QoS-Aware Load Balancing Algorithm for Joint Group Call Admission Control in Heterogeneous Networks

Rongrong Lian, Hui Tian, Wenchao Fei, Jie Miao and Canru Wang
Key Laboratory of Universal Wireless Communications, Ministry of Education
Wireless Technology Innovation Institute, WTI
Beijing University of Posts and Telecommunications, Beijing, P.R.China, 100876
E-mail: lianrrong@gmail.cn

***Abstract*—In heterogeneous network, it is likely that many mobile users send access requests at the same time. To avoid network congestion and provide users with good service experience, a novel joint group call admission control (JGCAC) algorithm is proposed. This algorithm includes two procedures: Firstly, the number of users allocated to each network is determined according to load balancing policy. Secondly, based on the evaluation values about network conditions and user requests, suitable users are admitted to each network by extended Hungary Algorithm. Simulation results indicate that the proposed algorithm can maintain load balancing among networks, reduce dropping rate and achieve a better user satisfaction.**

Keywords-heterogeneous network; joint call admission control; load balancing; QoS awareness

I. INTRODUCTION

Single wireless network can't fully satisfy user requests, so an important trend in the future is the integration of multiple networks, e.g., WLAN, GSM, UMTS, WiMAX and etc [1]. As an essential function of joint radio resource management (JRRM), joint call admission control (JCAC) remains a challenging problem. In heterogeneous networks, JCAC is aimed at two kinds of users, i.e., newly coming users and handover users. The JCAC algorithms have been studied considering various criteria and evaluation models. By selecting the network with the lowest load, load balancing can be maintained among networks, thus, call block rate can be significantly reduced [2]. In service-based algorithm, voice service is allocated to GERAN while active data service is allocated to UTRAN [3], which increases system throughput compared with the reverse policy. The user-centric solution is proposed to provide better user experience. It uses utility function about cost and latency to determine target network [4].

As shown in Fig. 1, in metropolitan area, group calls or handovers will occur frequently because of the high density of mobile users and their movements. In addition, lots of passengers take vehicle device such as bus and train every day. When the device leaves the coverage of a wireless network, those users in the vehicle have to select and switch to another network to continue their wireless access simultaneously [5][6].

This work was supported by National Natural Science Foundation of China (No.60832009), Funds for Creative Research Groups of China (No.61121001), Program for Changjiang Scholars and Innovative Research Team in University (No.IRT1049) and Major national S&T project (No.2010ZX03002-010-01).

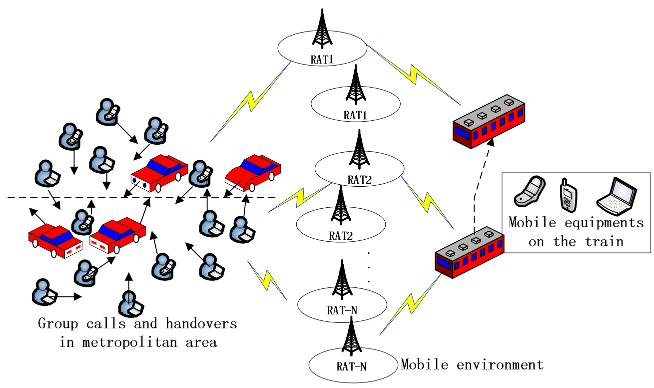

Figure 1. Typical GJCAR scenarios

Traditional algorithms [1-4] are based on the assumption that users send access requests one by one and the decision entity makes network selections for users by turn. However, in JGCAC scenario, since these algorithms don't consider the impact introduced by other users' decisions, users may choose improper networks, which may lead to network congestion and higher dropping rate. In [5], random delay is introduced to avoid making decisions simultaneously. Every requesting user delays a random period of time, and then sends access request again. However, it may cause long delays for some users and can't provide differential QoS guarantee according to user requirements. In [7], a centralized decision-making model is proposed to make network selection decisions for multiple users simultaneously, which can be used to solve the problem of incomplete information faced by JGCAC. The scheme formulates the JCAC problem to a mixed integer programming (MIP) problem, so the complexity of calculation is high in the case of mass requesting users. However, the centralized decision-making thought can be a good guidance to solve the problem of incomplete information for JGCAC.

