Radio Access Network Selection in a Heterogeneous Communication Environment

Weizhi Luo
Queen Mary, University of London, UK
weizhil.luo@elec.qmul.ac.uk

Eliane Bodanese
Queen Mary, University of London, UK
eliane.bodanese@elec.qmul.ac.uk

***Abstract*— In recent years, a variety of radio access networks have been developed and deployed, including UMTS, WiFi, and WiMAX. These networks overlap each other in many areas and create heterogeneous wireless communication environments, which can enable seamless communications, joint resource management and adaptive quality of service. For example, operators would not need to reject user requests, but redirect them to the most appropriate networks. The redirection could be through a different radio access network. However, such wireless heterogeneous system still has many pending issues to solve. One is the selection of the most appropriate radio access network when receiving a service request. This paper addresses this issue by proposing an adaptive and efficient algorithm. The simulation results show that the proposed radio access network selection algorithm can improve the network performance and capacity.**

I. Introduction

Nowadays, multiple Radio Access Networks (RANs) are available as commercial wireless systems. The RANs can be classified into three main categories: the mobile cellular networks (2G and 3G), the wireless local area networks (e.g. WiFi) and the wireless metropolitan area networks, such as WiMAX. In order to achieve seamless communications and joint resource management in a heterogeneous environment, interworking among networks is necessary. One solution is to introduce an IP-based backbone network. For example, the Simplified Architecture Evolution (SAE) project aims to develop a framework for supporting the 3GPP Long Term Evolution requirements [1]. The SAE concentrates on a packet switched (PS) core. The access to the 3GPP network is not restricted to the UMTS Terrestrial RAN (UTRAN), but it can even be via WiFi, WiMAX or wired technologies.

A PS core enables intelligence at the network edge and supports various business models [2]. An IP-based internetworking architecture enables providing diversified and flexible services which can fulfil different user requirements [3]. It also simplifies the network integration and provides a flatter network architecture [4]. Such an internetworking architecture is shown in Fig. 1. It includes different RANs, such as the evolved UTRAN, 2G/3G RAN, WiFi, and WiMAX. The infrastructure of the networks are maintained without modifications. The evolved UTRAN, the WiFi and the WiMAX networks can be directly connected to the PS core. In some situations (e.g. for security reasons [1]), a gateway may be introduced between the WiFi/WiMAX networks and the PS core. RANs like 2G RAN can access the PS core via an access router or a gateway, such as the Serving GPRS Service Node.

An integrated communication system based on the above

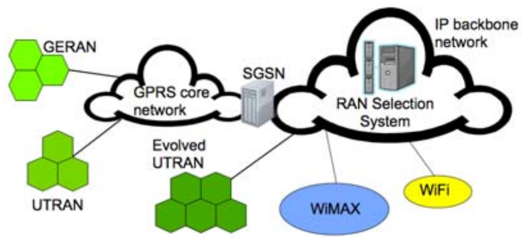

Fig. 1. Internetworking Architecture

internetworking architecture needs an intelligent RAN selection system in order to function efficiently. In [5], Zhu and McNair use a cost function to evaluate and compare available networks and select the most appropriate one to handoff to. A network-based access and interface selection system is introduced by Koundourakis et al. in [6]. Their work concentrates on optimally utilising overall system resources and providing acceptable QoS to end users. In [7], we propose that a RAN selection system should be based on the context information of the user/terminal and the network. The user/terminal context information includes the requested service, quality preference, terminal type, and user/terminal status. The network context information consists of available RANs, network capacity, resource availability, coverage area, and service costs. Currently, we propose a network-based system for RAN selection. As shown in Fig. 1, the selection system resides in the PS core. It receives and updates user/terminal and network context information, and accepts user service requests. Based on the context information, it implements an access optimisation algorithm to generate an optimal selection and then transfers the result to the user.

The remainder 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 algorithm 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 utilisation. The goal of the RAO algorithm is to maximise an objective function, which measures the benefits obtained from a candidate network. Each network covering a user request will be evaluated. The network that provides *overall* the greatest value for the objective function will be selected, if the request can be accepted. 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:

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

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 is the index representing the existing services, X_j is the index representing the networks that are not being evaluated, N_{X_j} is the number of the existing service sessions in network *j*, and S_k is the index representing the services provided by X_j. The objective function includes $US(RS, X_A)$, which is the user satisfaction experienced by the requested user when selecting network A for the requested service *RS*, $\sum_{i=1}^{N_{X_A}} US(S_i)$ is the impact upon the existing users in the network A, and $\sum_{X_j, j\neq A} \sum_{k=1}^{N_{X_j}} US(S_k)$ is the user satisfaction experienced by the services in the networks which are not being evaluated.

