Optimising Radio Access in a Heterogeneous Wireless Network Environment

Weizhi Luo and Eliane Bodanese
Queen Mary, University of London
E-mail: {weizhi.luo, eliane.bodanese}@elec.qmul.ac.uk

***Abstract*—A variety of wireless network technologies have been developed and deployed, including GSM, UMTS, WiFi and WiMAX. The advantages of having an integrated heterogeneous wireless network environment include seamless communications, joint resource management and adaptive quality of service. In such environment, operators would not need to reject the service requests, but redirect them to appropriate networks. However, the sought aims of a heterogeneous network system still have many pending issues. One of them is the selection of the most appropriate radio access network (RAN) according to the requested service and the context information about the user and the candidate networks. We aim to develop efficient RAN selection algorithms to facilitate radio access optimisation for future heterogeneous network system. The simulation results show that our RAN selection algorithm can improve the network performance.**

I. Introduction

Nowadays, multiple Radio Access Technologies (RATs) are available as commercial wireless systems. The wireless technologies can be classified into three main categories: the mobile cellular network technologies (2G and 3G), the wireless local area network technologies (e.g. WiFi) and the wireless metropolitan area network technologies, such as Digital Video Broadcasting and WiMAX. The next generation mobile communication systems foresee the existence of a heterogeneous wireless communication environment with seamless communications, joint resource management and adaptive quality of service. In an environment with multiple technologies, to make the Radio Access Networks (RANs) cooperate with each other to achieve those aims is a challenge.

The Next Generation Mobile Networks (NGMN) [1] provides a set of recommendations to support the work of standardization bodies and manufacturers towards a cost effective future integrated mobile communication system. There are three groups of recommendations. The first group are the functional recommendations which provide recommendations in several criteria enabling operators to offer flexible and attractive services. Some of the functional recommendations are closely related to RAN selection research. For example [1]:

1. End to end QoS in all segments and preferably optimum end to end QoS with service continuity.
2. Seamless mobility management, preferably based on intelligent infrastructure.
3. QoS based global roaming and interworking.
4. Real time conversational and streaming in packet switched across all required bearers.
5. Valued based charging for integrated network.
6. Scalable core throughput to allow for deployment options that match specific operators and traffic requirements, and optimise radio resources

The second group of recommendations are related to cost efficiency. Several recommendations of this group also influence the way RAN selection solutions should be considered. The main recommendations that would impact RAN selection solutions are [1]:

1. Fully integrated multi frequency sites IP backhaul and IP/MPLS backhaul.
2. Maximum throughput without proportional incremental costs.
3. One integrated network with RAN, Core and Transport with convergence fixed and mobile.
4. IMS like service management.
5. Negotiated access between the terminal and the network (under the guidance of the network), preferably optimised access for the application and terminal with user preferences.
6. Highly intelligent multipurpose handsets and devices.

The third group are the overarching recommendations which provide guidance to evaluate deployment suitability. The NGMN also expects the integrated network to maximise resource exploitation, where terminals are required to support other RATs. A Session Initiation Protocol (SIP) based subsystem may be implemented for the control of access, service and network function [1].

Our research aims to develop efficient RAN selection algorithms in a heterogeneous wireless network environment. The remaining of this paper is organised as follows. Section 2 describes the RAN selection algorithm. Section 3 presents the performance analysis of the proposed algorithm, and section 4 concludes the paper.

II. Radio Access Optimisation Algorithm

Our RAN selection is context aware and considers service type, quality preference, terminal type, user/terminal status, available RANs, network capacity, resource availability, coverage area, and service costs. The use of the above metrics makes the selection process more complex. Therefore, we propose an efficient Radio Access Optimisation (RAO) algorithm to simplify the selection process. The algorithm considers two conflicting factors, the increased carried traffic and the user satisfaction. It also performs data rate adaptation in order to optimise the network resource usage. The goal of the RAO algorithm is to maximise an objective function, which measures the level of user satisfaction obtained from a candidate network. Each network covering a user request will be evaluated. The network that provides the greatest value for the objective function will be selected. The next subsection explains the RAO objective function.

