Utility-based Dynamic Multi-service Bandwidth Allocation in Heterogeneous Wireless Networks

Wenchao Fei[1], Hui Tian[2] and Rongrong Lian[3]
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: [123]{wenchao.fei, bupttianhui, lianrrong}@gmail.com

Abstract—**This paper presents a dynamic bandwidth allocation scheme for multiple services in heterogeneous networks. A utility function is introduced first, whose parameters are determined by the modified multi-state Analytic Hierarchy Process (AHP) which adapts to different load levels according to dynamic thresholds. Based on this, a bandwidth allocation algorithm is proposed for Constant Bit Rate (CBR) and Variable Bit Rate (VBR) services depending on not only utility fairness among different networks, but also the fairness between new arrival and ongoing services. Simulation results show that the proposed algorithm is effective in guaranteeing both the fairness of the edge users and the network system performance, i.e. keeping low block probability in the wireless network system.**

Keywords-heterogeneous networks; utility function; fairness; bandwidth allocation; parallel transmission

I. INTRODUCTION

The heterogeneity will be an important characteristic of the future wireless communication systems [1][2], which may include multiple disparate wireless networks, such as WLAN, Bluetooth, GPRS, 3G cellular networks, etc. However, each single Radio Access Network (*RAN*) could not meet the requirements of all services because of the limited communication capability and isolated working style. The integration and cooperation of heterogeneous wireless networks as well as the heterogeneous radio resource management will become the major issues.

Recently, the issue of bandwidth allocation in heterogeneous environment has drawn much attention. The Always Best Connected goal [3] is the main principle in bandwidth allocation so that the service accesses the most suitable RAN [4][5][6], but this usually results in unbalanced load distribution and can't take the advantage of heterogeneous resources. Traffic splitting to make full use of the network resource is also a principle proposed in several literatures [7][8]. However, they cannot fairly allocate bandwidth ignoring the difference of the location of the user and the coverage of the RANs. Though a fair bandwidth allocation method for traffic splitting was proposed in [9], it mainly focused on the fairness of the whole network, leading to high block rate for some users and low network efficiency.

Therefore, both the resource utilization and the fairness of users should be taken into consideration in the issue of heterogeneous bandwidth allocation, which are seldom discussed in current literatures. This paper pays much attention to them.

In this paper, a utility-based dynamic bandwidth allocation algorithm is presented. To introduce the utility function, the multi-state AHP is proposed to adapt to different load levels according to dynamic thresholds. Thus, the utility function is able to dynamically synthesize the network parameters representing both network system performance and fairness of edge users. Based on the utility function, a bandwidth allocation algorithm is proposed for CBR and VBR services depending on utility fairness, in which the VBR services can degrade bandwidth on the condition of network congestion. The numerical simulation results indicate that the proposed scheme outperforms other existing mechanisms.

The rest of this paper is organized as follows: The utility function is described in Section II. Section III presents the proposed strategy in detail. Section IV illustrates the numerical simulation results. Finally, Section V concludes this paper.

II. THE UTILITY FUNCTION

In this section, a utility function is introduced at first. Then the network fittingness factor is proposed and discussed to introduce the traditional utility function to the heterogeneous network bandwidth allocation scenario.

In order to consider the influence of the location of users and the coverage of RANs, a heterogeneous wireless access network scenario is constructed, as shown in Fig.1.

The geographical area is totally covered by a WMAN base station (area 1), while partly covered by a 3G base station (area 2) and a WLAN access point (area 3). In area 3, a Mobile station (*MS*) can access any type of networks. Accordingly, a MS in area 2 can access 3G and WMAN, and, a MS in area 1 can only access WMAN. It is assumed that the MS is able to access any type of networks in its corresponding area, and perfect power control is supposed to ensure uniform available transmission rate.

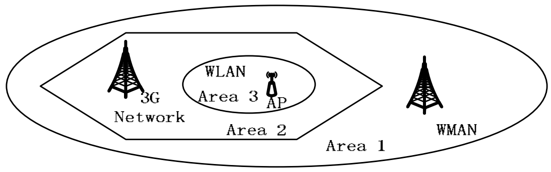

Figure.1 Heterogeneous Wireless Network Architecture

This work was supported by National Natural Science Foundation of China (No.60971125), 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).

