Network Selection Algorithm Based on Spectral Bandwidth Mapping and an Economic Model in WLAN&LTE heterogeneous networks

Pan Su [1,2], Zhou Weiwei [1], Gu Qingqing [1] and Ye Qiang [3]
[1] College of Telecommunications and Information Engineering
Nanjing University of Posts and Telecommunications, Nanjing 210003, China
[2] National Mobile Communications Research Laboratory
Southeast University, Nanjing 210096, China
[e-mail: supan@njupt.edu.cn]
[3] Electrical and Computer Engineering
University of Waterloo, Waterloo, ON, Canada, N2L 3G1
[e-mail: q6ye@uwaterloo.ca]
*Corresponding author: Pan Su

Abstract

Future wireless network aims to integrate different radio access networks (RANs) to provide a seamless access and service continuity. In this paper, a new resource denotation method is proposed in the WLAN and LTE heterogeneous networks based on a concept of spectral bandwidth mapping. This method simplifies the denotation of system resources and makes it possible to calculate system residual capacity, upon which an economic model-based network selection algorithm is designed in both under-loaded and over-loaded scenarios in the heterogeneous networks. The simulation results show that this algorithm achieves better performance than the utility function-based access selection (UFAS) method proposed in [12] in increasing system capacity and system revenue, achieving load balancing and reducing the new call blocking probability in the heterogeneous networks.

1. Introduction

With users' demands on telecommunication experience getting more diversified and personalized, it has posed a tough challenge to already overburdened existing networks and service providers. Consequently, the Beyond 3rd Generation (B3G) or 4th Generation (4G) wireless networks aim to integrate different radio access technologies (RATs) to provide a cooperative heterogeneous communication scenario [1]-[5]. This scenario where users can select the 'suitable' RAT can maximize the revenues of users and networks in some senses. The network selection mechanism, which is essentially used to allocate long term resources in the network selection and vertical handoff control stage, is thereby widely discussed in open literatures.

A network selection method based on Semi-Markov decision process is proposed in order to maximize the global revenue of the network in the whole network operational period [6]. In [7][8], a non-cooperative game theoretic framework for network selection is proposed in the CDMA, WLAN and WMAN integrated networks. Given the amount of bandwidth offered by other two players to a new call, this non-cooperative game theory based method aims to calculate the amount of bandwidth offered by the third player, which only maximizes the profit for this player. In [9], the author suggests two types of services: single-network and multi-homing services. The former refers to that an MT is assigned to the best available wireless network, and the latter means that an MT utilizes all available wireless access networks simultaneously. [10]-[12] present a policy or utility enabled network selection method in heterogeneous networks. This scheme establishes some suitable functions through the weighted combination of different QoS attributes for each candidate network. This network selection mechanism is designed based on the comparison of policy values (or utility values) calculated for each candidate network. In [13], the author puts forward a deferred acceptance algorithm to achieve optimum match between RATs and users in voice communication and a series of heuristic algorithms to obtain maximum throughput in video communication.

However, all of the researches above haven't paid enough attention to the denotation of resource granularity in heterogeneous networks which may cause the difficulty of resource comparison and managements between heterogeneous networks. For instance, the algorithms in [7] and [8] simply assume the bandwidth as transmission rate in the system. But the transmission rate in the physical layer cannot be sliced in the call admission control level to distribute among users in different networks. Because of the different multiple access technologies in the heterogeneous networks, there is no uniform denotation of resources in network selection. For example, WLAN, which is a statistical time division multiplexing system, seems to be difficult to allocate the spectral bandwidth directly to users since every call in the system occupies the whole system bandwidth when it has chance to transmit the data. While in OFDMA systems, time-frequency blocks are used to schedule resources. However, the capacity of time-frequency blocks dynamically changes under the network environment and also cannot be available for joint management with other networks. This observation motivates our work. Our goal is to obtain the denotation of resources practically that can be compared and managed in multiple networks, and on this basis we can design a practical rather than heuristic network selection algorithm.

Our contributions include two parts. First, we propose an equivalent bandwidth mapping concept to uniformly express and quantify resources in heterogeneous networks. The term

"equivalent bandwidth" or "effective bandwidth" has arised originally in the researches to describe the network resources occupied by users in ATM networks. In those researches, the fluctuation of the resources occupation caused by the bursty nature of users traffics is modeled by statistical methods and the occupied resource is represented by a certain statistical characteristic which is named as "effective bandwidth". Elwalid and D. Mitra etc. extended this method into wireless networks to describe the wireless resources occupation and design the "effective bandwidth"-based call admission control schemes [14][15][16]. The essential purpose of this "effective bandwidth" method is to solve the problem that the required resources of the call are uncertain for the bursty nature of the traffic whereas the network needs to know the resources required by a new incoming call before it admits the call. However, we have different objectives, i.e. to find a algorithm to uniformly quantify the occupied resources which are variedly presented when the call is in different wireless networks. The algorithm is based on the following fact: The only transmission medium for wireless communication networks is electromagnetic wave. The diverse resource forms, statistical time slots in WLAN and time-frequency blocks in LTE network, can be seen as the different partition of the electromagnetic spectrum in different orthogonal spaces, and can be somehow mapped into spectral bandwidth.

Second, based on the proposed concept of equivalent bandwidth, an economic model is then established as a revenue function to evaluate the revenue gain when admitting a new call to each RAT. Subsequently, upon the comparison of received revenue gain between two RATs, a call network selection algorithm is designed. Unlike most researches only take under-loaded scenario into consideration [10][11], we also take care of over-loaded scenario. By incorporating the balancing index into the revenue function, the algorithm could perform load balancing efficiently between two RATs in the heterogeneous networks.

The rest of the paper is organized as follows: The new resource allocation scheme in WLAN and LTE system is proposed in Section 2. In Section 3, a network selection algorithm is presented based on Section 2. In Section 4, parameter Settings and Simulation results are implemented. Section 5 concludes the paper.

2. System Model

Assuming that MTs are equipped with different RF modules of WLAN and LTE radio interfaces so as to connect with both networks technically. The coordination functionality in the two networks has to implement an effective network selection algorithm to maximize the system revenue while realizing load balancing between the integrated networks. The system revenue is closely related to the resources assigned to the calls. However, there is no uniform denotation of resources which we need to define the system revenue function in the WLAN and LTE heterogeneous networks. Therefore, in this model, we firstly manage to use the concept of equivalent bandwidth as a uniform denotation of resources in the heterogeneous system, upon which the proposed network selection algorithm is then designed to achieve more revenue, lower blocking probability and better load balancing compared with other network selection algorithms.