The user satisfaction *US* is calculated as:

$$US(S, X) = SNCL(S, X) \times \sum_i W_{S,i} \times NORM\left(Attr_{S,i}^X\right) \quad (2)$$

SNCL(S, X) is the *Service-Network Compatibility Level of service S in the network X*. This parameter measures the level of support a network provides for a specific service. For example, the UTRAN provides a better support for real-time services, such as speech, than the WiFi network. The WiFi network is better fitted to support non-real-time services, such as file transfer. $NORM\left(Attr_{S,i}^X\right)$ is the normalised value of a certain attribute *i* of service *S*, which represents the QoS provided by network *X*. The attribute *i* can be provided data rate, data transmission delay, etc. The normalised value ranges from 0 to 1 and the normalisation function will be presented later in this paper. $W_{S.i}$ is the weight representing the importance of attribute $Attr_{S,i}^X$ 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 weights. The value of $W_{S.i}$ can range from 0 to 1.

When the resource of network A is not sufficient to accept a 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 resource for the new requested service. The aim of the rate adaptation is to find out the best balance between the increased carried traffic and the user satisfaction. Before presenting the data rate adaptation scheme, we will present the methods for evaluating network resource availability. Two networks are studied: UTRAN and WiFi.

B. UTRAN Resource Availability Model

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 being consumed and the maximum power that the base station can use. To calculate the value of the transmission power being consumed, we first need to define the E_b/N_0 (the ratio of energy per user bit to the noise spectral density). For user *n* in the UTRAN cell, we derive the value of E_b/N_0 as:

$$\left(E_b/N_0\right)_n = \frac{W}{\upsilon_n \times R_n} \times \frac{P_n \times L_{BS,n}}{P_{total} \times (1-\alpha) \times L_{BS,n} + I_n} \quad (3)$$

where *W* is the chip rate of W-CDMA; υ_n is the activity factor of user *n*; R_n is the bit rate of user *n*; P_n is the base station transmission power for user *n*; $L_{BS,n}$ is the downlink attenuation between base station and user *n*; P_{total} is the total power being consumed by the base station; α is average orthogonality factor in the cell whose value ranges from 0.4 to 0.9 (1 means totally orthogonal); I_n is the thermal noise and inter-cell interference received by user *n*.

P_n can be calculated as:

$$P_n = \frac{(E_b/N_0)_n \times v_n \times R_n \times \lfloor (1-\alpha) \times L_{BS,n} \times P_{total} + I_n \rfloor}{W \times L_{BS,n}} \quad (4)$$

By summing up the transmission power of base station for every individual user, the total transmission power of base station (P_{total}) can be derived from (4) as:

$$P_{total} = \sum_{n=1}^{N} \frac{(E_b/N_0)_n \times v_n \times R_n \times \lfloor (1-\alpha) \times L_{BS,n} \times P_{total} + I_n \rfloor}{W \times L_{BS,n}} \quad (5)$$

Isolating P_{total} we have:

$$P_{total} = \frac{\sum_{n=1}^{N} \frac{(E_b/N_0)_n \times v_n \times R_n \times I_n}{W \times L_{BS,n}}}{1 - \frac{1-\alpha}{W} \times \sum_{n=1}^{N} (E_b/N_0)_n \times v_n \times R_n} \quad (6)$$

If P_{total} is greater than zero and smaller than or equal to the maximum base station transmission power P_m, it indicates that the UTRAN downlink network has sufficient resource to accept the new user request. Otherwise, the downlink network does not have sufficient resource and the data rate adaptation scheme needs to be performed.

C. *WiFi Resource Availability Model*

We have developed a simple but effective solution for evaluating capacity and resource availability in 802.11a/b based WiFi networks. We proposed a new parameter, the *expected number of contending packets* over the wireless channel, which is denoted by e_{ncp}. The full description of the model is presented in [7]. In this paper, we present the equation to calculate e_{ncp} and summarise the main information about the model.

$$e_{ncp} = \sum_{i=0}^{N} i \times \binom{N}{i} \times {e_p}^{i} \times (1-e_p)^{(N-i)} = N \times e_p \quad (7)$$

N is the number of existing connections in the network plus the new requested connection(s). e_p is the probability that the channel is occupied by a packet transmission. We calculate e_p by the equation below:

$$e_p = \sum_{s\_t} (p\_on_{s\_t} \times n_{s\_t} / N) \quad (8)$$

s_t represents the service type and $p\_on_{s\_t}$ is the probability that the channel is occupied by the transmission of the packets belonging to the service type *s_t*. n_{s_t} is the number of connections of service type *s_t*.