A. RAO objective function

Assuming network A is being evaluated for a requested service, the RAO objective function (OF) can be calculated as:

$$OF(RS, X_A) = US(RS, X_A) + \sum_{i=1}^{N_{X_A}} US(S_i) + \sum_{X_j, j \neq A} \sum_{k=1}^{N_{X_j}} US(S_k)$$

where *RS* represents the requested service, X_A represents the network A, N_{X_A} is the number of the existing service sessions in network A, S_i represents the existing service sessions in network A, X_j represents the *jth* network that is not being evaluated, N_{X_j} is the number of the existing service sessions in network X_j, and S_k represents the *kth* existing service session in network X_j. The objective function includes $US(RS, X_A)$, which is the user satisfaction experience by the requesting user when network A is selected for the service *RS*, $\sum_{i=1}^{N_{X_A}} US(S_i)$ is the impact upon the user satisfaction of the existing users in the network A if *RS* is accepted, and $\sum_{X_j, j \neq A} \sum_{k=1}^{N_{X_j}} US(S_k)$ is the collective user satisfaction experienced by the services in the networks which are not being evaluated.

The function *US* is calculated as:

$$US(S, X) = SNCL(S, X) \times \sum_{i} W_{S,i} \times NORM\left(Attr_i^X\right) \qquad (2)$$

SNCL(S,X) is the *Service-Network Compatibility Level for service S in the network X*. This parameter measures the level of support a network provides for a specific service. For example, a UTRAN provides a better support for real-time services, such as speech, than a WiFi network. A WiFi network is better fitted to support non-real-time services, such as file transfer. $NORM\left(Attr_i^X\right)$ is the normalised value representing the user satisfaction obtained from network *X* if service *S* is accepted. $Attr_i^X$ represents the attribute or characteristic *i* of network *X*, such as available data rate. The value of $NORM\left(Attr_i^X\right)$ is supplied by a normalisation function that ranges from 0 to 1. The greater value it has, the higher is the level of user satisfaction. The normalisation function is presented in section 3. $W_{S,i}$ is the weight representing the importance of attribute $Attr_i^X$ to service *S*. For example, real-time services, such as speech, are specified with high requirements for delay and jitter, therefore, attributes representing delay or jitter will have greater weight. The value of $W_{S,i}$ can range from 0 to 1.

When the resources of network A are not sufficient to admit the user request, a data rate adaptation scheme is performed. This scheme considers the data rate requirement from the new user request and the data rates allocated to the existing users. The scheme decreases the data rates to obtain sufficient network resources for the new requested service. The aim of the rate adaptation is to find the best balance between the increased carried traffic and the user satisfaction. Before presenting the data rate adaptation scheme, we will introduce the parameters for evaluating network resources availability. Two networks are studied: UTRAN and WiFi.

B. UTRAN Resource Availability Evaluation

The UTRAN uses the W-CDMA technology and it is an interference-limited cellular network. In the downlink, the network resource availability is determined by the amount of base station transmission power P_{total} being consumed and the maximum power P_{max} that the base station can have. When a new service request arrives, the algorithm calculates what would be the new P_{total} if this service was accepted. If the calculated P_{total} is greater than zero and smaller than or equal to P_{max}, it indicates that the UTRAN downlink network has sufficient resources to accept the new user request. Otherwise, the downlink network does not have sufficient resources and the data rate adaptation scheme needs to be performed. For the calculation of P_{total}, please refer to [2].

C. WiFi Resource Availability Evaluation

We have developed a simple but effective way for evaluating resource availability in 802.11a/b based WiFi networks. We proposed a new parameter, the *expected number of contending packets* (e_{ncp}) over the wireless channel. e_{ncp} provides a sum of probabilities of all the services sending packets to the shared WiFi channel. The description of the model and calculation of e_{ncp} is presented in [2].