We consider the heterogeneous networks supporting several classes of multi-media traffic (denoted as Φ_i, for $i = 1, 2, \cdots, S$.). For a call of class Φ_i, the activity factor is denoted as η_i and when the call is at the ON state, it can generate packets with a certain bit rate B_i. Assuming each packet contains P_s bit, then every call will generate packets with different rate

$$\lambda_i = \frac{B_i \eta_i T}{P_s} \quad (T \text{ is the time slot duration}).$$

2.1 Equivalent Bandwidth Formulation in WLAN

According to IEEE802.11 standard, if a call wants to transmit data, it must detect channels free for a while. When the channel is detected busy, the call will delay the transmission until the channel is idle for duration time of a DCF interframe space (DIFS). If the channel is detected idle for the time of DIFS, the call will start a counter. The size of the counter is random derived from $[0, CW-1]$, while CW represents the contention window length whose value is chose from $[CW_{min}, CW_{max}]$. The counter decreases by one when idle time slot is detected. Otherwise, it will stop counting. When the counter is decreased to 0, the user can send packets in the next coming time slot. If the receiver receives the packets successfully, it will send ACK after a short interframe space (SIFS) to the sender whose contention window length will be reset to CW_{min}. If the sender doesn't receive ACK for a duration time of 'ACK time out', which means a collision happened, the CW will be doubled until reaching CW_{max}. Then the sender will reschedule the transmission by randomly choosing a counter in $[0, CW-1]$. When the retransmission limit is reached, the packet will be discarded.

A packet that one user generates may experience channel detection, waiting time before transmitting and probable collision. The time is defined as the equivalent channel occupation time. During the equivalent channel occupation time, these following situations may happen [17]:

(1) The network is free, and the customer is detecting channels

(2) The network is busy, and other customers have sending packets successfully

(3) The network is busy, and some packets crushed

(4) The network is busy, and the customer send its packets successfully

We analyze the time of the above situations in multi-service scenario:

(1)Average waiting time

Let $CW_{i,min}$, i_r respectively represents the minimum contention window of service class Φ_i and the maximum retry limit of all service classes, the maximum contention window of service class Φ_i : $CW_{i,max} = 2^{i_r} CW_{i,min}$.

Then the contention window $CW_i(k)$ of a service class Φ_i call when it is trying the k_{th} transmission is:

$$CW_i(k) = \min(CW_{i,max}, 2^{k-1} CW_{i,min}) \qquad k = 1, 2, \cdots i_r + 1 \tag{1}$$

Therefore, the total waiting time before transmitting successfully is:

$$\sum_{n=1}^{k} \frac{CW_i(n) - 1}{2} \tag{2}$$

The probability of sending successfully until trying k times is $p_i^{k-1}(1 - p_i)$, where p_i represents the probability of failure-sending. Therefore, we get the average waiting time:

$$\overline{W_i} = \sum_{k=1}^{i_r+1} p_i^{k-1}(1 - p_i) \sum_{n=1}^{k} \frac{CW_i(n) - 1}{2}, \quad i = 1, 2, \cdots, S \tag{3}$$

And

$$p_i = 1 - (1 - \frac{\lambda_i}{1 - p_i})^{N_i - 1} \prod_{n=1,n\neq i}^{S} (1 - \frac{\lambda_n}{1 - p_n})^{N_n} \quad , \qquad i = 1, 2, \cdots, S \tag{4}$$

where $(1 - \frac{\lambda_i}{1 - p_i})^{N_i - 1}$ represents no other calls in this class Φ_i attempt to send a packet, and

$\prod_{n=1,n\neq i}^{S} (1 - \frac{\lambda_n}{1 - p_n})^{N_n}$ represents all calls in other classes also do not attempt to send a packet.

The derivation process of this equation can be seen in appendix A.

(2)Average sending time of other customers

For class Φ_i, the average number of packets generated by each call within time T_i is $\lambda_i T_i$, if the system is at equilibrium state, the number of packets generated by each call should statistically be transmitted successfully, the transmission time needed is $\lambda_i T_i T_s'$. T_s' is the time duration when the channel is sensed busy because of a successful transmission, which includes DIFS, SIFS , transmission time of ACK (T_{ACK}) and a data packet transmission time in physical layer T_s. Where, $T_s = P_s / W$, W is the transmission bit rate supported by the physical layer. The total transmission time of all calls in the system except the investigated one is:

$$(N_i - 1)\lambda_i T_i T_s' + T_i T_s' \sum_{n=1,n\neq i}^{S} N_n \lambda_n \tag{5}$$

Given that the values of SIFS, DIFS and the T_{ACK} are all constant, we have $T_s' = T_s + SIFS + DIFS + T_{ACK} = \beta T_s$, β is a constant.

(3)Average collision time

Before a packet is transmitted successfully, it may undergo several possible collisions. For the invested a call in service class Φ_i, the packet will go through $\sum_{k=1}^{i_r+1} (k-1) p_i^{k-1}(1 - p_i)$ retries before being sent successfully. So its average collision time is:

$$\overline{T_{ci}} = T_f \sum_{k=1}^{i_r+1} (k-1) p_i^{k-1}(1 - p_i) \tag{6}$$

where T_f is the time duration when the channel is sensed busy due to failed transmissions. T_f includes transmission time of a packet, DIFS and the time of 'ACK timeout'. We get: $T_f = T_s + ACK_{time-out} + DIFS = \gamma T_s$.

Therefore, the total collision-occupation time of all calls in WLAN system can be expressed as:

$$\frac{1}{2}[(1 + (N_i - 1)\lambda_i T_i)\overline{T_{c_i}} + T_i \sum_{n=1,n\neq i}^{S} N_n \lambda_n \overline{T_{c_n}}] \tag{7}$$

From the above, we can get the equivalent channel occupation time T_i to transmit a packet

successfully, that is:

$$T_i = (N_i - 1)\lambda_i T_i T_s^{'} + T_i T_s^{'} \sum_{n=1,n\neq i}^{S} N_n \lambda_n + \frac{1}{2}[(1 + (N_i - 1)\lambda_i T_i)\overline{T_{c_i}}$$

$$+ T_i \sum_{n=1,n\neq i}^{S} N_n \lambda_n \overline{T_{c_n}}] + T_s^{'} + \overline{W_i} \tag{8}$$

To keep the system steady, we have to make sure that: $\lambda_i T_i < 1$, then we get :

$$\lambda_i T_i = \frac{\lambda_i(2T_S^{'} + 2\overline{W_i} + \overline{T_{c_i}})}{2 + \lambda_i(2T_S^{'} + \overline{T_{c_i}}) - \sum_{n=1}^{S} N_n \lambda_n (2T_S^{'} + \overline{T_{c_n}})} < 1 \tag{9}$$