Based on the above analysis, how to make network selections for users considering the impact of other users is the main challenge for JGCAC, and effective algorithm should be developed to avoid network congestion and to provide QoS guaranteed service for users. In addition, the complexity of JGCAC algorithm should be low.

In this paper, a novel JGCAC algorithm is proposed to deal with the problems discussed above. In the proposed centralized decision-making algorithm, the number of allocated users for each network is determined based on load balancing policy, and then distribute users to networks by user satisfaction. Moreover, with the introduction of Hungary Algorithm, the complexity of the algorithm is low. The proposed algorithm can be implemented in a centric entity located at the network side, e.g., the Common Radio Resource Management (CRRM) server proposed by 3GPP [8].

The rest of this paper is organized as follows. In section II, the network load model and the user evaluation model related to the proposed algorithm are introduced. In section III, the proposed JGCAC algorithm is described. The performance of the proposed JGCAC scheme is evaluated in section VI. Section V concludes the paper.

II. MODEL FORMULATION

In JGCAC scenarios, multiple users send access requests simultaneously. To avoid network congestion, load balancing is of vital importance from the perspective of network operators. As for users, better service experience is preferred, e.g., lower latency, less cost and etc. Therefore, in this section, two models are constructed, i.e., network load model and user evaluation model.

A. Network Load Model

Suppose at the given time, there are N requesting users and M available networks. Let $U = \{u_1, ..., u_N\}$ and $A = \{n_1, ..., n_M\}$ denote the set of requesting users and the set of available networks, respectively. Among the N requesting users, the number of newly coming users and the number of handover users are N' and N'', respectively. Let b_i be the requested data rate of user u_i. The total available bandwidth and occupied bandwidth of network n_i are supposed to be C_i and O_i, respectively. Let η_i be the load factor of network n_i. Then, η_i is defined as

$$\eta_i = \frac{O_i}{C_i} \tag{1}$$

Before JGCAC, to evaluate the load of each network, the initial load factor is denoted by

$$\eta_i' = \frac{O_i - \sum_{k \in P_i} b_k}{C_i} \tag{2}$$

where P_i is the set of users who want to handoff from network n_i.

The average load factor of the whole system is given by

$$\bar{\eta} = \frac{\sum_{i=1}^{M} O_i + \sum_{j=1}^{N'} b_j}{\sum_{i=1}^{M} C_k} \tag{3}$$

where $\bar{\eta}$ is the ratio of total requested bandwidth to the total bandwidth the system can provide.

B. User Evaluation Model

Service latency, bit error rate (BER), handover frequency and cost are important to users, so network load factor, received signal strength (RSS), mobility speed and cost are selected to reflect user satisfaction. In order to select the best networks for users, TOPSIS [9] is introduced to evaluate the candidate networks by calculating the similarity of the candidate networks and the optimal network.

The coding and modulation method as well as radio coverage are different among RATs. To compare the RSS from different RATs and to match the mobility speed with the network coverage, fuzzy evaluation method is used [10]. As shown in Fig. 2, for all the available networks, if the RSS is bigger than a certain threshold, the fuzzy evaluation value is 1, and if the RSS is smaller than a certain threshold, the fuzzy evaluation value is 0, otherwise the value is between 0 and 1. Let the RSS evaluation thresholds be different among RATs. Similarly, as shown in Fig. 3, the best-matched speeds are different among RATs. NETWORK 1 is fit for low speed users, so when the speed is slower than 15 km/h, the fuzzy evaluation value is 1, otherwise the value is smaller than 1. NETWORK 2 and NETWORK 3 are fit for medium and high speed users, respectively.