When a new connection is requested to the network, the algorithm calculates the new value of e_{ncp} considering the new requested service and the existing services. If the new e_{ncp} is less than or equal to 1, the connection can be admitted. That means, on average, there is less than one packet in contention to access the network channel. However, if the value is greater than 1, it means that some packets will content with each other and we have to consider the requested and existing service types within the network before performing any decision. If the requested and existing service types are UDP based or hybrid (coexistence of UDP based and TCP based services), the requested connection will be rejected. This is because packet collisions will cause delays and packet loss for the real-time UDP based service session. There are no more guarantees that the delay and packet loss will be acceptable according to the requirements of the services. However, if the requested and existing service types are all TCP based, the value of e_{ncp} can be viewed as the number of *'long-live' TCP sessions* [3], and the analysis method proposed by Bruno et al. in [3] can be implemented to calculate the effective transmission rate (excluding traffic and protocol overheads) of each packet generated by the requested service and the existing service sessions. If all the services can be delivered inside the constraints of the calculated effective transmission rate, then the new request can be accepted. Otherwise, it will be rejected.

D. Data Rate Adaptation Scheme

When the resources in a network are insufficient to admit a user request, a data rate adaptation scheme is performed. The data rate adaptation scheme is an iterative process and it works as follows.

The required data rate from the new user request and the data rates of the existing users served by the same network are grouped together into a vector V_X. In each iteration, the scheme selectively decreases the data rate of one user service, and the data rate vector V_X is updated with the decreased data rate. This process stops when the one of the following conditions is reached:

1. The required data rate for supporting all the users in the vector V_X is reached within the network constraints.
2. The adaptation is found infeasible.

The data rate adaptation scheme aims to effectively consume the network resources and maximise the user satisfaction.

Adaptation in a UTRAN network

Given a data rate vector V_X, when there are not sufficient resources in the UTRAN network, the users are limited by inter-cell interference and thermal noise, I_n. The base station has to increase the power to overcome I_n, and P_{total} becomes greater than P_{max}.

If the V_X leads to the above overload phase, the data rate adaptation scheme will decrease the data rates of certain users to obtain a feasible value of P_{total}. The pseudo code of the scheme is presented in Fig. 1.

When the UTRAN is in the overload phase, a subset DS_X will be extracted from V_X. The DS_X subset includes the data rates of the services which are capable of suffering degradation but still comply with minimum QoS requirements. They are the candidates for adaptation. In each round, the scheme will hypothetically decrease the data rate of each candidate service belonging to DS_X by one level and calculate the ratio of the difference of the base station power, $P_{total} - P'_{total}$, to the outcome difference of the objective function, $OF - OF'$ (difference of the values before and after decreasing the data rate). The service whose hypothetical data rate degradation results in the greatest ratio will be selected for actual rate degradation. The selected data rate is denoted as R_m. After the degradation, a new value, R'_m, is obtained. Then, R_m will be deleted from DS_X and the data rate vector V_X will be updated with R'_m. The next round of adaptation will proceed until a feasible P_{total} is reached, or no more data rates can be decreased. The adaptation scheme aims to maximise the reduction of power consumption and minimise the loss of user satisfaction.

Given data rate vector $V_X = [R_1, R_2, \ldots, R_{N_X}, R_{N_X+1}]$, calculate the Objective Function *OF*
Take V_X and calculate P_{total}
Extract the subset $DS_X = [R_1, R_2, \ldots, R_i, \ldots, R_m, \ldots]$
While the UTRAN is in the overload phase
  For each user *i* in DS_X
    Decrease the data rate R_i to a lower level as R'_i
    Form a vector $V'_X = [R_1, R_2, \ldots, R'_i, \ldots, R_{N_X+1}]$, calculate P'_{total} and OF'
    Calculate the ratio $RT(R_i)$:

$$RT(R_i) = \frac{P_{total} - P'_{total}}{OF - OF'}$$

  End For
  Among the calculated *RT*s, select the service whose degraded data rate (R_m) supplies the greatest RT value: $R_m = argmaxRT(R_i)$
  Delete R_m from DS_X
  Form a new $V_X = [R_1, R_2, \ldots, R'_m, \ldots, R_{N_X+1}]$
  Take V_X and calculate P_{total}
End While