A. The Utility Function

In economics, utility is a measurement of relative satisfaction, which refers to the total satisfaction received by a consumer from consuming a good or service. Utility function is adopted in this paper is to map bandwidth provided by RAN into performance change as perceived by the networks. The utility function is related with the allocated bandwidth P and it can be defined as follows [9]:

$$U(P) = \omega \log(1 + \alpha \bullet P) \tag{1}$$

where $U(P)$ is the utility of a new connection, ω and α are constants indicating the scale and the shape of the utility function.

However, compared with a single network, the same bandwidth change to different networks should lead to different performance effect on the corresponding utility. Thus the utility function (1) should be redefined.

The utility function is modified by introducing a fittingness factor ε , which adapts P to the heterogeneous bandwidth allocation scenario, i.e.

$$U(P) = \omega \log(1 + \alpha \bullet P / \varepsilon) \tag{2}$$

B. The Fittingness Factor

Compared with the traditional wireless environment, the different coverage of different RAN results in a new fairness problem, especially for the edge users in the large coverage networks. As shown in Fig.1, in area 3 users could access three RANs while only one can be accessed in area 1. As a result, if large number of users exist in area 3 and consume large part of the capacity of WMAN, the block rate will be high in area 1. On the other hand, an improper way to guarantee the fairness of the edge users usually leads to the low efficiency of the network. e.g., reserving fixed bandwidth for users in area 1 and 2 will cause the lack of bandwidth for users in area 3 and the unbalance load distribution.

Based on discussions above, the fittingness factor should take account of the network parameters indicating both the network system performance and the fairness of the edge users in the large coverage networks. More importantly, it should be adjusted dynamically. In this paper, four network parameters are selected in the two aspects, i.e.

- Network parameters for system performance: wireless network capacity C and available capacity C' . The bandwidth allocation should balance the load to ensure the throughput and avoid block. C stands for the initial state and the potentiality of bearing load, while C' represents the present ability of bearing load.
- Network parameter for the edge user fairness: the Coverage of wireless networks S and Service density of network coverage area ρ . The service should be allocated more bandwidth from small coverage network than the large coverage network to reserve bandwidth for edge users. Similarly, the bearing load proportion of large coverage network should be decreased if service density in small area gets higher.

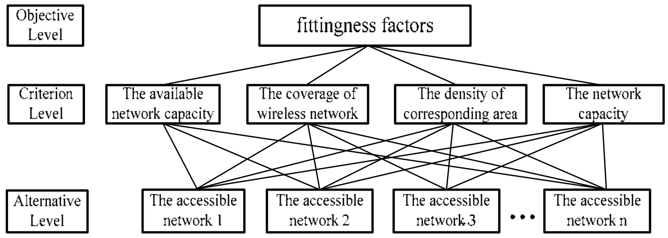

Figure.2 AHP decision hierarchy for the integrated system

Therefore, to synthesize these network parameters and gain the fittingness factors corresponding to different wireless networks, AHP is an option. However, the importance of each parameter is hard to evaluate due to the varying network load. Therefore, the AHP is modified as multi-state AHP to adapt to different load levels according to dynamic thresholds.

1) Multi-state AHP

With the network parameters mentioned above, the AHP decision hierarchy for the 3G/WLAN/WMAN integrated system is constructed as in Fig.2.

The judgment matrix between the criterion level and alternative level is based on the value of the network parameter. However, the judgment matrix W of the criterion level under objective level is based on the experience or willingness of operator, and it should agree with the consistency requirement. Assume that the importance of the criterion level to the objective level can be solved as $w = [w_{c'}, w_s, w_\rho, w_c]$.

In order to evaluate the influence of the network load on the judgment matrix W , the wireless networks are divided into three states: high load, middle load and light load. Denote the judgment matrixes of the three states as W^h , W^m , and W^l respectively.

Since the bandwidth of new arrival service could be allocated by its maximum requirement in light or middle load state, the importance of fairness should increase as the load level becomes higher. Meanwhile, when the networks are in low or middle load, the bandwidth allocation considers the load balance principle to improve the system performance, i.e.

$$w_c^l + w_{c'}^l \geq w_c^m + w_{c'}^m \geq w_c^h + w_{c'}^h \tag{3}$$

$$w_s^l + w_\rho^l \leq w_s^m + w_\rho^m \leq w_s^h + w_\rho^h \tag{4}$$

2) Dynamic thresholds

The thresholds are the crucial factors in the determination of network load state. To further guarantee the fairness, a dynamic threshold scheme is proposed. The wireless networks that overlay the current area are divided into two classes: the large coverage networks (class 1) and the smallest coverage network(class 2). Taking users in area 3 for example, the WMAN and the 3G network are the class 1, and WLAN is the class 2.

