Optimising Access in an Integrated Wireless Network Environment

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

***Abstract*—Some of the important advantages of having an integrated mobile network environment would be seamless communications, joint resource management and adaptive quality of service. In such environment, operators would not need to reject the user requests, but redirect them to appropriate networks. However, the sought aims of an integrated 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 networks. We aim to develop efficient RAN selection algorithms inside a realistic internetworking system architecture. Also, as part of our research, we developed models for UMTS and WiFi networks which facilitate the evaluation of the network resource availability in these networks, which provide a more realistic input of network context information into RAN selection algorithms. The simulation results show that our RAN selection algorithms can improve the network performance.**

***Index Terms*—Intelligent radio access network selection, internetworking, resource availability evaluation**

I. Introduction

Nowadays, multiple Radio Access Technologies (RATs) coexist including GSM/GPRS, UMTS, IEEE 802.11 based Wireless Local Area Network (e.g. WiFi) and IEEE 802.16 based Wireless Metropolitan Area Network (e.g. 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 and an internetworking system architecture. Also, as part of our research, we developed models for UMTS and WiFi networks which facilitate the evaluation of the network resource availability in these networks when carrying more realistic scenarios of service requests. Realistic network context input is important when evaluating RAN selection algorithm's performance.

This paper is organized as follows. Section 2 briefly describes our proposed system architecture. Section 3 explains the methods for calculating resource availability of different networks. Section 4 describes two RAN selection algorithms. Section 5 presents a performance analysis of the proposed

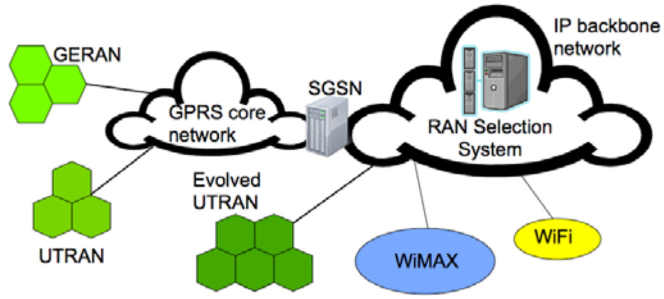

Fig. 1. Internetworking Architecture

algorithms and section 6 concludes the paper.

II. Proposed System Architecture

We proposed an internetworking architecture (Fig. 1) with different RANs (e.g. evolved UMTS Terrestrial RAN (UTRAN), 2G/3G RAN, WiFi, and WiMAX). The infrastructures of the networks are maintained without modifications. The evolved UTRAN, the WiFi and the WiMAX networks can be directly connected to the IP backbone network. In some situations (e.g. for security reasons [2]), a gateway may be introduced between the WiFi/WiMAX networks and the IP backbone network. RANs like 2G RAN can access the IP backbone network via an access router or a gateway, such as the Serving GPRS Service Node.

Our proposed internetworking architecture takes into account the NGMN recommendations for the near future. It simplifies the complexity in the network integration and presents a flatter network architecture [3]. Moreover, the proposed internetworking architecture is IP based, which enables the provision of diversified and flexible services to fulfil different user requirements [4].

Based on the proposed internetworking architecture, an intelligent RAN selection system ought to be deployed. The RAN selection 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 mechanism for the RAN selection. As shown in Fig. 1, a Decision Maker (DM) resides in the IP backbone network. It receives and updates user/terminal and network context information, and accepts user service requests. Based on the context information, the DM implements an algorithm to generate an optimized selection and then transfers the result to the user. In order to exchange the service requests and transmit the context information, a specific signalling mechanism should be implemented in a heterogeneous communication environment. For the signalling we use the SIP protocol, because it is simple, extensible and it can be integrated with the IP technology [5]. In addition to SIP, a signalling network is required to enable the communications between the DM and the user terminals, especially when the user terminal is still not connected to any RAN. Public Land Mobile Networks (PLMNs) have a broad coverage area and reliable link [6]. Therefore a PLMN is a better candidate over other RANs to work as a signalling network for the intelligent RAN selection system. Together with a database of network locations and user location information, using a PLMN as initial signalling network even can eliminate the need for thorough RAN scanning. However, this does not mean that the PLMN is the only choice. Users can communicate with the DM via their existing non-PLMN based connections if a PLMN is not available.