This means:

$$\sum_{n=1}^{S} N_n \lambda_n (2T_S^{'} + \overline{T_{c_n}}) < 2 - 2\lambda_i \overline{W_i} \tag{10}$$

Since $\lambda_i \overline{W_i} \square 1$, the above inequation can be simplified as:

$$\sum_{n=1}^{S} N_n \lambda_n (2T_S^{'} + \overline{T_{c_n}}) < 2 \tag{11}$$

With the $\overline{T_{c_i}}$ and $T_S^{'}$ formulations, we can finally get:

$$\sum_{n=1}^{S} N_n \lambda_n (\beta + \gamma \frac{p_n}{2 - 2p_n}) < \frac{1}{T_S} \tag{12}$$

We define:

$$\partial_n = \lambda_n (\beta + \gamma \frac{p_n}{2 - 2p_n}) \ , \quad n = 1, 2, \cdots, S \tag{13}$$

as the Equivalent Bandwidth of a call in class Φ_n in WLAN system, then our final inequation becomes $\sum_{n=1}^{S} N_n \partial_n < \frac{1}{T_S}$. Obviously, the left formulation $\sum_{n=1}^{S} N_n \partial_n$ is the total spectrum occupied by all users and the right side $1/T_S = w / P_s$ is a constant weighted spectrum of the system, which represents the total system bandwidth.

Equation (12) means that the system is stable only when the equivalent bandwidth occupied by all calls in the system is less than the system bandwidth.

2.2 Equivalent Bandwidth Formulation in LTE

The OFDMA technology used in the LTE system with total amount of C sub-carriers is considered in our scenario. Each call in the system has its required transmission power respectively to maintain different levels of QoS satisfaction. According to [19][20], the achievable rate on sub-carrier j allocated to the n th user in service class of Φ_i can be defined as:

74

Pan et al.: Network Selection Algorithm Based on Spectral Bandwidth Mapping
and an Economic Model in WLAN&LTE heterogeneous networks

$$c_{ijn} = W \log_2 (1 + a \cdot \frac{G_{ijn} p_{ijn}}{\sigma^2}) \tag{14}$$

, where $a \approx -1.5 / \log(5 * BER)$, BER is the required bit error rate for certain class of service, G_{ijn} denotes the channel gain on sub-carrier j allocated to n th user in service class Φ_i, and G_{ijn} is the normally distributed random variable, p_{ijn} denotes the power allocated to the n th user at sub-carrier j in service class Φ_i, W is the bandwidth of single sub-carrier, σ^2 is the thermal noise power.

From equation (14), the user's data rate in service class $\Phi_i (i = 1, 2, ..., S)$ is:

$$R_{in} = C_{in} W \log_2 (1 + a_i \cdot \frac{\overline{G_{in}} \cdot \overline{p_{in}} \cdot \eta_i}{\sigma^2 \cdot C_{in}}) \quad (i = 1, 2, ..., S; n = 1, 2, ... N_i) \tag{15}$$

Where, R_{in} is the average data rate of n th user $(n = 1, 2, ... N_i)$ received by service class Φ_i, N_i denotes the number of activated users in service class Φ_i, C_{in} is the number of allocated sub-carriers to the n th user in service class Φ_i, $\overline{p_{in}}$ denotes the average power allocated to the n th user, η_i is the activity factor for service Φ_i, we use $\overline{G_{in}}$ [20] as the average channel gain for user n in service.

After truncating appropriately Taylor series expansion for equation (15), R_{in} is reformulated as:

$$R_{in} \approx C_{in} W (\frac{1}{\ln 2} \cdot \frac{a_i \overline{G_{in} p_{in}} \eta_i}{\sigma^2 C_{in}} - \frac{1}{2 \ln 2} \cdot \frac{a_i^2 \overline{G_{in}}^2 \overline{p_{in}}^2 \eta_i^2}{\sigma^4 C_{in}^2}) \quad (i = 1, 2, ..., S; n = 1, 2, ... N_i) \tag{16}$$

Given B_i as the required bit rate for a class Φ_i call, which is determined by the QoS profile of the service class, the number of sub-carriers needed for the bit rate B_i can be obtained after some algebraic operations:

$$C_{in} \approx \frac{W a_i^2 \overline{G_{in}}^2 \overline{p_{in}}^2 \eta_i^2}{2\sigma^2 W a_i \overline{G_{in} p_{in}} \eta_i^2 - 2 B_i \sigma^4 \ln 2} \tag{17}$$

The equivalent bandwidth ∂_{in} is defined by multiplying C_{in} with single sub-carrier bandwidth W:

$$\partial_{in} = C_{in} W \approx \frac{W^2 a_i^2 \overline{G_{in}}^2 \overline{p_{in}}^2 \eta_i^2}{2\sigma^2 W a_i \overline{G_{in} p_{in}} \eta_i^2 - 2 B_i \sigma^4 \ln 2} \tag{18}$$

∂_{in} denotes the equivalent spectral bandwidth occupied by a call of class Φ_i to satisfy its QoS metrics (including BER and bit rate requirements) under certain channel conditions.

As we know in the OFDMA system, the number of allocated sub-carriers to all ongoing calls cannot outnumber the total number of system sub-carriers. This can be formulated algebraically as follows:

$$\sum_{i=1}^{k}\sum_{n=1}^{N_i} C_{in} \leq C \tag{19}$$

Multiplying the inequality (19) by single sub-carrier bandwidth W yields:

$$\sum_{i=1}^{k}\sum_{n=1}^{N_i} WC_{in} = \sum_{i=1}^{k}\sum_{n=1}^{N_i} \partial_{in} \leq CW = W_0 \tag{20}$$

where W_0 is total bandwidth in the LTE system.

Equation (20) as well as equation (12) expresses the equivalent spectral bandwidth restriction of LTE and WLAN systems: The prerequisite for guaranteeing the QoS of calls and stability of the systems is that the equivalent bandwidth occupied by all calls in the system should be less than the system bandwidth. Equation (20) and equation (12) convert different resource allocation schemes in LTE and WLAN systems into a quasi Frequency Division Multiple Access (FDMA) mode where each call occupies a part of systems spectral bandwidth. This equivalent spectral bandwidth can be used to uniformly define the revenue function for both systems.