The calculation of $p\_on_{s\_t}$ is service type dependent. The metrics include requested data rate and service characteristics (e.g. packet payload size, packet arrival interval). For full explanation, please refer to [7].

Assuming a new connection is made to the network, if 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 collide with each other and we have to consider the requested and existing service types within the network before performing any action. If the requested and existing service types are UDP based or hybrid (coexistence of UDP and TCP based services), the connection will be rejected. This is because packet collisions will cause delays and packet loss for the real-time UDP based service sessions. 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, by viewing the value of e_{ncp} as the number of '*long-live*' TCP sessions (please refer to [8]), the analysis method proposed in [8] can be implemented to calculate the effective transmission rate (excluding traffic and protocol overheads) of each packet generated by the requested connection and the existing users. Bruno et al. [8] investigated the performance of the TCP connections over WiFi networks. They assumed the size of the TCP advertised window as equal to one. This assumption ensures the TCP flows will have a fair access to the channel bandwidth. The calculation of the effective transmission rate is described in [7] and [8].

In a situation where the service types are hybrid and e_{ncp} is equal to or less than 1, the effective bandwidth for the requested connection and the existing users can be obtained as follows: As e_{ncp} is not greater than 1, on average there is less than one packet in contention to access the network channel. The quality requirements for UDP based services can be fulfilled. The effective bandwidth values for the UDP based services are identical to their transmitting rates. For the TCP based services, each packet can be transmitted at the maximum effective rate of about 4400 kbps (according to the proposed analytical method and simulation results in [8]). This is because, on average, there is less than one packet in contention to access the network channel.

D. *Data Rate Adaptation Scheme*

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

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

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

1) *Adaptation in a UTRAN network:*

Given a data rate vector V_X, the required base station power P_{total} can be obtained according to (6). When there is not sufficient resource in the UTRAN network, the value of P_{total} will depend on the following two overload phases:

- *Negative overload phase*: In this case, the users are limited by intra-cell interference, which cannot be overcome by increasing the base station power. In (6), $\frac{1-\alpha}{W} \times \sum_{n=1}^{N} (E_b/N_0)_n \times v_n \times R_n$ (denoted as D_{total} for simplicity) will be greater than 1; P_{total} becomes negative because the data rates required by all the users in the cell are too high and/or the number of users is too large. Infinitely increasing the base station power will not satisfy the demands.

- *Positive overload phase*: In this case, the users are limited by inter-cell interference and thermal noise, I_n. The base station has to increase its power P_n to overcome I_n. When there is not sufficient resource, P_{total} becomes greater than P_m .

If the V_X leads to an overload phase, the data rate adaptation scheme will decrease the data rates of certain users so as to obtain a feasible value of P_{total}. The pseudo code of the scheme is presented in Fig. 2.

When the UTRAN is in the negative overload phase, the scheme will extract a subset DS_X from V_X . The DS_X subset includes the data rates of the services which are capable to suffer degradation but still comply with minimum QoS requirements. They are candidates for adaptation. In each round, the scheme will theoretically decrease the data rate of each candidate service belonging to DS_X by one level and calculate the ratio of D_{total} - D'_{total} to the outcome difference of the objective function, *OF – OF'* (difference of the values before and after decreasing the data rate). The ratio for user service *i* is denoted as $RT(R_i)$. The service whose theoretical data rate degradation results in the greatest ratio, denoted as R_m, will be selected for actual degradation. After 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 the UTRAN moves to the positive overload phase or a feasible P_{total} is reached or no more data rates can be decreased.

When the UTRAN is in the positive overload phase, similarly to the process in the negative overload phase, a subset DS_X will be extracted from V_X . In each round, the scheme will theoretically 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'*. The candidate service whose degraded data rate results in the greatest ratio (R_m) will be selected for actual rate degradation. Then, R_m will be deleted from DS_X and the data rate vector V_X is updated with R'_m and the next round of adaptation will proceed until a feasible P_{total} is reached or no more data rates can be decreased. In both overload phases, the adaptation scheme aims to maximise the reduction of power consumption and minimise the loss of user satisfaction.