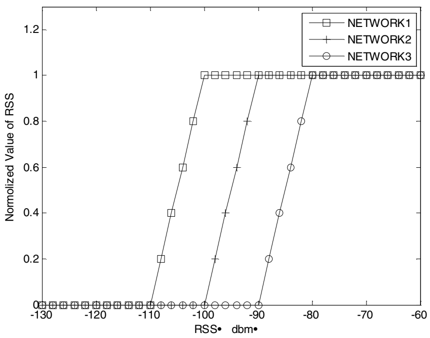

Figure 2. Evaluation of RSS

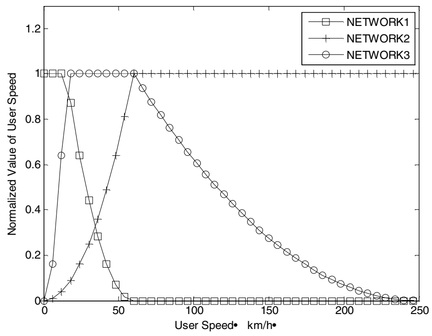

Figure 3. Evaluation of speed match

As smaller-the-better parameters, network load factor and cost are normalized as follows

$$x_i^*(k) = \frac{\max_k - x_i(k)}{\max_k - \min_k} \tag{4}$$

where $x_i(k)$ is the value of parameter k for network n_i, $\max_k$ and $\min_k$ are the maximum and minimum values among all the available networks, respectively.

Let c_{ij} be the evaluation value of network n_i to user u_j. c_{ij} is defined as the Euclidean distance between the network parameters and the optimal parameters denoted by 1. In this paper, c_{ij} is used to reflect user satisfaction, i.e., the user is more satisfied with smaller c_{ij}. We formulate c_{ij} as follow

$$c_{ij} = \sqrt{\sum_{k=1}^{4} w_k (x_i^{j*}(k) - 1)^2} \tag{5}$$

where ω_k is the weight of parameter *k*. $x_i^{j*}(k)$ is the normalized evaluation value of parameter *k* for network n_i to user u_j.

III. Proposed Algorithm for JGCAC

Maintaining load balance and providing user with satisfied service are the two main objectives for JGCAC. Actually, when each user selects the preferred network, load balancing is hard to realize. Moreover, to keep load balance means that some users can't select the best networks for them. Since the two objectives can't be satisfied simultaneously, and congestion avoidance and dropping rate reduction are more important in JGCAC scenario, in this paper, we give more priority to load balancing and then seek to meet user requirements under load balancing.

Based on the above analysis, the proposed algorithm includes two procedures as follows

- Determining the number of users distributed to each network based on load balancing policy.
- Under the fixed number of allocated users for each network, selecting the target network for users to maximize the total user satisfaction with extended Hungary Algorithm.

1) Determining the number of users distributed to each network

In order to realize load balancing, the proposed method is as follows

a) With the requested bandwidths of users, the initial load factor η_i' for each network is given by (2).

b) The average load factor of the whole system $\bar{\eta}$ is indicated in (3).

c) ①If the load factor of each network is equal or less than $\bar{\eta}$, the number of allocated users denoted by t_i is given by

$$t_i = \left\lfloor C_i \times (\bar{\eta} - \eta_i) / \bar{b} \right\rfloor \tag{6}$$

where $\bar{b}$ is the average requested bandwidth, and t_i is calculated as the number of allocated users to make the load factor of network n_i equal to $\bar{\eta}$.

②If the load factor of network n_j is bigger than $\bar{\eta}$, t_j of network n_j is set to be zero, and repeat step *b)* to calculate the average load factor of the whole system, exclusive of network n_j. Then repeat step *c)*.

d) For constraint $\sum_{i=1}^{M} t_i = N$, if $\Delta N = N - \sum_{i=1}^{M} t_i > 0$, continue the following steps, otherwise the number of allocated users is set by the above steps.

e) The new load factor of network n_i is given by

$$\eta_i'' = \frac{O_i - \sum_{k \in P_i} b_k + \bar{b} \times t_i'}{C_i} \tag{7}$$

Assuming that network n_k has the minimum value η_k'', add one allocated user for network n_k, i.e., $t_k = t_k + 1$.

f) Repeat step *e)* for ΔN times.