Adaptation in a WiFi network

Given a data rate vector V_X and the characteristics of the services, the *expected number of contending packets* e_{ncp} can be obtained. When there are not sufficient resources in the WiFi network, the value of e_{ncp} will depend on the following two overload phases:

- *Hybrid overload phase*: In this case, the service types are hybrid. The users are limited by packet collisions. The packet collisions will cause delays and packet loss for the real-time UDP based service sessions and there are no more guarantees that the delay and packet loss will be acceptable according to the requirements of the services. e_{ncp} is greater than 1.
- *Non-real-time service overload phase*: In this case, the service types are all TCP based. The users are limited by packet contention to access the network channel. The effective packet transmission rate and the end-to-end bandwidth calculated according e_{ncp} cannot satisfy the lowest service level.

If the WiFi network enters the hybrid overload phase similarly to the scheme in the UTRAN, the data rate adaptation scheme will selectively decrease the data rates of certain real-time services in order to obtain a value of e_{ncp} less than 1. The data rate of some real-time services can be lowered by adjusting their encoders. The pseudo code of the scheme is presented in Fig. 2.

However, the adaptation scheme will not be applied to the non-real-time service overload phase. This is because, the data rates of non-real-time services depend on the channel contention in the WiFi network, but not the encoder. Also, the TCP advertised window is assumed to be 1 [3] and no further adjustment is available.

The adaptation scheme aims to minimise the channel contention and maximise user satisfaction.

Given data rate vector $V_X = [R_1, R_2, \ldots, R_{N_X}, R_{N_X+1}]$, calculate the Objective Function *OF*

Hybrid overload phase:

Take V_X and service characteristic, calculate e_{ncp}
Extract the subset $DS_X = [R_1, R_2, \ldots, R_i, \ldots, R_m, \ldots]$
While the WiFi network is in the hybrid overload phase
  For each user *i* in DS_X
    Decrease the data rate R_i to a lower level as R'_i
    Form a vector $V'_X = [R_1, R_2, \ldots, R'_i, \ldots, R_{N_X+1}]$, calculate e'_{ncp} and OF'
    Calculate the ratio $RT(R_i)$:

$$RT(R_i) = \frac{e_{ncp} - e'_{ncp}}{OF - OF'}$$

  End For
  Among the calculated *RT*s, select the service whose degraded data rate (R_m) supplies the greatest RT value: $R_m = argmaxRT(R_i)$
  Delete R_m from DS_X
  Form a new $V_X = [R_1, R_2, \ldots, R'_m, \ldots, R_{N_X+1}]$
  Take V_X and characteristics, calculate e_{ncp}
End While

Fig. 2. Pseudo Code of the Data Rate Adaptation Scheme for WiFi

III. Radio Access Network Selection Simulation

In order to compare the network performance obtained from the different RAN selection algorithms, we have implemented call level simulations. We simulate a heterogeneous communication environment with two RANs, UTRAN and WiFi. The objective is to gauge the performance of the RAO algorithm and compare with the Multiservice Vertical Handoff Decision Algorithm (MUSE-VDA) presented in [4]. The RAO algorithm is not confined to UTRAN and WiFi networks, but it can be performed in more complex situations where more technologies are involved. The way in which the objective function is adopted to measure and evaluate the UTRAN and the WiFi network can be applied to other wireless network technologies.