2) Adaptation in a WiFi network:

Given a data rate vector V_X and service characteristics, the expected number of contending packets e_{ncp} can be obtained. When there is not sufficient resource 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.

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

Negative overload phase:

Take V_X and calculate D_{total} according to (6)
Extract the subset $DS_X = [R_1, R_2, ...R_i, ...R_m, ...]$
While the UTRAN is in the negative overload phase
  For each service i in DS_X
    Decrease the data rate R_i to a lower level as R'_i
    Form a vector $V'_X = [R_1, R_2, ..., R'_i, ..., R_{N_X+1}]$, calculate D'_{total} and *OF'*
    Calculate the ratio $RT(R_i)$:

$$RT(R_i) = \frac{D_{total} - D'_{total}}{OF - OF'}$$

  End For
  Among the calculated *RT*s, select the service whose degraded data rate (R_m) supplies the greatest RT value: $R_m = \arg\max RT(R_i)$
  Delete R_m from DS_X
  Form a new $V_X = [R_1, R_2, ..., R'_m, ..., R_{N_X+1}]$
  Take the new V_X and calculate P_{total} according to (6)
End While

Positive overload phase:

Take V_X and calculate P_{total} according to (6)
Extract the subset $DS_X = [R_1, R_2, ...R_i, ...R_m, ...]$
While the UTRAN is in the positive overload phase
  For each service i in DS_X
    Decrease the data rate R_i to a lower level as R'_i
    Form a vector $V'_X = [R_1, R_2, ..., R'_i, ..., R_{N_X+1}]$, 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: $R_m = \arg\max RT(R_i)$
  Delete R_m from DS_X
  Form a new $V_X = [R_1, R_2, ..., R'_m, ..., R_{N_X+1}]$
  Take the new V_X and calculate P_{total} according to (6)
End While

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

- *Non-real-time service overload phase*: In this case, the service types are all TCP based. The users are limited by packets 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, similar to 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. 3.

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 [8] and no further adjustment is available.

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

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

Hybrid overload phase:

Take V_X and service characteristics, calculate e_{ncp}
Extract the subset $DS_X = [R_1, R_2, ...R_i, ...R_m, ...]$
While the WiFi network is in the hybrid overload phase
  For each service i in DS_X
    Decrease the data rate R_i to a lower level as $R_i^{'}$
    Form a vector $V_X^{'} = [R_1, R_2, ..., R_i^{'}, ..., R_{N_X+1}]$, calculate $e_{ncp}^{'}$ and OF'
    Calculate the ratio $RT(R_i)$:

$$Rt_n = \frac{e_{ncp} - e_{ncp}^{'}}{OF - OF'}$$

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

Fig. 3. Pseudo Code of the Data Rate Adaptation Scheme for WiFi

TABLE I. PARAMETERS AND TYPICAL VALUES

Service Type	VC	VS	AS	WB	FT
Required Data Rate (kbps)	64/128	64/128	32/64	128/256	64/128
Transport Protocol	UDP	UDP	UDP	TCP	TCP
Number of Users	7	13	2	13	9

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 algorithms (A1 and A2) presented in [7]. The RAO algorithm is not confined to UTRAN and WiFi networks, but it can be performed in more complex situations. There are five types of services: Video Call (VC), Video Streaming (VS), Audio Streaming (AS), Web Browsing (WB), and File Transfer (FT). The first three 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 sufficient. 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.