2) selecting the target networks for users to maximize the total user satisfaction

After deciding the number of allocated users for each network, how to distribute users to networks under fixed t_i is the subsequent problem. In this paper, in order to maximize the total user satisfaction, the well-known Hungary Algorithm is used. The details are as follows.

As stated in section II, the evaluation value of network n_i to user u_j, which is denoted by c_{ij}, is used to reflect user satisfaction, i.e., the user is more satisfied with smaller c_{ij}. Therefore, the optimization model of Hungary Algorithm [11] can be formulated as

$$\begin{aligned} & min \quad z = \sum_i \sum_j c_{ij} x_{ij} \\ & s.t. \begin{cases} \sum_i x_{ij} = 1, & i = 1, 2, \ldots, M \\ \sum_j x_{ij} = 1, & j = 1, 2, \ldots, N \\ x_{ij} = 1 \ or \ 0 \end{cases} \end{aligned} \tag{8}$$

where $x_{ij} = 1$ means user u_j selects network n_i, and $x_{ij} = 0$ means user u_j doesn't select network n_i.

Since the premise of Hungary Algorithm is that one network can only admit one user, on the condition $t_i > 1$, it is assumed that there are $t_i - 1$ extended networks to admit these users, and these extended networks have the same evaluation values with network n_i . The final evaluation matrix is extended as

$$\begin{bmatrix} c_{11} & c_{12} & \cdots & \cdots & c_{1N} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ c_{11} & c_{12} & \cdots & \cdots & c_{1N} \\ c_{21} & c_{22} & \cdots & \cdots & c_{2N} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ c_{21} & c_{22} & \cdots & \cdots & c_{2N} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ c_{M1} & c_{M2} & \cdots & \cdots & c_{MN} \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix} a \\ a \\ a \end{matrix}}\right\} t_1 \\ \left.\vphantom{\begin{matrix} a \\ a \\ a \end{matrix}}\right\} t_2 \\ \vdots \\ \vdots \end{matrix} \tag{9}$$

Then the classic Hungary Algorithm is used to solve the optimization statement in (8).

In this paper, evaluation value c_{ij} is determined by TOPSIS concerning load factor, RSS, mobility speed and cost. Actually, other QoS related criteria and evaluation methods can also be used according to the preference of network operators. Therefore, the proposed algorithm can easily combine with other JCAC algorithms to achieve the required performance.

IV. SIMULATION RESULTS AND ANALYSES

Besides the proposed algorithm in section III, two traditional JACAC algorithms for single user are also given for comparison. Random delay policy [5] is introduced to both of the compared algorithms to avoid simultaneous requests. Compared algorithm 1 selects the network with the minimum load factor as target network. Compared algorithm 2 selects the network with the minimum evaluation value as target network.

For the simulation parameters, it is supposed that there are two WLAN networks, two UMTS networks and one WiMAX network in current area. The available bandwidths of WLAN, UMTS and WiMAX are denoted by vector $C = \{11, 2, 20\}$ (*MHz*). The set of occupied bandwidth of each network is $O = \{10, 9, 1.5, 1, 18\}(MHz)$ .The requested bandwidth of user u_i can be selected from the set $B = \{64, 128, 192\}(KHz)$. It is supposed that there are multiple users sending access requests at the given time. For the calculation of the evaluation value, the weights of load factor, RSS, mobility speed and cost are 0.2, 0.3, 0.3 and 0.2, respectively. For each combination of user and network, we select a randomized average RSS value. The mobility speed of each user is also set a random value.

In Fig. 4, the coefficient of variation of loads (COV) versus the number of requesting users is plotted. The coefficient of variation of loads is defined as the standard deviation of load factors observed in networks divided by the mean load factor. This definition has extensively been used as a metric in the literature for the illustration of the distribution of load [12]. As expected, the compared algorithm 1 achieves the best performance due to the fact that each user selects the network with minimum load factor after random delay. The compared algorithm 2 achieves the worst performance because each user selects the target network according to the evaluation values, which may lead to mass users select the same network. The proposed algorithm provides performance that lies between those of the compared algorithms. The reason of the slightly degraded performance of the proposed algorithm compared with the compared algorithm 1 is that the number of allocated users for each network is determined using the average requested bandwidth. However, users with various requested bandwidths may be distributed to networks unevenly. The proposed algorithm is still better than the compared algorithm 2 because in the proposed algorithm, the number of allocated users for each network is determined based strictly on the principle of load balancing.