There are six types of services: Speech (SP), Video Call (VC), Video Streaming (VS), Audio Streaming (AS), Web Browsing (WB), and File Transfer (FT). The first four services are real-time and UDP based. The others are non-real-time and TCP based. The services can belong to two service classes: basic and premium. The basic service class has a lower data rate requirement, which provides the minimum quality constraint and threshold that the service should meet. The premium service class has a higher data rate requirement, which provides better service quality when resources are available. The number of users for each service type, the service parameters and their typical values are listed in Table 1. For the UDP based real-time services, the data rate values are fixed. For the TCP based non-real-time services, the data rate values are the minimal requirements for each service class.

The parameters of the user satisfaction function (*US*) of each service type are presented in Table 2.

The value of SNCL ranges from 0 to 1, where 0 means minimum compatibility and 1 means maximum compatibility. For example, we define the SNCL of the speech service in an UTRAN network as *high*, and in the WiFi network as *very low*. These linguistic terms are converted to crisp numbers. For the linguistic terms '*high*' and '*very low*', the corresponding crisp numbers are *0.717* and *0.091*, respectively. For a detailed explanation on this conversion process, please refer to [5].

TABLE I. Parameters and Typical Values

Service	SP	VC	VS	AS	WB	FT
Required Data Rate (kbps)	6.7/12.2	64/128	64/128	32/64	128/256	64/128
Protocol	UDP	UDP	UDP	UDP	TCP	TCP
Number of Users	5	5	5	5	10	10

For speech, video call, video and audio streaming services, the *data rate* attribute provided by the network is normalised based on a Sigmoid function, when the minimum required data rate is achieved. The Sigmoid function has been used before to estimate user satisfaction (perceived QoS) [6]. As the data rate increases, the user satisfaction also increases. A Sigmoid curve has a convex and a concave characteristic. When the data rate is quite low or very high for these types of service, an increase of data rate will not improve the user satisfaction significantly. This is because, at low data rates, the increase in data rate needs to be significant in order to change the perceived QoS by the user. At high data rates, the perceived QoS is good or excellent, an increase in data rate will hardly affect further the user's perception. Web browsing and file transfer services possess a bursty pattern. The increase in data rate has a significant positive effect on the user satisfaction up to the high data rate values. For these service types, the *data rate* attribute is normalised by an Exponential function [7].

The mobility support provided by the WiFi network is

TABLE II. Parameters of User Satisfaction (US) Functions

Service Type	**Service/Network Compatibility Level (*SNCL*)**	**Considered Attributes**	**Normalisation Function**	**Weights**
SP	*SNCL(VC, UTRAN)* = 0.717 *SNCL(VC, WiFi)* =0.091	Supplied date rate ***dr*** (kbps); mobility support ***m***	$NORM(dr) = \begin{cases} 0, & dr < 6.7 \\ 0.3 + \frac{1}{2+2\times e^{[illegible]}}, & dr \geq 6.7 \end{cases}$ $NORM(m, WiFi) = 0.091$ $NORM(m, UTRAN) = 1$	$Weight_{dr} = 0.717$ $Weight_{m} = 0.5$
VC	*SNCL(VC, UTRAN)* = 0.717 *SNCL(VC, WiFi)* =0.091	Supplied date rate ***dr***; mobility support ***m***	$NORM(dr) = \begin{cases} 0, & dr < 64 \\ 0.3 + \frac{1}{2+2\times e^{[illegible]}}, & dr \geq 64 \end{cases}$ $NORM(m, WiFi) = 0.091$ $NORM(m, UTRAN) = 1$	$Weight_{dr} = 0.717$ $Weight_{m} = 0.5$
VS	*SNCL(VS, UTRAN)* = 0.717 *SNCL(VS, WiFi)* =0.091	Supplied date rate ***dr***; mobility support ***m***	$NORM(dr) = \begin{cases} 0, & dr < 16 \\ \frac{1}{1+e^{[illegible]}}, & dr \geq 16 \end{cases}$ $NORM(m, WiFi) = 0.091$ $NORM(m, UTRAN) = 1$	$Weight_{dr} = 0.909$ $Weight_{m} = 0.5$
AS	*SNCL(AS, UTRAN)* = 0.717 *SNCL(AS, WiFi)* = 0.283	Supplied date rate ***dr***; mobility support ***m***	$NORM(dr) = \begin{cases} 0, & dr < 32 \\ \frac{1}{1+e^{[illegible]}}, & dr \geq 32 \end{cases}$ $NORM(m, WiFi) = 0.091$ $NORM(m, UTRAN) = 1$	$Weight_{dr} = 0.909$ $Weight_{m} = 0.283$
WB	*SNCL(WB, UTRAN)*=0.5 *SNCL(WB, WiFi)* =1	Supplied date rate ***dr***; mobility support ***m***	$NORM(dr) = \begin{cases} 0, & dr < 32 \\ 1 - e^{[illegible]}, & dr \geq 32 \end{cases}$ $NORM(m, WiFi) = 0.091$ $NORM(m, UTRAN) = 1$	$Weight_{dr} = 0.717$ $Weight_{m} = 0.091$
FT	*SNCL(FT, UTRAN)* =0.283 *SNCL(FT, WiFi)* =0.909	Supplied date rate ***dr***; mobility support ***m***	$NORM(dr) = 1 - e^{[illegible]}$ $NORM(m, WiFi) = 0.091$ $NORM(m, UTRAN) = 1$	$Weight_{dr} = 0.717$ $Weight_{m} = 0.091$