It can be seen from (20) that the less ∂_{in} is, the more calls can be admitted. To find the smallest ∂_{in}, let $\dfrac{\partial\left(\overline{\partial_{in}}\right)}{\partial \overline{p_{in}}} = 0$, it follows:

$$\overline{p_{in\,opt}} = \frac{2B_i\sigma^2 \ln 2}{\eta_i W a_i \overline{G_{in}}} \tag{21}$$

Substituting $\overline{p_{in\,opt}}$ into equation (18), we get the optimal equivalent bandwidth for a certain class of service:

$$\partial_i = 2B_i \ln 2 \tag{22}$$

From the deduction of equation (13) and (22), the physical layer resources, i.e. statistical time slots in WLAN, power and sub-carriers in OFDMA, occupied by calls according to their QoS requirements have been mapped into equivalent bandwidth. We can thereby define the revenue function depended on it in the next section and design the revenue-based network selection algorithm.

3. Proposed Network selection Algorithm

In this section, we propose a network selection algorithm based on a designated revenue function in the wireless heterogeneous networks. The revenue function evaluates the revenue receive by a network when it allocates bandwidth to calls (i.e. admit calls). Then based on the results, the network selection algorithm chooses the most suitable network for the new call. It was shown in [18] that the adopted revenue function should have the characteristics of continuous, strictly concave and twice differentiable [12]. Thus, we manage to use a kind of Bell function as the revenue function, which has got the above characteristics, to denote the corresponding revenue received when allocating bandwidth to the incoming call [12]:

$$B(x,\partial_i) = \exp\left[-\frac{(x-\partial_i)^2}{2\partial_i^2}\right] \tag{23}$$

Where x denotes the bandwidth resource allocated to a new call, ∂_i is the equivalent bandwidth requirement of service class Φ_i, $B(x,\partial_i)$ denotes the received revenue when allocating bandwidth x to the new call. **Fig. 1** is depicted below to show the main characteristics of the revenue function:
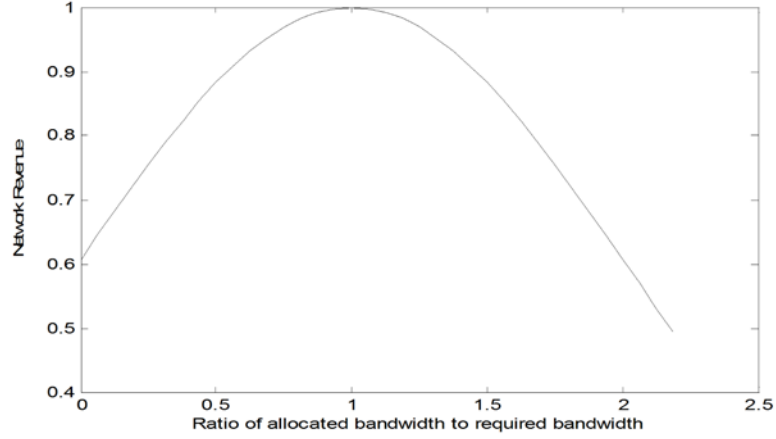
Fig. 1. Network Revenue

From **Fig. 1** we can see that the received revenue grows with the increase of bandwidth allocation to a new call, which reaches the highest value when the allocated bandwidth equals the bandwidth requirement. Therefore, this function, as a reference revenue function, is proved to be rational to reflect the difference between the allocated equivalent bandwidth and the equivalent bandwidth requirement.

Additionally, in order to perform the load balancing issue between the coupled integrated networks, we define a load balancing index to maintain certain level of traffic balance in the system. The index for RAT j is considered as follows:

$$\theta_j = \exp\left(-\frac{\sum_{i=1}^{M}\sum_{n=1}^{n_{ji}} x_{j_{in}}}{W_j}\right) \tag{24}$$

where n_{ji} is the current existed number of class Φ_i users in RAT j, $x_{j_{in}}$ denotes the bandwidth allocated to each user of class Φ_i and W_j is the system bandwidth of RAT j. It is obviously that the balancing index θ_j decreases gradually as the system load is getting heavy. By introducing the balancing index, a new revenue function with load balancing for each RAT is established as follows:

$$U_j\left(x\right)=\exp\left(-\frac{\sum_{i=1}^{M}\sum_{n=1}^{n_{j_i}}x_{j_{in}}}{W_j}\right)\exp\left[-\frac{\left(x_{j_i}-\partial_i\right)^2}{2\partial_i^2}\right]\tag{25}$$

Thus, the revenue gain for each RAT when admitting a new incoming call can be formulated as:

$$\Delta U_j\left(x\right)=\exp\left(-\frac{\sum_{i=1}^{M}\sum_{n=1}^{n_{j_i}}x'_{j_{in}}}{W_j}\right)\left(\sum_{m\neq i,m=1}^{M}\sum_{n=1}^{n'_{j_m}}\exp\left[-\frac{\left(x'_{j_{mn}}-\partial_m\right)^2}{2\partial_m^2}\right]+\sum_{n=1}^{n'_{j_i}+1}\exp\left[-\frac{\left(x'_{j_{in}}-\partial_i\right)^2}{2\partial_i^2}\right]\right.$$

$$\left.-\sum_{i=1}^{M}\sum_{n=1}^{n_{j_i}}\exp\left[-\frac{\left(x_{j_{in}}-\partial_i\right)^2}{2\partial_i^2}\right]\right)\tag{26}$$

n'_{j_i} means the changed number of class Φ_i users after admitting the new call in RAT j. $x'_{j_{in}}$ denotes the bandwidth allocation for every call after admitting the new call in RAT j. $x_{j_{in}}$ denotes previously allocated bandwidth for every ongoing call before admitting the new call in RAT j. If RAT j is over-loaded which means there remains insufficient residual bandwidth for the incoming call, every ongoing call is required to make their effort to contribute some of their bandwidth to the incoming call according to our proposed network selection algorithm.

Based on the above analysis, we can design the network selection algorithm in the heterogeneous networks under the following conditions:

3.1 Under-loaded Condition

(1). Algorithm Description

In this condition, the residual system bandwidth resources calculated based on equivalent bandwidth in either RAT are sufficient to accept a new incoming call. Therefore, our algorithm states that this incoming call is expected to be admitted in one of the RATs based on the comparison between the revenue gains of RATs and the call will be allocated with its requested amount of equivalent bandwidth. The revenue gain of per RAT after accepting this new call can be calculated by the equation (26).