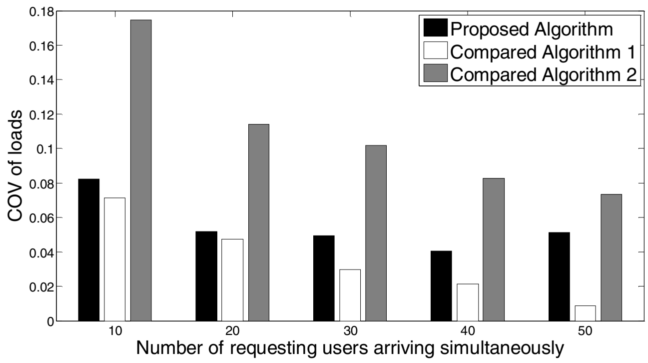

Figure 4. Distribution of load across networks

TABLE I. COMPARISON OF THE NUMBER OF ALLOCATED USERS FOR DIFFERENT ALGORITHMS

Number of Allocated users	**Network ID**				
	n_1	n_2	n_3	n_4	n_5
Proposed algorithm	0	2	2	6	0
Compared algorithm 1	0	2	2	6	0
Compared algorithm 2	0	0	3	5	2
Initial load factor	0.909	0.818	0.750	0.500	0.900
Average load factor under load balancing	0.887				

Table I shows the number of allocated users for each network when there are 10 requesting users. From the table, it can be observed that the simulation result of the proposed algorithm is the same with that of the compared algorithm 1, which means the proposed algorithm can realize load balancing efficiently. Furthermore, for the above two algorithms, when the initial load factor of a certain network is greater than the average load factor, the number of allocated users for that networks is zero. In contrast, for the compared algorithm 2, two users select network n_5 with initial load factor greater than the average load factor, which will make network n_5 vulnerable to congestion.

Fig. 5 gives the performance comparison of user satisfaction which is denoted by the average evaluation value of admitted users. As mentioned before, smaller average evaluation value of admitted users means the admitted users are more satisfied. As indicated from the curve, the proposed algorithm can achieve a better user satisfaction. The user

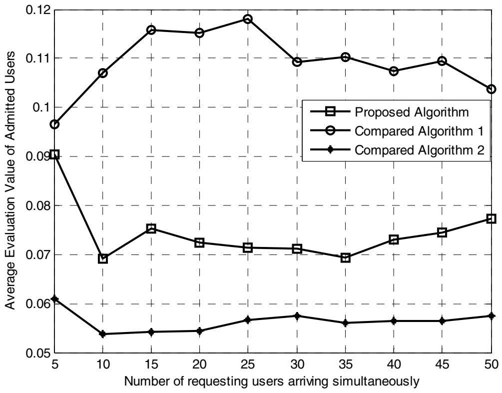

Figure 5. Average evaluation value of admitted users comparison

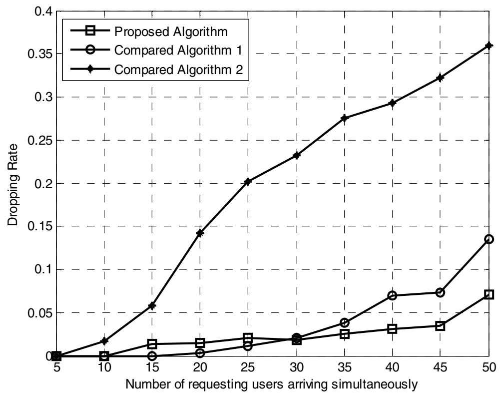

Figure 6. Dropping rate comparison

satisfaction of the compared algorithm 1 is the worst due to the ignorance of user requirements. Although the compared algorithm 2 achieves the best user satisfaction for admitted users, Fig. 6 shows that its dropping rate is extremely high due to selecting the best network irrespective of load balancing, which means more users are not satisfied because of the rejection of service. As indicated in Fig. 6, the proposed algorithm and the compared algorithm 2 can achieve a lower dropping rate because of the consideration of load balancing to distribute the requesting users to networks. Furthermore, when the number of requesting users is larger than 30, the proposed algorithm can achieve the best performance, because with more requesting users, the users with various requested bandwidths are distributed to networks more uniformly.