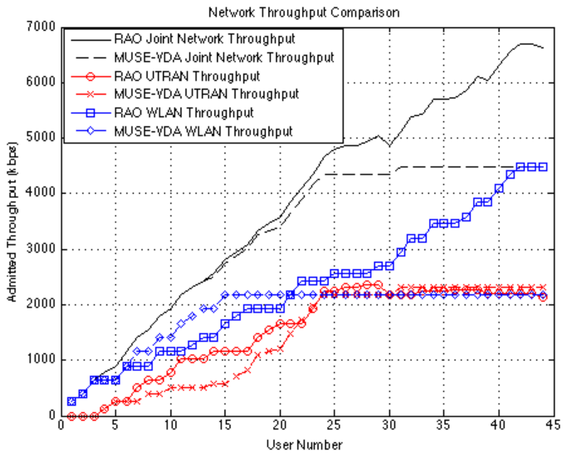

Fig. 3. Network Throughputs of Different Algorithms

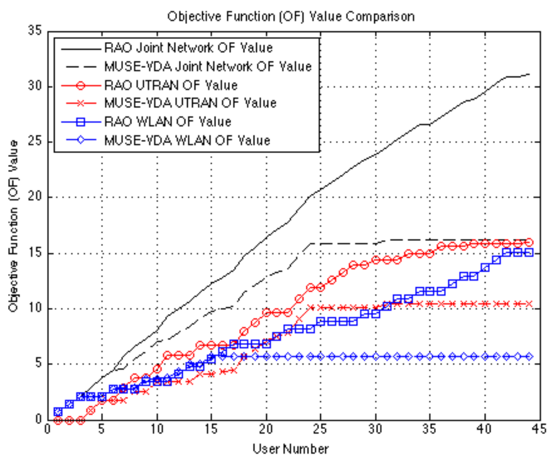

Fig. 4. Objective Function Values of Different Algorithms

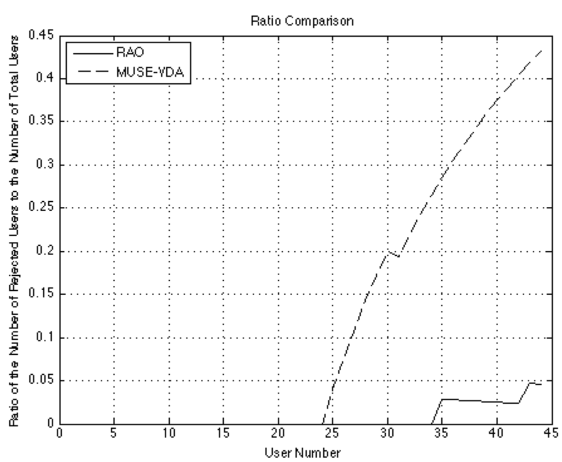

Fig. 5. Ratio of Blocked Requests to the Total Number of Requests of Different Algorithms