(2). Core Algorithm Implementation

Let ΔU_W and ΔU_L denote the revenue gain for WLAN and LTE network respectively if a new call is accepted in its network. Then the network selection algorithm for under-loaded scenario is proposed as follows:

If $\Delta U_W > \Delta U_L$

　　Select the WLAN network

Else if $\Delta U_W < \Delta U_L$

　　Select the LTE network

Else

　　Select the WLAN network (it is the RAT with higher priority)

End

3.2 Over-loaded Condition

(1). Algorithm Description

When the system residual bandwidth resources are not enough to provide for one new call, our proposed algorithm indicates that all of the ongoing calls have to contribute some of their equivalent bandwidths to the incoming call according to certain method in order to accept this call under bearable degradation of each user's QoS. In other words, the system bandwidth resources are facing the redistribution and the assigned bandwidth for every ongoing call is varied after admitting the new user.

Our algorithm designs that the bandwidth allocation for all calls including the incoming call is updated according to the weights of their bandwidth requirements, upon which the total bandwidth assignment is renewed and every call receives its weighed linear division of total system bandwidth respectively after admitting the new call. Specifically, the total system bandwidth is redistributed algebraically as follows:

$$ x_{j_{in}}^{'} = \frac{W_j \partial_i}{\partial_i + \sum\limits_{m=1}^{M} n_{j_m} \partial_m} \tag{27} $$

Where, $\partial_i + \sum\limits_{m=1}^{M} n_{j_m} \partial_m$ represents the total bandwidth needed by all users in the network j to satisfy the Qos after admitting a new user of class Φ_i, and W_j denotes the total bandwidth of the network j. This mechanism could increase the system capacity of the heterogeneous networks, simultaneously making every effort to control each call's QoS degradation under the acceptable level.

With the increasing number of admitted calls in the over-loaded condition, the system revenue gain and users' QoS level in each RAT decrease gradually. Therefore, two important adjustable thresholds E_W and E_L can be defined as the revenue gain lower bound for WLAN and LTE respectively in order to control the QoS degradation level for certain class of service. Since different services have diverse QoS requirements, their tolerance of revenue degradation varies. Therefore, E_W and E_L are defined based on the strictest QoS requirement.

Then, our core network selection algorithm for over-loaded scenario is proposed as follows (for one class of service):

(2). Core Algorithm Implementation:

Step 1: Initialize: $N_j = 0$ $(j = WLAN, LTE)$

Step 2: If $N_j == 0$ $(j = WLAN, LTE)$

> For $n = 1 : n_{j_i}$
>
> > $x_{j_{in}} = x_{j_{in}}^{'}$;
>
> End
>
> For $n = 1 : n_{j_i} + 1$
>
> > $$ x_{j_{in}}^{'} = \frac{W_j \partial_i}{\partial_i + \sum\limits_{m=1}^{M} n_{j_m} \partial_m} ; $$
>
> End

$$\Delta U_j(x) = \exp\left(-\frac{\sum\limits_{i=1}^{M}\sum\limits_{n=1}^{n_{j_i}} x_{j_{in}}}{W_j}\right)\left(\sum\limits_{m\neq i,m=1}^{M}\sum\limits_{n=1}^{n_{j_m}}\exp\left[-\frac{\left(x'_{j_{mn}}-\partial_m\right)^2}{2\partial_m^2}\right]+\sum\limits_{n=1}^{n_{j_i}+1}\exp\left[-\frac{\left(x'_{j_{in}}-\partial_i\right)^2}{2\partial_i^2}\right]\right.$$

$$\left.-\sum\limits_{i=1}^{M}\sum\limits_{n=1}^{n_{j_i}}\exp\left[-\frac{\left(x_{j_{in}}-\partial_i\right)^2}{2\partial_i^2}\right]\right)$$

        If $\Delta U_j < E_j \, (E_j > 0)$

          $N_j = n_j$;

        End

    End

Step 3: If $N_w \wedge = 0$ && $N_L == 0$

    Select LTE network; Go to Step 2;

   Else if $N_w == 0$ && $N_L \wedge = 0$

    Select WLAN network; Go to Step 2;

   Else if $N_w \wedge = 0$ && $N_L \wedge = 0$

    Stop;

   Else

    If $\Delta U_W > \Delta U_L$

     Select WLAN network; Go to Step 2;

    Else if $\Delta U_W < \Delta U_L$

     Select LTE network; Go to Step 2;

    Else

     Select WLAN network; Go to Step 2;

    End

   End

4. Performance Evaluation

4.1 Simulation Environment

We consider the WLAN and LTE heterogeneous scenario. These two cells are collocated together to form a kind of coupled integrated network. The Okumura-Hata model [18] is adopted as a channel propagation model to simulate the transmission loss in the micro-cell system, in which the transmitting and receiving antenna gains are normalized to 1. We assume perfect power control scheme is implemented in either RAT between base stations and MTs in order to achieve desired QoS requirements in different services.

In our system, we consider two classes of services: Voice and Data with transmission rates of 16kbps and 64kbps respectively. The other parameters are listed in **Table 1**.

Table 1. Simulation parameters

RAT j / Parameters	WLAN	OFDMA
System Bandwidth	-	5MHz
Micro-cell Radius	100m	1km
Channel Gain	-	$N(10^{-10}, 10^{-10})$
Thermal Noise Power	-	$10^{-12}\,\mathrm{W}$
Channel rate	11Mbps	100Mbps
Slot time	20μs	-
$CW_{i,min}$	16(voice), 32(data)	-
$CW_{i,max}$	$2^{i_r} CW_{i,min}$	-
Retransmission Limit(i_r)	7	-
Packet size	1023 byte	-
SIFS	10μs	-
DIFS	50μs	-
ACK	10.2μs	-
$ACK_{timeout}$	20.8488μs	-
System Sub-carriers Number	-	100
Single Sub-carrier Bandwidth	-	5kHz
Bit Error Rate(BER)	-	10^{-3} (Voice), 10^{-4} (Data)
Require Bit Rate	16kbps(Voice), 64kbps(Data)	
Service Activity Factor	0.4 (Voice), 0.25(Data)	

* $N(10^{-10}, 10^{-10})$ represents a Gaussian random variable, with mean value of 10^{-10} and variation of 10^{-10}.