Each DM serves a certain area/domain and cooperates with a limited number of RANs so as to avoid excessive signalling traffic within the backbone network, as well as managing better scalability and reliability. Compared to the related research, such as the Composite Radio [7], our proposed DM is centralized and situated in the IP backbone network. Assuming two requests are in process, after making a reply for the first one, the context information is stored in the database, the DM can dynamically get the updated information and use it for the second request. This avoids signalling overhead for querying the network status information and decreases the response time.

III. Network Modelling

The system architecture introduced above enables a context-aware RAN selection. Network resource availability is an important attribute of context information. They are dynamic and dependent on the access technology and the service type. In order to develop a context aware RAN selection algorithm and investigate the RAN selection algorithm's performance, we develop simple but effective models for evaluating resource availability in UMTS and WiFi networks. The models try to treat more realistic network scenarios. Our models analyze various service types, including speech, video call, audio streaming, video streaming, web browsing, and file transfer. When a request arrives, the models consider the characteristics of the requested service, as well as the access technology, and efficiently obtain network resource availability, which is used by the RAN selection algorithm.

A. *Mathematical Model for UMTS*

The UTRAN implements the W-CDMA technology and it is an interference-limited cellular network. The modelling can be divided into two parts: uplink and downlink. For duplex services, such as speech and video call, both uplink and downlink should be considered. For asymmetrical services, e.g. video streaming and web browsing, the downlink traffic plays a major role and the model neglects the uplink traffics and concentrates on the downlink.

In the uplink, we firstly consider two parameters: interference margin and noise rise. The interference margin is a predefined parameter in the link budget and its value determines the loading of the cell [8]. The size of the cell coverage is limited by the maximum allowed path loss defined in the link budget. A typical value for the interference margin in coverage-limited cells ranges from 1 to 3 dB [8]. In our model, the radius of an UTRAN is 1000 metres and the interference margin is defined as 3 dB. The noise rise measures the ratio between the base station total received wideband power and the base station thermal noise power. Assuming a new uplink connection is made to the network, it will increase the total wideband power received at the base station and lead to a new value of noise rise. If the noise rise is greater than the interference margin, that means the uplink resource is not enough and the requested uplink connection will be rejected. Otherwise, the requested uplink connection can be admitted. Here, we will explain how to calculate the value of the received wideband power. First, we need to define the E_b/N_o, the ratio between energy per user bit and the noise spectral density. The E_b/N_o value is service type dependent, e.g. 5.0 dB for speech [8]. The E_b/N_o value of the admitted user *j* can be derived as follows [8]:

$$(E_b/N_0)_j = \frac{W}{v_j * R_j} * \frac{i_j}{I_{total} - i_j} \quad (1)$$

W is the chip rate of W-CDMA; v_j is the activity factor of user *j*; R_j is the bit rate of user *j*; i_j is the received signal power from user *j* at the base station; I_{total} is the total wideband power at the base station, which includes the total received wideband signal power from the users who are being served and the thermal noise power in the base station.

Assuming the number of the admitted users in a UTRAN cell is *n*, the connection request from a new user, which has not been admitted, will be the *n+1*th user in the network and (1) can be modified as:

$$(E_b/N_0)_{n+1} = \frac{W}{v_{n+1} * R_{n+1}} * \frac{i_{n+1}}{I_{total\_n}} \quad (2)$$

i_{n+1} is the estimated value of the signal power from the new connection. I_{total_n} is the wideband power received at the base station, before the *n+1*th user is admitted. The reason for such modification is that, before the request is admitted, the current total received wideband power at the base station includes the wideband signal power from the users who have already been admitted, but it does not include the signal power from the new connection request. From the perspective of the new connection request, the current total interference can be regarded as noise against its signal power at the base station. Isolating i_{n+1} we get:

$$i_{n+1} = \frac{(E_b/N_0)_{n+1} * v_{n+1} * R_{n+1} * I_{total\_n}}{W} \quad (3)$$

Then, we estimate the new total received wideband power I_{total_n+1} as:

$$I_{total\_n+1} = I_{total\_n} + i_{n+1} + i_{increased\_intercell\_interference} \quad (4)$$

$i_{increased\_intercell\_interference}$ represents the increased inter-cell interference. During the development of the mathematical model, we discover that inter-cell interference brings significant influence to the capacity of the interference-limited