defined as *very low*, and by the UTRAN network as extremely high. These linguistic terms are converted to the crisp numbers 0.091 and 1, respectively. Table 2 also shows examples for defining the weighting values.

In the simulation, users start the services gradually. When a service request is received, the RAO and MUSE-VDA RAN selection algorithms assign the service to an appropriate network and adjust the data rates if necessary. We compare the performance of the algorithms in three aspects: network throughput, objective function value, and the ratio of blocked requests to the total number of requests.

Fig. 3 depicts the throughputs of the joint UTRAN and WiFi networks, the UTRAN network, and the WiFi network, when the different RAN selection algorithms are implemented. Fig. 4 presents the values of the objective function in relation to the number of accepted users. The value of the objective function represents the user satisfaction. The greater is the value of the objective function, the higher is the level of user satisfaction. The simulation results show that the RAO algorithm provides greater throughput and higher user satisfaction. The MUSE-VDA algorithm can just perform better at network throughput in the UTRAN network. By implementing the RAO algorithm, the overall network resources are effectively used to carry more traffic and improve the user satisfaction. Fig. 5 depicts the ratio of blocked requests to the total number of requests received by the joint networks (UTRAN and WiFi). Before the 25th service request arrives, the ratio in all algorithms is zero. When MUSE-VDA is used, the ratio increases after the 25th service request and reaches 43% in the end of simulation time. In contrast, when RAO algorithm is used, the ratio remains zero, until the arrival of the 35th service request. In the end, the ratio generated by RAO reaches around 5%, which is significantly lower than the ratio generated by MUSE-VDA. The use of RAO algorithm provides a good performance, because it can dynamically and properly adjust the service classes of the existing users, and allow more requests to be admitted.

IV. Conclusion

In this paper, we first introduce the trend and requirements for an integrated heterogeneous network system in the next generation mobile networks. We also present a RAN selection algorithm to simplify the selection process and consider the increased resource usage and the user satisfaction. The simulation results show that the RAO algorithm can effectively use network resources, improve user satisfaction, and admit more requests. In future research, we will study the performance of the RAO algorithm in more complex scenarios and compare RAO against other possible algorithms.

References

[1] Board of NGMN Limited. Next Generation Mobile Networks Beyond HSPA & EVDO A White Paper http://www.ngmn-cooperation.com.

[2] Weizhi Luo and Eliane Bodanese, “Optimizing Access in an Integrated Wireless Network Environment”, Proceeding of SPECTS 2008.

[3] R. Bruno, M. Conti and E. Bregori, “Modeling TCP Throughput Over Wireless LANs”. Proc. 17th IMACS World Congress Scientific Computation, Applied Mathematics and Simulation, 2005.

[4] Fang Zhu and Janise McNair, “Multiservice Vertical Handoff Decision Algorithms.”, EURASIP Journal on Wireless Communications and Networking, Vol. 2006, pp. 1-13

[5] Shu-Jen Chen and Ching-Lai Hwang, Fuzzy Multiple Attribute Decision Making: Methods and Applications, Springer-Verlag, 1992, pp. 465-486.

[6] Mainak Chatterjee, et al,, “Rate Allocation and Admission Control for Differentiated Services in CDMA Data Networks”, IEEE Transactions on Moblie Computing, vol. 6, NO. 2, February 2007.

[7] Veselin Rakocevic, et al., “Performance Analysis of Bandwidth Allocation Schemes in Multisevice IP Networks using Utility Functions”, Proceedings of the 17th International Teletraffic Congress, 2001.