Considering the area k with m kinds of access networks, and the network n is the smallest coverage one. Denote thresholds $\theta_1^h, \theta_2^h (\theta_1^h \leq \theta_2^h)$, $\theta_1^m, \theta_2^m (\theta_1^m \leq \theta_2^m)$, θ_1^l, θ_2^l $(\theta_1^l \leq \theta_2^l)$ as the thresholds for class 2 corresponding to different load levels of class 1, and $\theta_1^{class1}, \theta_2^{class1}$ ($\theta_1^{class1} \le \theta_2^{class1}$) as the load level thresholds for class1, θ_i as the load level of network i.

The procedure of determining the W in multi-state AHP by using dynamic thresholds is depicted as below.

Algorithm 1 Dynamic Threshold Algorithm to determine W

1. **If** $1 \le \forall i \le m$ *and* $i \ne n, \theta_i \ge \theta_2^{class1}$,then
2. Determine W with $\theta_1^h, \theta_2^h (\theta_1^h \le \theta_2^h)$ and θ_n

$$\left(\begin{array}{l} \theta_n \ge \theta_2^h, W = W^h \\ \theta_1^h < \theta_n \le \theta_2^h, W = W^m \\ \theta_n \le \theta_1^h, W = W^l \end{array} \right)$$

3. **Else if** $1 \le \forall i \le m$ *and* $i \ne n, \exists \theta_i \le \theta_1^{class1}$
4. Determine the W with $\theta_1^l, \theta_2^l (\theta_1^l \le \theta_2^l)$ and θ_n
5. **Else**
6. Determine the W with $\theta_1^m, \theta_2^m (\theta_1^m \le \theta_2^m)$ and θ_n
7. **End if**

Based on the definition of parameter importance in W^h, W^m, and W^l, when networks of class 1 have higher load level, the bandwidth allocation procedure of class 2 should have larger probability to select W^h or W^m to reserve bandwidth for the edge users, therefore:

$$\theta_1^h \le \theta_1^m \le \theta_1^l \quad (5)$$

$$\theta_2^h \le \theta_2^m \le \theta_2^l \quad (6)$$

III. Utility-Fair Bandwidth Allocation

In this section, a bandwidth allocation strategy is proposed, depending on not only utility fairness within several different networks, but also the fairness between the new arrival service and all the ongoing services.

Since the bandwidth requirement of VBR services is not fixed, when a VBR service access networks, a procedure is proposed to calculate its exact bandwidth value firstly, which is called bandwidth supply plan in this paper.

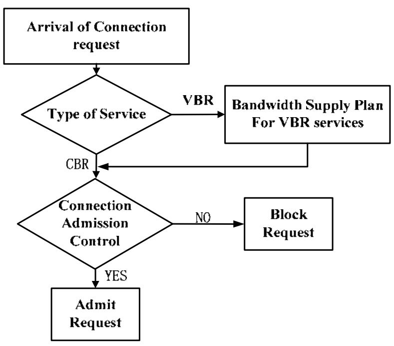

Figure.3 Process of Bandwidth Allocation for Multiple services

The procedure of bandwidth allocation for multiple services is shown in Fig.3. When a new connection arrives at one area, its type of services will be checked firstly. If it is a VBR connection, the bandwidth supply plan is run to obtain the exact VBR bandwidth. Then, the connection admission control is run to decide whether to admit the request or not.

A. Bandwidth Supply Plan

Considering the area k with a heterogeneous wireless access environment where there are m classes of available access networks containing N ongoing connections among which N_1 is for the CBR services and N_2 for the VBR services. For the j th connection of the i th network in the k th area, denote its current allocated bandwidth as $b_{i,j,k}$ and its utility as $U_{i,j,k}(b_{i,j,k})$. In the process of bandwidth supply plan for a new arrival VBR services, a method of average plan is implemented.

For the first arrival VBR service in the area, the connection can be allocated bandwidth according to its maximum bandwidth requirement, i.e.

$$N_2 = 0, B_{N+1,k} = B_{\max}^{VBR} \quad (7)$$

where $B_{N+1,k}$ is the bandwidth value of the new arrival VBR service. It should be allocated bandwidth by m class networks, i.e.

$$B_{N+1,k} = \sum_i b_{i,N+1,k}, i = 1, 2, \ldots, m \quad (8)$$

where $b_{i,N+1,k}$ is the bandwidth that the i th network allocates to the new arrival connection.