Fig. 2. The Layout of cells (Cell 1 is the cell of interest)

cellular network. However, the authors in [8] only assume the ratio of other cell to own cell interference as 55%. In order to develop a more realistic model, for the uplink analysis, we consider the interference from the first tier of cells surrounding the cell of interest. In the downlink, the first and second tiers of cells are considered. The layout of the cells is shown in Fig. 2. Currently, our research focuses on investigating the network resource availability in one cell (cell 1 in Fig. 2). In order to simplify the interference analysis, we assume that all the surrounding cells possess the same features as the cell of interest. They are identical in network configuration, link budget, user distribution, services power consumption, received interference and radio environment. Therefore, as a new connection request is made in the cell of interest, identical requests will correspondently occur in the surrounding cells and introduce new inter-cell interference. Similar assumptions can be found in [9]. We obtained $i_{increased\_intercell\_interference}$ from:

$$i_{new\_intercell\_interference} = \frac{i_{n+1} * \sum_{j=2}^{7} L_{n+1th\_user\_in\_cell\_j,BS\_in\_cell\_1}}{L_{n+1th\_user\_in\_cell\_1,BS\_in\_cell\_1}} \quad (5)$$

$L_{n+1th\_user\_in\_cell\_1,BS\_in\_cell\_1}$ is the radio propagation attenuation between the *n+1*th user and the base station, both of which are in the cell of interest. $L_{n+1th\_user\_in\_cell\_j,BS\_in\_cell\_1}$ is the radio propagation attenuation between the *n+1*th user in cell *j* and the base station in the cell of interest. Radio propagation attenuation can be calculated based on [10].

From the point of the existing user's view, the new admitted connection increases the interference against their signal power. In order to maintain their E_b/N_o, they have to raise their signal power. For example, before the new user is admitted, the signal power from the *n*th user is denoted as i_n. After the new user is admitted, the signal power has to be recalculated based on (1), (4), and (5) as:

$$i_n' = \frac{\frac{(E_b/N_0)_n * v_n * R_n}{W} * \left(I_{total\_n+1} - i_n - \sum_{j=2}^{7} \frac{i_n * L_{nth\_user\_in\_cell\_j,BS\_in\_cell\_1}}{L_{nth\_user\_in\_cell\_1,BS\_in\_cell\_1}} \right)}{1 - \frac{(E_b/N_0)_n * v_n * R_n}{W} * \sum_{j=2}^{7} \frac{L_{nth\_user\_in\_cell\_j,BS\_in\_cell\_1}}{L_{nth\_user\_in\_cell\_1,BS\_in\_cell\_1}}} \quad (6)$$

The new value of the signal power generated by the *n*th user also increases the total received interference at the base station according to (4). In order to maintain the required E_b/N_o, the other users have to recalculate and raise their signal power, which further increase the total received wideband power at the base station. Then, considering the new value of the wideband power, the connection request has to recalculate its signal power and the total wideband power based on equations (6) and (4). These recalculations will be repetitively carried out until the value of the total wideband power stabilizes and achieves an equilibrium state or the noise rise value is greater than the interference margin.

In the downlink, the network resource availability is determined by the base station transmission power. Our model assumes that the maximum value of the base station transmission power is 20 Watts [8]. 20 percent of this power is used for signalling and the remaining (16 Watts) is used for traffic [8]. Considering a new downlink connection request, if the base station transmission power is still less than the maximum transmission power, the connection can be admitted. If not, it will be rejected. Similar to the uplink, assuming the number of the admitted users in a UTRAN cell is *n*, we derive the value of E_b/N_0 of user *n+1*th user in the cell of interest as:

$$(E_b/N_0)_{n+1} = \frac{W}{v_{n+1} * R_{n+1}} * \frac{P_{BS\_in\_cell\_1,n+1th\_user\_in\_cell\_1} * L_{BS\_in\_cell\_1,n+1th\_user\_in\_cell\_1}}{\left[\begin{array}{l} P_{BS\_in\_cell\_1} * (1-\alpha) * L_{BS\_in\_cell\_1,n+1th\_user\_in\_cell\_1} + \\ \sum_{k=2}^{19} P_{BS\_in\_cell\_k} * L_{BS\_in\_cell\_k,n+1th\_user\_in\_cell\_1} + P_N \end{array}\right]} \quad (7)$$

where $P_{BS\_in\_cell\_1,n+1th\_user\_in\_cell\_1}$ is the transmission power of base station in the cell of interest for the *n+1*th user; $L_{BS\_in\_cell\_1,n+1th\_user\_in\_cell\_1}$ is the downlink attenuation between base station and the *n+1*th user, both of which are in the cell of interest; $P_{BS\_in\_cell\_1}$ is the total transmission power of base station in the cell of interest; α is average orthogonality factor in the cell whose value ranges from 0.4 to 0.9 (1 means totally orthogonal); $P_{BS\_in\_cell\_k}$ is the total transmission power of base station in cell *k*, which is one of the cells (first and second tiers) surrounding the cell of interest; $L_{BS\_in\_cell\_k,n+1th\_user\_in\_cell\_1}$ is the attenuation between base station in cell *k* and the *n+1*th user in the cell of interest, and P_N is the average thermal noise at the user terminal.