TABLE II. Parameters of USER SATISFACTION (US) FUNCTIONS

Service Type	Service-Network Compatibility Level (*SNCL*)	Considered Attributes	Normalisation Function	Weights
VC	*SNCL(VC, UTRAN)* = 0.717 *SNCL(VC, WiFi)* =0.091	Supplied date rate ***dr***; delay ***d***; jitter ***j***; mobility support ***m***	$NORM(dr)=\begin{cases}0, & dr<64\\ 0.5+\frac{1}{2+2*e^{-2*(dr-9.6)}}, & dr\geq 64\end{cases}$ $NORM(d)=\frac{1}{1+e^{0.04*(d-275)}}$ $NORM(j)=\begin{cases}0.5+\frac{0.5}{1+e^{0.55*(j-40)}}, & j\leq 60\\ \frac{0.5}{1+e^{0.55*(j-75)}}, & j>60\end{cases}$ *NORM(m, WiFi)* = 0.091 *NORM(m, UTRAN)* = 1	$Weight_{dr}=0.717$ $Weight_{d}=0.717$ $Weight_{j}=0.717$ $Weight_{m}=0.5$
VS	*SNCL(VS, UTRAN)* = 0.717 *SNCL(VS, WiFi)* =0.091	Supplied date rate ***dr***; mobility support ***m***	$NORM(dr)=\begin{cases}0, & dr<16\\ \frac{1}{1+e^{-0.06*(dr-48.5)}}, & dr\geq 16\end{cases}$ *NORM(m, WiFi)* = 0.091 *NORM(m, UTRAN)* = 1	$Weight_{dr}=0.909$ $Weight_{m}=0.283$
AS	*SNCL(AS, UTRAN)* = 0.717 *SNCL(AS, WiFi)* = 0.283	Supplied date rate ***dr***; mobility support ***m***	$NORM(dr)=\begin{cases}0, & dr<32\\ \frac{1}{1+e^{-0.1743*(dr-24.25)}}, & dr\geq 32\end{cases}$ *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***	$NORM(dr)=\begin{cases}0, & dr<32\\ 1-e^{-5.85*dr/384}, & dr\geq 32\end{cases}$ *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***	$NORM(dr)=\begin{cases}0, & dr<32\\ 1-e^{-11.4*dr/384}, & dr\geq 32\end{cases}$ *NORM(m, WiFi)* = 0.091 *NORM(m, UTRAN)* = 1	$Weight_{dr}=0.717$ $Weight_{m}=0.091$

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 the crisp numbers. For the linguistic term '*high*' and '*very low*', the crisp numbers are *0.717* and *0.091*. For detailed explanation on this conversion process, please refer to [9].

For 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) [10]. As the data rate increases, the user satisfaction also increases. As shown in Fig. 4, a Sigmoid curve has a convex and a concave characteristic. When the data rates are quite low or very high for these types of services, 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 further the user's perception.

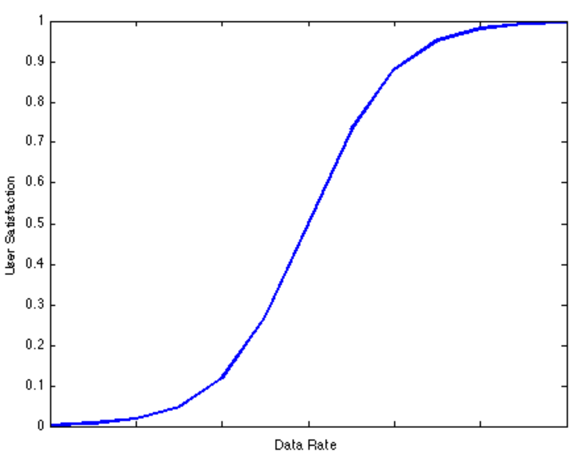

Fig. 4. Sigmoid Function

We also use the Sigmoid function to normalise the *delay* and *jitter* attributes. However, in contrast to the data rate, as the delay or the jitter increases, the user satisfaction deteriorates.

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. Therefore, for these service types, the *data rate* attribute is normalised by an Exponential function [11].

The mobility support provided by the WiFi network is 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 RAN selection algorithms assign the service to an appropriate network and adjust the