In the case VBR services exist in the area, the average bandwidth of all ongoing VBR services is taken as the new arrival VBR service bandwidth, i.e.

$$N_2 > 0, B_{N+1,k} = \sum_n B_{n,k} / N_2, n = 1, 2, \ldots N_2 \quad (9)$$

Therefore, compared to the existing VBR services in area k, the new arrival VBR service obtains fair bandwidth according to bandwidth supply plan.

B. Utility-Fair Bandwidth Allocation

The objective of the utility-fair bandwidth allocation scheme is to guarantee the bandwidth allocation fairness between RANs when they provide bandwidth to a MS simultaneously. Since the new arrival service could obtain bandwidth by decreasing VBR service bandwidth, two network states are discussed: light load and full capacity occupied.

Light Load: For the light load case, all VBR connections can be allocated bandwidth according to their maximum bandwidth requirement (actually, based on the VBR bandwidth supply plan, when RANs are in light load state, the arrival VBR service is allocated the maximum requiring bandwidth just as the first one), while all CBR connections are allocated fixed bandwidth as the service requirement.

For both CBR and VBR services, to achieve the bandwidth allocation fairness between RANs, the utility of different RANs should be equal, i.e.

$$1 \le \forall i_1, i_2 \le m, U_{i_1,N+1,k}(b_{i_1,N+1,k}) = U_{i_2,N+1,k}(b_{i_2,N+1,k}) \qquad (10)$$

By solving the equations (10), (2), (8) and (7)or(9), the allocated bandwidth $b_{i,N+1,k}$ and the multiple network bandwidth allocation proportion $p_{i,N+1,k}$ can be obtained, shown as equation (11).

$$p_{i,N+1,k} = b_{i,N+1,k} / B_{N+1,k} \qquad (11)$$

Since $\alpha_{i_2,k}$, $\omega_{i_1,k}$ are constant and known, $\varepsilon_{i_2,N+1,k}$ is the factor that indicates the bandwidth allocation proportion. Therefore, the allocation procedure takes account of the fairness of edge users and network system performance.

Full Capacity Occupied: For the full capacity occupied case, the RAN cannot afford the bandwidth for the new arrival connection. Hence, all ongoing VBR services in the area degrades bandwidth fairly to admit the arrival one, until there exist ongoing VBR services reach its minimum bandwidth requirement. The procedure follows the steps blow.

Step 1: The bandwidth which needs to be allocated by RAN i $(1 \le i \le m)$, namely $b_{i,N+1,k}$, is calculated based utility fairness just like the way in the light load state.

Step 2: In each RAN, all ongoing VBR connections decrease bandwidth fairly, following the equation (12) and (13), where denotes $\beta_{i,j,k}$ as the decreased bandwidth of the i th network and j th VBR connection in area k .

$$1 \le \forall i \le m, 1 \le \forall j_1, j_2 \le N_2, \beta_{i,j_1,k} / b_{i,j_1,k} = \beta_{i,j_2,k} / b_{i,j_2,k} \quad (12)$$

$$b_{i,N+1,k} = \sum_j \beta_{i,j,k}, j = 1, 2, \ldots, N_2 \qquad (13)$$

Step 3: If all the ongoing VBR services in area k satisfy equation (14), the new arrival connection is admitted, otherwise it is rejected.

$$1 \le \forall j \le N_2, B'_{j,k} = (B_{j,k} - \sum_i \beta_{i,j,k}) \ge B_{\min}^{VBR} \qquad (14)$$

IV. SIMULATION RESULTS

A. Parameter illustration

The simulation scenario is shown as in Fig.1. In case of IEEE 802.16e-based wireless access, the data transmission rate is 20Mbps in a single cell. For the 3G cellular network, the total transmission rate is 2Mbps in a cell. And for the WLAN, the IEEE 802.11b physical layer is used and the average bit rate is assumed to be 11Mbps.

One class of video traffic is considered as the CBR service with bandwidth requirement of 256Kbps. And we assume that the VBR service is multimedia services with bandwidth requirement range from 256Kbps to 512Kbps.