Based on the above simulation results, it can be concluded that the proposed algorithm has better overall performance than the two compared traditional algorithms. Compared algorithm 1 doesn't consider the user requirements, so it can't provide users with QoS guaranteed service. Compared algorithm 2 doesn't pay much attention to the congestion problems faced by JGCAC, which leads to unbalanced loads and higher dropping rate. The compared algorithm combines the advantages of the two compared algorithms and avoids the problems faced by the compared algorithms. It can effectively maintain load balance as well as improve total user satisfaction, and the dropping rate is the lowest. As for the JGCAC latency, obviously the proposed algorithm can achieve a better performance due to the centralized decision-making method without the random delay introduced by the compared algorithms.

V. CONCLUTION AND FURTHER WORK

In JGCAC scenario, how to select the proper networks for mass requesting users timely and efficiently is a challenging problem. This paper proposed a centralized algorithm to make network selection decisions for users simultaneously to avoid excess delays. The proposed algorithm is developed to distribute users into different networks to avoid network congestion, meanwhile, user satisfaction is considered to provide better service experience. Simulation results indicate that the proposed algorithm have a better overall performance compared with the traditional algorithms. It can effectively achieve load balancing and better satisfaction of admitted users, meanwhile the dropping rate is the lowest. In the future, we will work to consider other QoS parameters to reflect the various requirements of multiple services.

REFERENCES

[1] A. H. Zahran, B. Liang, and A. Saleh, "Signal threshold adaption for vertical handoff in heterogerous wireless networks," Mobile Networks and Applications. vol. 11, pp. 625-640, 2006.

[2] X. Gelabert, J. Pérez-Romero, O. Sallent, R. Agustí, "On the suitability of load balancing principles in heterogeneous wireless access networks," IWS2005/WPWC'05, Aalborg, Denmark.

[3] X. Gelabert, J. Pérez-Romero, O. Sallent, R. Agustí, F. Casadevall, "Radio resource management in heterogeneous networks".

[4] Olga Ormond, John Murphy, "utility-based intelligent network selection in beyond 3G systems," EEE ICC, 2006.

[5] Cai Xuejun, Liu Fang, "Network selection for group handover in multi-access networks," , 2008.

[6] Hamid Aghvami, Nima Sattari, Oliver Holland, "Group handover among heterogeneous wireless networks," IEEE MELECON, May 16-19, 2006.

[7] SuKyoung Lee, Kotikalapudi Sriram, Kyungsoo Kim, Yoon Hyuk Kim and Nada Golmie, "Vertical handoff decision algorithms for providing optimized performance in heterogeneous wireless networks," IEEE Transactions on Vehicular Technology, vol. 58, no. 2, February 2009.

[8] Improvement of RRM across RNS and RNS/BSS(Release 5)//3GPP TR25.881 v5.0.0., Sophia Antipolis, France, 2001.

[9] Enrique Stevens-Navarro and Vincent W.S. Wong, "Comparision between vertical handoff decision algorithms for heterogeneous wireless networks," 2006 IEEE.

[10] Lorenza Giupponi, Jordi Perez-Romero, "A novel approach for joint radio resource management based on fuzzy neural methodology," IEEE Transactions on Vehicular Technology, vol. 57, no. 3, May 2008.

[11] Yunquan HU, Yaohuang Guo, Operations Research Tutorial, 3rd ed., Tsing Hua University Press, April 2007.

[12] Sukyoung Lee, Yoon Hyuk Kim, Nada Golmie, "Vertical handoff decision algorithm for providing optimized performance in heterogeneous wireless networks," IEEE Transactions on Vehicular Technology, vol. 58, no.2., February 2009.