEEE Wireless Commun.*, vol. 24, no. 3, pp. 146–153, Jun. 2017, doi: 10.1109/MWC.2016.1600317WC.

[4] K. Vasudeva, M. Simsek, D. López-Pérez, and I. Güvenç, "Analysis of handover failures in heterogeneous networks with fading," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 6060–6074, Jul. 2017, doi: 10.1109/TVT.2016.2640310.

[5] E. Lee, C. Choi, and P. Kim, "Intelligent handover scheme for drone using fuzzy inference systems," *IEEE Access*, vol. 5, pp. 13712–13719, 2017, doi: 10.1109/ACCESS.2017.2724067.

[6] S. Wang, C. Fan, C.-H. Hsu, Q. Sun, and F. Yang, "A vertical handoff method via self-selection decision tree for Internet of vehicles," *IEEE Syst. J.*, vol. 10, no. 3, pp. 1183–1192, Sep. 2016, doi: 10.1109/JSYST.2014.2306210.

[7] M. Polese, M. Giordani, M. Mezzavilla, S. Rangan, and M. Zorzi, "Improved handover through dual connectivity in 5G mmWave mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 2069–2084, Sep. 2017, doi: 10.1109/JSAC.2017.2720338.

[8] Y. Xu, R. Q. Hu, L. Wei, and G. Wu, "QoE-aware mobile association and resource allocation over wireless heterogeneous networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2014, pp. 4695–4701, doi: 10.1109/GLOCOM.2014.7037549.

[9] I. V. S. Brito and G. B. Figueiredo, "Improving QoS and QoE through seamless handoff in software-defined IEEE 802.11 mesh networks," *IEEE Commun. Lett.*, vol. 21, no. 11, pp. 2484–2487, Nov. 2017, doi: 10.1109/LCOMM.2017.2735958.

[10] L. Yala, P. A. Frangoudis, and A. Ksentini, "QoE-aware computing resource allocation for CDN-as-a-service provision," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–6, doi: 10.1109/GLOCOM.2016.7842182.

[11] A. B. Zineb, M. Ayadi, and S. Tabbane, "QoE-fuzzy VHO approach for heterogeneous wireless networks (HWNs)," in *Proc. IEEE 30th Int. Conf. Adv. Inf. Netw. Appl. (AINA)*, Mar. 2016, pp. 949–956, doi: 10.1109/AINA.2016.35.

[12] A. Ksentini, T. Taleb, and K. B. Letaif, "QoE-based flow admission control in small cell networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2474–2483, Apr. 2016, doi: 10.1109/TWC.2015.2504450.

[13] A. Farrokhi and O. Ercetin, "QoE based random sleep-awake scheduling in heterogeneous cellular networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2016, pp. 1–6, doi: 10.1109/WCNC.2016.7565172.

[14] Z. Deng, Y. Liu, J. Liu, X. Zhou, and S. Ci, "QoE-oriented rate allocation for multipath high-definition video streaming over heterogeneous wireless access networks," *IEEE Syst. J.*, vol. 11, no. 4, pp. 2524–2535, Dec. 2017, doi: 10.1109/JSYST.2015.2430893.

[15] Z. Zhang and Y. Zhang, "Layered admission control algorithms with QoE in heterogeneous network," *Ad Hoc Netw.*, vol. 58, pp. 179–190, Apr. 2017, doi: 10.1016/j.adhoc.2016.07.003.

[16] Q. Wu, Z. Du, P. Yang, Y.-D. Yao, and J. Wang, "Traffic-aware online network selection in heterogeneous wireless networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 1, pp. 381–397, Jan. 2016, doi: 10.1109/TVT.2015.2394431.

[17] Z. Du, Q. Wu, P. Yang, Y. Xu, J. Wang, and Y.-D. Yao, "Exploiting user demand diversity in heterogeneous wireless networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4142–4155, Aug. 2015, doi: 10.1109/TWC.2015.2417155.

[18] J. Lee *et al.*, "LTE-advanced in 3GPP Rel-13/14: An evolution toward 5G," *IEEE Commun. Mag.*, vol. 54, no. 3, pp. 36–42, Mar. 2016, doi: 10.1109/MCOM.2016.7432169.