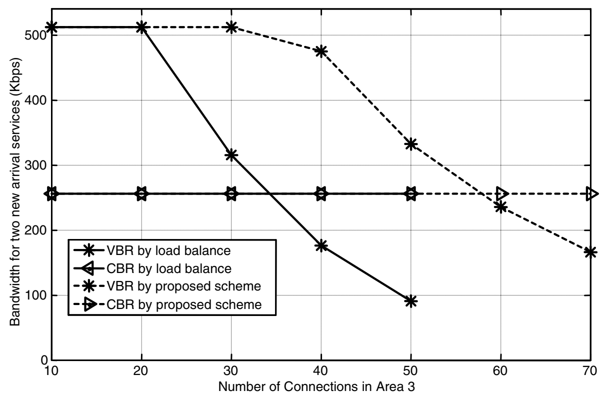

Figure.4 Bandwidth for two new arrival services in area 3

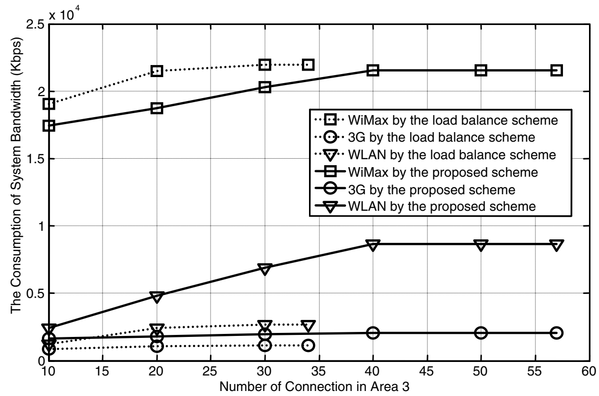

Figure.5 The total consumption of system bandwidth

For simplicity, the parameters for the utility function are set as follows: $\omega = \alpha = 1$. The thresholds are set: θ_1^l =0.4, θ_2^l =0.8, θ_1^m =0.3, θ_2^m =0.7, θ_1^h =0.2, θ_1^h =0.6, and weight of the criterions is w^l =[0.42,0.3,0.18,0.1], w^m =[0.33,0.36,0.2,0.1], w^h =[0.3, 0.39,0.21,0.1] by calculating W^l , W^m , W^h .

B. Simulation Results

Above all, we simulate the solution obtained from bandwidth allocation in two cases. The first case is the static scenario, in which the connections access the networks and don't leave until the simulation is ended. The second case is the dynamic scenario that the services are served after admitted, and then leave the system. For the sake of comparison, the load balancing strategy with fixed bandwidth allocation proportion among different networks is also presented, in which bandwidth allocated by a network is proportional to its network capacity.

In case 1, the number of connections in area 1 and 2 is 25 and 20, respectively, whereas that of area 3 increases, until the system cannot allocate bandwidth for the new arrival requests.

Fig.4 illustrates variations in the amount of bandwidth for the new arrival connection of two service types in area 3. For VBR services, as expected, when there are few connections in area 3, the new arrival connection get its bandwidth as its maximum requirement, and VBR services decrease bandwidth for the new arrival connection while system capacity is fully occupied. And for CBR services, the CBR services get fixed

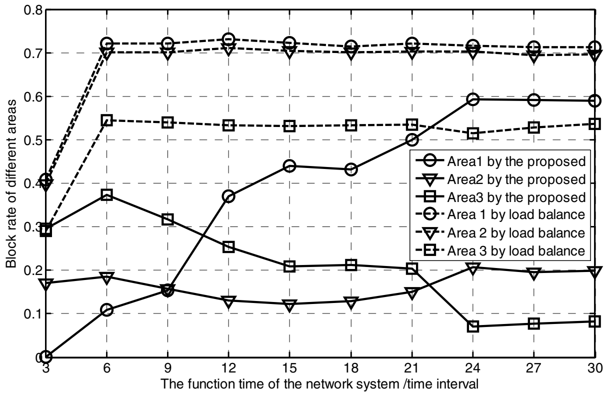

Figure.6 The block rate of users in different areas

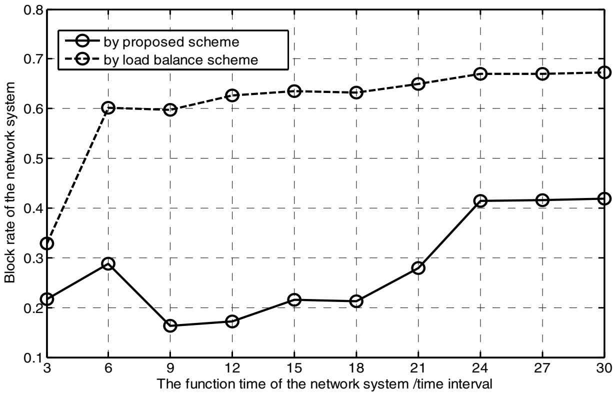

Figure.7 The block rate of users in different areas

bandwidth as the service requires. It also can be noticed that the proposed scheme can support more users.