4.2 Simulation Results

Fig. 2 shows that the equivalent bandwidth occupied by a call increases with the rising of system load (in terms of number of voice users, where the number of data users is fixed at 10) . The reason is that the collision probability of a call increases with the system load and the collisions lead to the rising of the equivalent channel occupation time for the call to transmit a packet successfully.
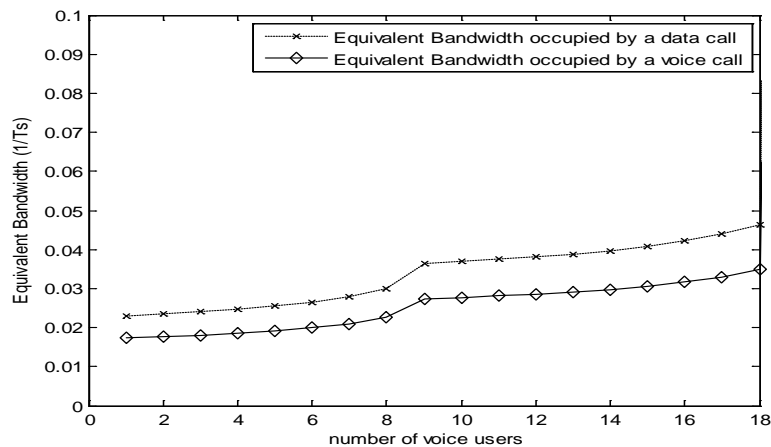
Fig. 2. occupied Equivalent Bandwidth VS number of users

Fig. 3 depicts the equivalent bandwidth required for voice calls varying with their allocated power levels in LTE system. It can be seen from the figure that there appears an optimal value of power allocation which corresponds to a minimum equivalent bandwidth requirement for voice calls in LTE. Additionally, we can also find that the optimal value of equivalent bandwidth is independent with the variation of BER requirement, which has been proved by equation (15) that the optimal equivalent bandwidth only relies on call's transmission rate. Another distinct observation is that the optimal value of allocated power gets small when BER requirement degrades, which is quite reasonable in the sense that less power allocation might cause BER degradation. The results in the figure provide a set of optimal power control and sub-carrier allocation strategy under different BER requirements.
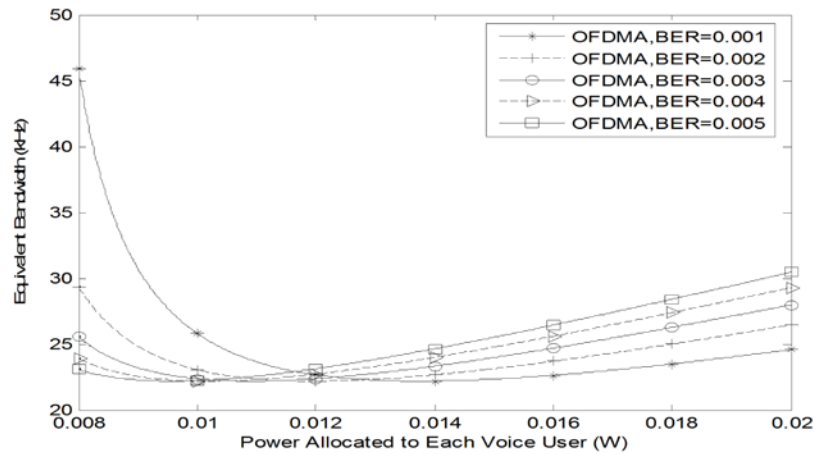
Fig. 3. Equivalent Bandwidth Requirements for Voice Users in LTE

Fig. 4 and **Fig. 5** show the load balancing efficiency between two network selection methods for both voice and data services. In the figures, for admitting either voice or data calls, our proposed method can achieve better load balancing than the utility function-based access selection (UFAS) method in [12] since our algorithm introduces a more efficient balancing index into the network selection method. By contrast, the algorithm implementation result for UFAS method indicates that when the load condition in WLAN is extremely heavy, there still appears no new voice user being admitted into LTE system.

For the collision reason, the required equivalent bandwidth for a single call in WLAN increases with the rising number of admitted calls. Thus, the consumption rate of the system spectral bandwidth in WLAN is consequently higher than that in LTE network. Therefore, it is observed in **Fig. 4** that the incremental rate of calls accessing to WLAN is lower than that of LTE, it means that the proposed algorithm is inclined to direct users accessing to LTE network when user numbers and collisions grow in the WLAN network.

Fig. 5 demonstrates similar algorithm load balancing between two methods for admitting data users into the heterogeneous system. And we can observe in the UFAS method that when all the system remaining resources in LTE are totally unoccupied by any user, there remains only 25% residual capacity available in WLAN system, which appears to be inferior to the load condition under our proposed algorithm.
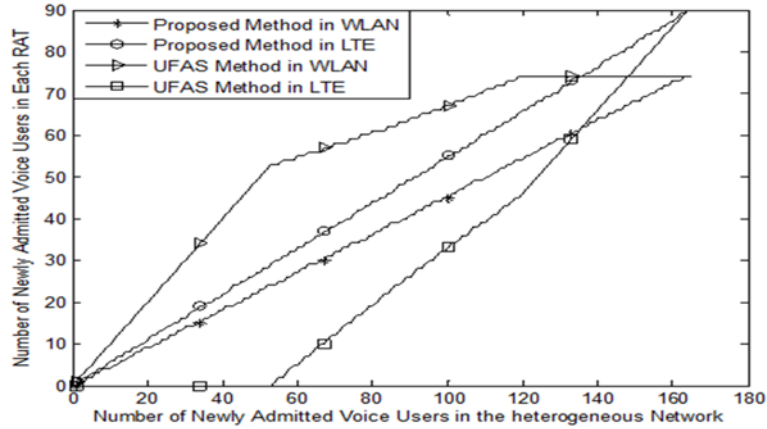
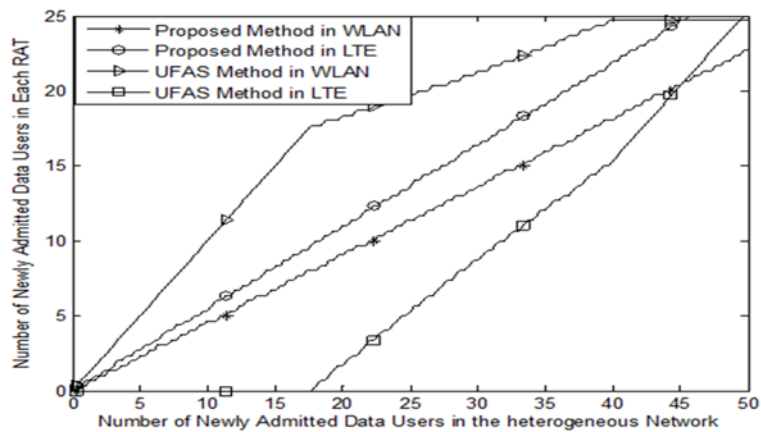**Fig. 4.** Comparison of load balancing Efficiencies for Voice Service

Fig. 5. Comparison of load balancing Efficiencies for Data Service

In **Fig. 6** and **Fig. 7**, the new call blocking probability of both voice and data users is analyzed in each RAT in the heterogeneous system. Here, we suppose that three-fifths of the system bandwidth resources have already been occupied by ongoing voice or data calls in either RAT.