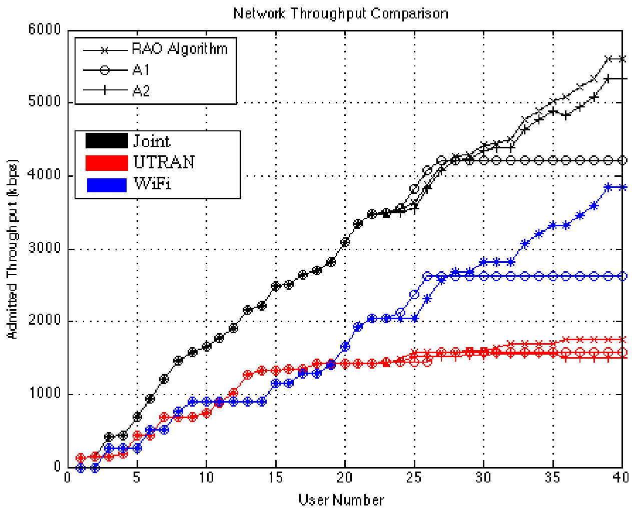

Fig. 5. Network Throughputs of Different Algorithms

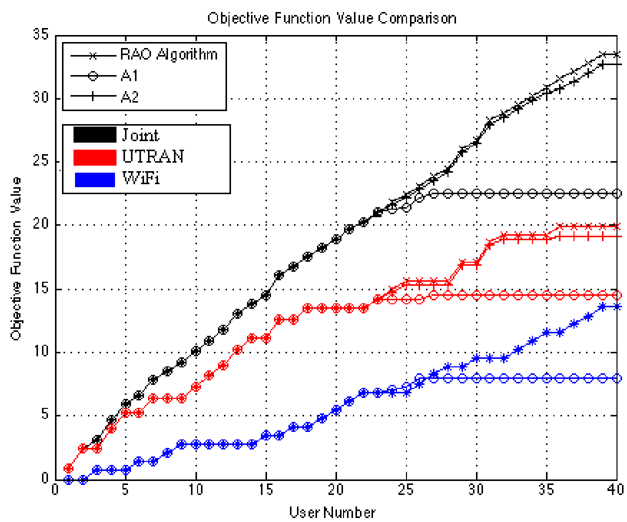

Fig. 6. Objective Function Values of Different Algorithms

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. 5 depicts the joint (UTRAN and WiFi), UTRAN, and WiFi network throughputs when different RAN selection algorithms are implemented. Fig. 6 presents the values of objective function in relation to the number of accepted users. These simulation results show that the RAO algorithm outperforms the A1 and A2 algorithms. The use of the RAO algorithm provides greater throughput and higher user satisfaction. The A2 algorithm can achieve the same performance as RAO only in the WiFi network. By implementing the RAO algorithm, the overall network resources are effectively used to carry more traffic and the user satisfaction is improved.

Fig. 7 depicts the ratio of blocked requests to the total

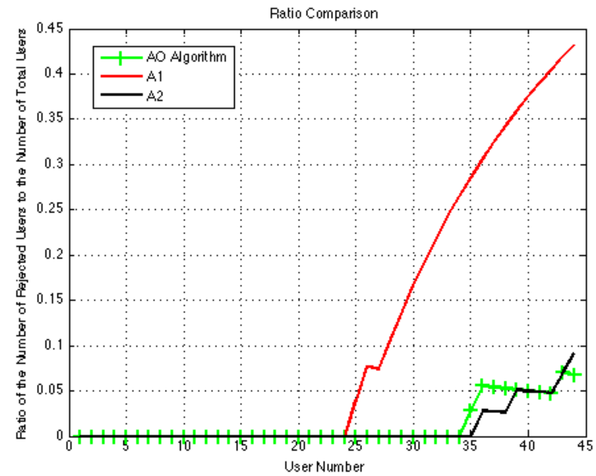

Fig. 7. Ratio of Blocked Requests to the Total Number of Requests of Different Algorithms

number of requests received by the networks (when using the different RAN selection algorithms). Before the 25th service request arrives, the ratio in all algorithms is zero. When A1 is used, the ratio increases after the 28th service request and reaches 43% in the end of simulation time. In contrast, when A2 and RAO algorithms are used, the ratio remains zero, until the arrival of the 36th and 35th service requests, respectively. In the end, the ratio generated by A2 reaches 9%. For the RAO algorithm, the ratio reaches 7%, which is lower than the ratios generated by A1 and A2. 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 architecture of an integrated communication system for the next generation mobile networks. We also propose a RAN selection scheme which can facilitate seamless communications, joint resource management, and adaptive quality of service. The RAO algorithm presented in section 2 simplifies the selection process and considers the increased resource utilisation 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 the RAO algorithm against other proposed algorithms in the field.

References

[1] 3GPP. (2006, 4 Oct.). UTRA-UTRAN Long Term Evolution (LTE) and 3GPP System Architecture Evolution (SAE) [Online]. Available: http://www.3gpp.org/Highlights/LTE/LTE.htm

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

[3] 3GPP. TR 22.978 v7.1.0 (2005-06) Technical Report, “All-IP Network (AIPN) feasibility study (release 7)”.

[4] P. Bosch, et al., "Flat Cellular (UMTS) Networks." in the proceeding of the 2007 IEEE Wireless Communications & Networking Conference, pp. 3861-3866.

[5] G. Koundourakis, et al., "Network-based Access Selection in Composite Radio Environment.", in the proceeding of the 2007 IEEE Wireless Communications & Networking Conference, pp. 3877-3883

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

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

[8] 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.

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

[10] M. 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

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