As discussed before, all the surrounding cells possess the same features as the cell of interest. So, the transmission power of every base station should be the same and $P_{BS\_in\_cell\_k}$ can be replaced by $P_{BS\_in\_cell\_1}$. After the modification, we can calculate $P_{BS\_in\_cell\_1,n+1th\_user\_in\_cell\_1}$:

$$P_{BS\_in\_cell\_1,n+1th\_user\_in\_cell\_1} = \frac{(E_b/N_0)_{n+1} * v_{n+1} * R_{n+1}}{W} * \left[P_{BS\_in\_cell\_1} * (1-\alpha) + \frac{\sum_{k=2}^{19} P_{BS\_in\_cell\_1} * L_{BS\_in\_cell\_k,n+1th\_user\_in\_cell\_1} + P_N}{L_{BS\_in\_cell\_1,n+1th\_user\_in\_cell\_1}}\right] \quad (8)$$

By summing up the transmission power of base station in the cell of interest for every individual user, such as $P_{BS\_in\_cell\_1,n+1th\_user\_in\_cell\_1}$, the total transmission power of base station in cell of interest can be derived from (8) as:

$$P_{BS\_in\_cell\_1} = \sum_{m=1}^{n+1} \left\{ \frac{(E_b/N_0)_m * v_m * R_m}{W} * \left[P_{BS\_in\_cell\_1} * (1-\alpha) + \frac{\sum_{k=2}^{19} (P_{BS\_in\_cell\_1} * L_{BS\_in\_cell\_k,mth\_user\_in\_cell\_1}) + P_N}{L_{BS\_in\_cell\_1,mth\_user\_in\_cell\_1}} \right] \right\} \quad (9)$$

At the end, we can calculate $P_{BS\_in\_cell\_1}$ as:

$$P_{BS\_in\_cell\_1} = \frac{\sum_{m=1}^{n+1} \left[\frac{(E_b/N_0)_m * v_m * R_m}{W} * \frac{P_N}{L_{BS\_in\_cell\_1,mth\_user\_in\_cell\_1}} \right]}{1 - \sum_{m=1}^{n+1} \left[\frac{(E_b/N_0)_m * v_m * R_m}{W} * \left[(1-\alpha) + \sum_{k=2}^{19} \frac{L_{BS\_in\_cell\_k,mth\_user\_in\_cell\_1}}{L_{BS\_in\_cell\_1,mth\_user\_in\_cell\_1}} \right] \right]} \quad (10)$$

If $P_{BS\_in\_cell\_1}$ is smaller than or equal to the maximum base station transmission power, the downlink connection request can be accepted. Otherwise, the request will be rejected.

B. *Model for WiFi*

The WiFi network implements the IEEE 802.11b standard. The performance of IEEE 802.11 network has been investigated in [11] and [12]. The models presented in these two papers have been used as the basic analytical methods for investigating IEEE 802.11 network performance and cited in several latter publications in this field.

Bruno et al. [13] investigated the performance of the TCP connections over the WiFi network. 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. Bruno et al. also assumed that each station in the WiFi network possess a single *'long-live'* [13] TCP session which has an unlimited amount of data in the source and at least one packet to transmit. However, the proposed analysis method cannot be directly applied to our model when considering the characteristics of different applications. For example, a typical web browsing service session is not a *'long-live'* TCP session and it can be divided into ON/OFF periods [14]. The ON period represents web page download comprising packet calls. The OFF period represents the intermediate reading time. Furthermore, real-time services, like VoIP, are becoming more popular and that is urging evaluations of IEEE 802.11 network performance for hybrid service types.