[19] A. Javed, H. Larijani, A. Ahmadinia, R. Emmanuel, M. Mannion, and D. Gibson, "Design and implementation of a cloud enabled random neural network-based decentralized smart controller with intelligent sensor nodes for HVAC," *IEEE Internet Things J.*, vol. 4, no. 2, pp. 393–403, Apr. 2017, doi: 10.1109/JIOT.2016.2627403.

[20] X. Fu, S. Li, M. Fairbank, D. C. Wunsch, and E. Alonso, "Training recurrent neural networks with the levenberg–marquardt algorithm for optimal control of a grid-connected converter," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 9, pp. 1900–1912, Sep. 2015, doi: 10.1109/TNNLS.2014.2361267.

[21] Z. Gao, B. Wen, L. Huang, C. Chen, and Z. Su, "Q-learning-based power control for LTE enterprise femtocell networks," *IEEE Syst. J.*, vol. 11, no. 4, pp. 2699–2707, Dec. 2016, doi: 10.1109/JSYST.2016.2535461.

[22] M. Srinivasan, V. J. Kotagi, and C. S. R. Murthy, "A Q-learning framework for user QoE enhanced self-organizing spectrally efficient network using a novel inter-operator proximal spectrum sharing," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 11, pp. 2887–2901, Nov. 2016, doi: 10.1109/JSAC.2016.2614952.

[23] C. J. C. H. Watkins and P. Dayan, "Technical note: Q-learning," *Mach. Learn.*, vol. 8, nos. 3–4, pp. 279–292, 1992, doi: 10.1023/A:1022676722315.

[24] Y.-H. Chen, C.-J. Chang, and C. Y. Huang, "Fuzzy Q-learning admission control for WCDMA/WLAN heterogeneous networks with multimedia traffic," *IEEE Trans. Mobile Comput.*, vol. 8, no. 11, pp. 1469–1479, Nov. 2009, doi: 10.1109/TMC.2009.65.

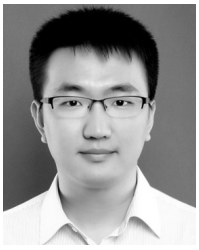
[25] S. Doltsinis, P. Ferreira, and N. Lohse, "An MDP model-based reinforcement learning approach for production station ramp-up optimization: Q-learning analysis," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 44, no. 9, pp. 1125–1138, Sep. 2014, doi: 10.1109/TSMC.2013.2294155.

[26] S. Singh, T. Jaakkola, M. L. Littman, and C. Szepesvári, "Convergence results for single-step on-policy reinforcement-learning algorithms," *Mach. Learn.*, vol. 38, no. 3, pp. 287–308, 2000, doi: 10.1023/A:1007678930559.

[27] F. Fischer and H. Mandl, "Knowledge convergence in computer-supported collaborative learning: The role of external representation tools," *J. Learn. Sci.*, vol. 14, no. 3, pp. 405–441, 2005, doi: 10.1207/s15327809jls1403_3.

[28] H. P. Young, "The evolution of conventions," *Econometrica*, vol. 61, no. 1, pp. 57–84, 1993, doi: 10.2307/2951778.

JIAMEI CHEN received the Ph.D. degree in information and communication engineering from the Harbin Institute of Technology, Harbin, China, in 2015. She was a Visiting Scholar with Purdue University from 2011 to 2012. She is currently an Assistant Professor with the College of Electrical and Information Engineering, Shenyang Aerospace University. Her research interests are heterogeneous wireless networks, resource allocation, and green communication.



YAO WANG received the Ph.D. degree in information and communication engineering from the Harbin Institute of Technology, Harbin, China, in 2015. He is currently an Assistant Professor with the Department of Air Defense Forces, Noncommissioned Officer Academy, Institute of Army Artillery and Air Defense Forces. His major research interests are spectrum resource management and power control in cognitive radio and spread spectrum communication technology.

YUFENG LI received the Ph.D. degree from the Changchun Institute of Optics, Fine Mehcanics and Physics, Chinese Academy of Sciences, China, in 2006. He was a Visiting Scholar with the Beijing Institute of Technology from 2009 to 2010. He is currently a Professor with the College of Electrical and Information Engineering, Shenyang Aerospace University. His research interests are image compression and transmission technology and wireless communication theory.

ERSHEN WANG received the Ph.D. degree from Dalian Maritime University, China, in 2009. He is currently an Associate Professor with the College of Electrical and Information Engineering, Shenyang Aerospace University. His research interest is the satellite navigation and positioning technology.