The total consumption of system bandwidth is depicted in Fig.5. It can be found that to guarantee fairness of edge users, the services in area 3 consume more bandwidth in WLAN but lower in WMAN in the proposed scheme. Therefore the proposed scheme provides higher wireless resource utilization and fairness to the edge users.

In case 2, the arrival time of the new connection obeys Poisson distribution in all areas, and the serve time of connection obeys negative exponential distribution. After being served, the connection leaves the system. 30 time intervals are simulated, and the average arrival rates per time interval corresponding to the different areas are shown in Table I. The rate settings can reflect various network load levels.

Fig.6 plots the block rate of different areas. It can be observed that each of the block rate curves in proposed scheme is lower than the corresponding curve in load balance scheme, respectively. In the 0-10 time intervals, when a great number of services exist in area3, the block rate of users in area 1 still maintain low level. Block rate of the whole network system is shown in Fig.7. Notice that by the proposed scheme, the block rate is much lower compared with the load balance scheme.

TABLE I. THE ARRIVAL RATE OF DIFFERENT NETWORK

Arrival Rate	**Number of Area**		
	Area 1	***Area 2***	***Area 3***
1-10 time interval	20	10	60
10-20 time interval	40	10	40
20-30 time interval	60	10	20

V. CONCLUSION

In this paper, a dynamic bandwidth allocation algorithm is developed for multiple services in heterogeneous networks consisting of a WMAN, a 3G cellular network and a WLAN. Based on the proposed multi-state AHP and dynamic threshold schemes, it can provide dynamic bandwidth allocation for both CBR and VBR services by the utility function. Meanwhile, under the condition of network congestion, the VBR services can degrade their bandwidth to other types of services. The numerical simulation results show that the proposed algorithm is effective in supporting more services and keeping block probability substantially low, while guaranteeing the fairness of edge users. In the future, we will work to simplify the parameter setting and apply mathematic methods to optimize the block probability.

REFERENCES

[1] F.Ivanek, ”Convergence and Competition on the Way Toward 4G”,IEEE Microwave Magzine,vol.9, Issue 4, pp.6-14,Aug.2008

[2] A.Jamalipour, T. Wanda and T.Yamazato, ”A tutorial on multiple access technologies for beyond 3G mobile networks”, IEEE Communications Magazine, vol.43, Issue 2, pp.110-117, Feb. 2005

[3] Chen Yiping, Yang Yuhang, ”A new 4G architecture providing multimode terminals always best connected services,” IEEE Wireless Comm. Magazine, pp.36 – 41, Apr. 2007.

[4] Qingyang Song, Jamalipour. A , A network selection mechanism for next generation networks, ICC 2005, May 2005

[5] Sallent. O, Perez-Romero. J, Ljung. R, Karlsson. P, Barbaresi. A, ” Operator's RAT Selection Policies Based on the Fittingness Factor Concept”, Mobile and Wireless Communications Summit, 2007. 16th IST, July 2007

[6] E.Z. Tragos, et al., “Admission control for QoS support in heterogeneous 4G wireless networks,” IEEE Network Journal, vol. 22, no. 3, Jun. 2008, pp. 30 – 37.

[7] Furuskar. A, Zander.J, ” Multiservice allocation for multiaccess wireless systems”, Wireless Communications, IEEE Transactions, vol. 4, Issue 1, pp. 174 – 184, Jan. 2005

[8] Louha. K, Jun. J.H., ”Exploring Load Balancing in Heterogeneous Networks by Rate Distribution”, IPCCC 2008, Dec. 2008

[9] Changqing Luo, Hong Ji, Yi Li, ”Utility-Based Multi-Service Bandwidth Allocation in the 4G Heterogeneous Wireless Access Networks” , WCNC 2009. IEEE, April 2009