As shown in the two figures, under certain level of QoS constraint, the new call blocking probability in either voice and data service increases with the growth of new call arrival rate since the increasing number of new calls arrivals leads to more system resources occupied, thereby increasing the call blocking opportunity. From the simulation, we can find the blocking probabilities for both services in either RAT in the proposed method less than those in UFAS method, which means that our proposed network selection algorithm can achieve better resource utilization.
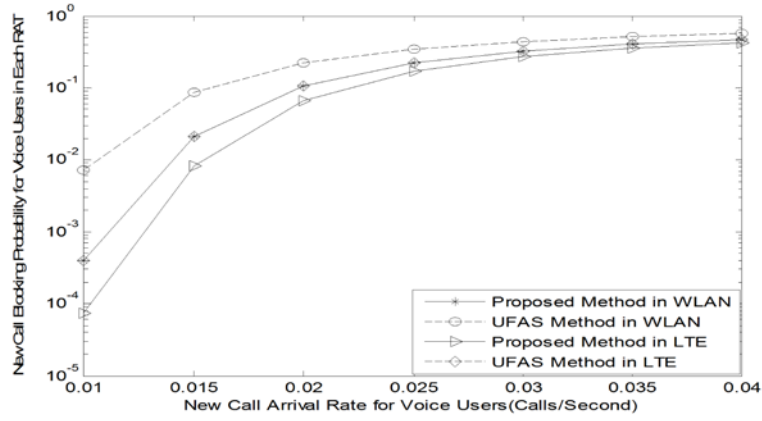
Fig. 6. New Call Blocking Probability for Voice Users
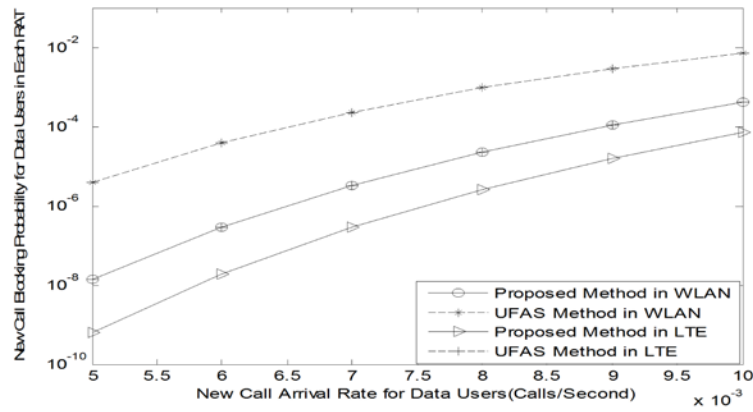

Fig. 7. New Call Blocking Probability for Data Users

5. Conclusions

In this paper, we firstly propose a new resource denotation method based on equivalent bandwidth mapping concept in order to uniform the expression of resources in WLAN and LTE heterogeneous networks. Based on this mechanism, all the QoS requirements of certain service can be mapped into the equivalent bandwidth, which makes it possible to define system revenue function related to the system resources. We subsequently design network selection algorithm based on an economic model. Simulation results show that introducing efficient load balancing index makes this algorithm achieve better load balancing performance. And by accepting small level of QoS degradation, our proposed algorithm also achieves lower new call blocking probability and larger system capacity than the utility function-based access selection (UFAS) method.

APPENDIX A

To obtain the equation (4), we define q_i as the probability of a call in class Φ_i sending a packet in a time slot. The collision occurs when at least one of the remaining calls of all service

classes also transmits in the same slot and we get the following equation:

$$p_i = 1 - (1 - q_i)^{N_i - 1} \prod_{n=1, n \neq i}^{S} (1 - q_n)^{N_n} \tag{a-1}$$

where $(1 - q_i)^{N_i - 1}$ represents no other call in this class Φ_i attempt to send a packet, and

$\prod_{n=1, n \neq i}^{S} (1 - q_n)^{N_n}$ represents all calls in other classes also do not attempt to send a packet.

Conditional on a nonempty queue of a call in class Φ_i, the transmission probability of the call can be approximated as

$$\frac{E[A_i]}{T_i} \tag{a-2}$$

where $E[A_i]$ is the average numbers of transmission attempts of a call in Φ_i made during T_i.

$$E[A_i] = \sum_{k=1}^{m_r + 1} k p_i^{k-1} (1 - p_i)^{I\{k < m_r + 1\}} = \frac{1 - p_i^{m_r}}{1 - p_i} \approx \frac{1}{1 - p_i} \tag{a-3}$$

Where m_r is the retransmission limit and its value is 7 in the IEEE 802.11standard. And $I\{A\} = 1$ when A is a positive value, on the contrary, $I\{A\} = 0$.

The probability of nonempty queue is $\lambda_i T_i$, so we have:

$$q_i = \frac{E[A_i]}{T_i} \lambda_i T_i = E[A_i] \lambda_i = \frac{\lambda_i}{1 - p_i} \tag{a-4}$$

By substituting (a-4) into (a-1), the collision probability is represented as

$$p_i = 1 - (1 - \frac{\lambda_i}{1 - p_i})^{N_i - 1} \prod_{n=1, n \neq i}^{S} (1 - \frac{\lambda_n}{1 - p_n})^{N_n} \quad , \quad i = 1, 2, \cdots, S \tag{a-5}$$

Acknowledgments

References

[1] 3rd Generation Partnership Project (3GPP). [Online]. Available: http://www.3gpp.org/
[2] 3rd Generation Partnership Project 2 (3GPP2). [Online]. Available: http://www.3gpp2.org/
[3] 3GPP Improvement of RRM across RNS and RNS/BSS (Release 5), TR 25.881(v5.0.0), Dec.2001.
[4] 3GPP, Requirements on 3GPP System to Wireless Local Area Network (WLAN) Inter-working, TS 22.234 (v8.0.0). Mar. 2007.
[5] 3GPP Feasibility study for evolved Universal Terrestrial Radio Access (UTRA) and Universal Terrestrial Radio Access Network (UTRAN) (Release 8), TR 25.912(v8.0.0), December. 2008.
[6] Yu, F.; Krishnamurthy, V., "Optimal Joint Session Admission Control in Integrated WLAN and CDMA Cellular Networks with Vertical Handoff," *Mobile Computing, IEEE Transactions on* , vol.6, no.1, pp.126-139, Jan. 2007. Article (CrossRef Link)
[7] Niyato, D.; Hossain, E., "A Noncooperative Game-Theoretic Framework for Radio Resource Management in 4G Heterogeneous Wireless Access Networks," *Mobile Computing, IEEE*

Transactions on, vol.7, no.3, pp.332-345, March 2008. Article (CrossRef Link)

[8] Niyato, D.; Hossain, E., "WLC04-5: Bandwidth Allocation in 4G Heterogeneous Wireless Access Networks: A Noncooperative Game Theoretical Approach," *IEEE Global Telecommunications*, pp.1-5, November 27,2006-December 1,2006. Article (CrossRef Link)

[9] Ismail, M.; Weihua Zhuang; Ming Yu, "Radio Resource Allocation for Single-Network and Multi-Homing Services in Heterogeneous Wireless Access Medium," in *Proc. of Vehicular Technology Conference (VTC Fall), 2012 IEEE,* vol., no., pp.1-5, 3-6 September. 2012. Article (CrossRef Link)

[10] Liu Sheng-mei; Pan Su; Mi Zheng-kun; Meng Qing-min; Xu Ming-hai, "A Simple Additive Weighting Vertical Handoff Algorithm Based on SINR and AHP for Heterogeneous Wireless Networks," in *Proc. of 2010 International Conf. on Intelligent Computation Technology and Automation*, vol.1, no., pp.347-350, 11-12 May 2010. Article (CrossRef Link)

[11] Xuebing Pei; Tao Jiang; Daiming Qu; Guangxi Zhu; Liu, Jian, "Radio-Resource Management and Access-Control Mechanism Based on a Novel Economic Model in Heterogeneous Wireless Networks," *Vehicular Technology, IEEE Transactions on* , vol.59, no.6, pp.3047,3056, July 2010. Article (CrossRef Link)

[12] Yung-Han Chen; Nuan-Yu Yang; Chung-Ju Chang; Fang-Ching Ren, "A Utility Function-based Access Selection Method for Heterogeneous WCDMA and WLAN Networks," *Personal, Indoor and Mobile Radio Communications, 2007. PIMRC 2007. IEEE 18th International Symposium on*, vol., no., pp.1-5, 3-7 Sept. 2007. Article (CrossRef Link)

[13] Yiyue Wu; Viswanathan, H.; Klein, T.; Haner, M.; Calderbank, R., "Capacity Optimization in Networks with Heterogeneous Radio Access Technologies," in *Proc. of Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, vol., no., pp.1-5, 5-9 Dec. 2011. Article (CrossRef Link)

[14] Elwalid, A.I.; Mitra, D., "Effective bandwidth of general Markovian traffic sources and admission control of high speed networks," in *Proc. of INFOCOM '93. Proceedings.Twelfth Annual Joint Conference of the IEEE Computer and Communications Societies. Networking: Foundation for the Future, IEEE*, vol.1, pp.256-265, 1993. Article (CrossRef Link)

[15] Stahlschmidt, T.; Godoy, W.; Foronda, A., "Effective bandwidth based connection admission control for stored variable bit rate video streams," in *Proc. of Communications (LATINCOM), 2010 IEEE Latin-American Conference on,* pp.1-6, 15-17 Sept. 2010. Article (CrossRef Link)

[16] Spitler, S.L.; Lee, D.C., "Integration of Explicit Effective-Bandwidth-Based QoS Routing With Best-Effort Routing," *IEEE/ACM Transactions on Networking*, vol.16, no.4, pp.957-969, Aug. 2008. Article (CrossRef Link)

[17] Cai, L.X.; Xuemin Shen; Mark, Jon W.; Lin Cai; Yang Xiao, "Voice capacity analysis of WLAN with unbalanced traffic," *IEEE Transactions on Vehicular Technology*, vol.55, no.3, pp.752-761, May 2006. Article (CrossRef Link)

[18] Wei Wang; Soung-Chang Liew; Li, V.O.-K., "Solutions to performance problems in VoIP over a 802.11 wireless LAN," *Vehicular Technology, IEEE Transactions on,* vol.54, no.1, pp.366-384, Jan. 2005. Article (CrossRef Link)

[19] Keunyoung Kim; Youngnam Han; Seong-Lyun Kim, "Joint subcarrier and power allocation in uplink OFDMA systems," *Communications Letters, IEEE,* vol.9, no.6, pp.526-528, Jun 2005. Article (CrossRef Link)

[20] Ki-Dong Lee, Victor C.M.Leung, "Capacity Planning for Group-Mobility Users in OFDMA Wireless Networks," *EUEASIP Journal on Wireless Communications and Networking*, pp.1-12, 2006. Article (CrossRef Link)

Pan Su is a Professor in the Faculty of Telecommunication Engineering of Nanjing University of Posts and Telecommunications, China. He received his Bachelor degrees of Engineering from Huazhong University of Science and Technology, China and his Master of Engineering degree, from Nanjing University of Posts and Telecommunications Nanjing, China. After graduation, he held appointment in Motorola, Inc. and obtained the Training Instructor Certification in Chicago, US. in 1995. He obtained his Ph. D. degree in electrical and electronic engineering from The University of Hong Kong, Hong Kong in 2004. Since August 2004, he has been with the Nanjing University of Posts and Telecommunications.
Professor Pan has published more than 40 papers in journals and conferences. His re search interests are in the general areas of wireless communications, particularly, wir eless resource allocation and QoS guarantee in OFDM, MIMO and heterogeneous ne tworks.
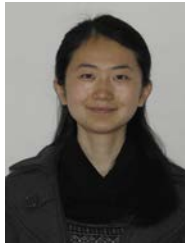
Zhou Weiwei received her B.Sc. degree in Communication Engineering from Nanjing University of Posts and Telecommunications, Nanjing, China, in 2014. She is currently working toward the Ph.D. degree in Nanjing University of Posts and Telecommunications. Her research interests include resource management, network selection and wireless resource allocation.
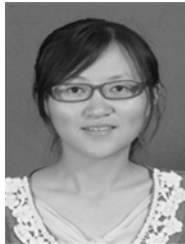
Gu Qingqing received her B.Sc. degree in Communication Engineering in 2011 and Master's degree in Communication and Information Engineering in 2014 respectively, all from Nanjing University of Posts and Telecommunications. Her research interests include radio resource management and wireless resource allocation.

Ye Qiang received his B.Sc. degree in Communication Engineering in 2009 and Master's degree in Communication and Information Engineering in 2012 respectively, all from Nanjing University of Posts and Telecommunications. He is currently working toward the Ph.D. degree at the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research interests include medium access control and performance optimization in mobile ad hoc networks.