We realize that the assumptions of [13] are inconsistent with the reality. Therefore, we have developed a simple but effective solution for evaluating capacity and resource availability in WiFi networks. We propose a new parameter, the expected number of contending packets over the wireless channel, which is denoted by e_{ncp}. 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 larger 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 (UDP based and TCP based services coexist), the connection will be rejected. This is because packet collisions will cause delays and packet loss for the real-time UDP based service sessions and there is no more guarantee that the delay and packet loss will be acceptable according to the requirements of the services. If the requested and existing service types are all TCP based, the analysis method proposed in [13] can be implemented to calculate the effective WiFi network transmission rate (excluding traffic and protocol overheads) of each packet generated by the requested connection and the existing users. Based on the effective packet transmission rate, the effective end-to-end bandwidth for the TCP based request and the existing service sessions can be estimated. If the effective bandwidth satisfies the requirements of all the users, the connection request will be admitted. If not, the request will be rejected.

Before presenting how to investigate the effective WiFi network packet transmission rate for the requested connection and the existing users, we will first explain the method to calculate e_{ncp}:

$$e_{ncp} = \sum_{i=0}^{N} i * \binom{N}{i} * e_p{}^i * (1-e_p)^{(N-i)} = N * e_p \qquad (11)$$

N is the number of existing connections in the network plus the new requested connection(s). For a speech service user, he/she introduces two connections, because the speech service is duplex and bi-directional. For an audio streaming service or web browsing user, he/she is assumed to introduce one connection. This is because the above services are asymmetrical and the downlink traffic plays a major role. Therefore, for simplicity reasons, the analysis neglects the uplink traffics and concentrates on the downlink. e_p is the probability that the channel is occupied by a packet transmission. We calculate e_p by the equation below:

$$e_p = \sum_{s\_t} \left(p\_on_{s\_t} * n_{s\_t} / N\right)$$

(12)

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*. The calculation of $p\_on_{s\_t}$ is service type dependent and we will present it later. n_{s_t} is the number of connections of service type *s_t*.

Assuming e_{ncp} is 2, that means, on average, there are two packets in contention for access the network channel at any time. In a situation where all the service types are TCP based, we suppose these packets are generated by ***two*** *TCP sessions* coexisting in the network. These *TCP sessions* always have packets in their queues and they are ready to send packets at any time. Compared to the *'long-live'* TCP session assumed in [13], these two *TCP sessions* behave in the same way. Therefore, the analytical method presented in [13] can be implemented. These two *TCP sessions* can have a fair access to the WiFi network channel whose effective packet transmission rate is about 4400 kbps. This value is derived from the analytical model in [13]. It is also validated through realistic discrete-event simulations in [13]. Therefore, each *TCP session* can obtain an effective packet transmission rate of about 2200 kbps. That means the payload of each packet generated by these two *TCP sessions* can be transmitted at the throughout 2200 kbps. The calculation of e_{ncp} considers all the service types in the network. As mentioned before, not all service sessions behave as the *'long-live'* TCP session, such as web browsing. Therefore, the *TCP session*, which is considered above, may comprise several different service sessions. As a consequence, the packets generated by the *TCP session* may belong to different service sessions and, on average, the effective transmission rate of each packet generated by each service session is 2200 kbps. The calculation of the effective end-to-end bandwidth for the TCP based service session should consider the characteristics of the services (such as packet arriving interval), as well as the data rate and transmission delay in the core network.

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. Because 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. For services like VoIP, generally speaking, their encoders do not adjust the transmitting rates to the available channel bandwidth, except some technologies such as AMR. Therefore, the effective bandwidth values for UDP based services are identical with their transmitting rates. For TCP based services, each packet can be transmitted at the maximum effective packet transmission rate of about 4400 kbps (according to the proposed analytical method in [13]). This is because, on average, there is less than one packet in contention to access the network channel.

$p\_on_{s\_t}$ can be derived as:

$$p\_on_{s\_t} = t\_p_{s\_t} / in\_p_{s\_t} \qquad (13)$$

$in\_p_{s\_t}$ represents the packet inter arrival time of service type *s_t*. $t\_p_{s\_t}$ is the mean time spent in completing the transmission for a packet of service type *s_t*. It consists of not only the time used for transmitting the packet payload and the headers but also the overheads introduced by the CSMA/CA mechanism [15].

For example, a VoIP codec generates 50 packets per second at the data rate of 8 kbps, the packet inter arrival time $in\_p_{s\_t}$ is 20 ms and the packet size is 20 bytes. As the IP/UDP/RTP header size is 40 bytes, $t\_p_{s\_t}$ can be calculated as follows:

$$\begin{aligned} t\_p_{s\_t} &= DIFS + e\_idle + phy\_mac\_hdr + \\ &= t\_b * (20+40) + delay + SIFS + delay + t\_ack \end{aligned} \qquad (14)$$

DIFS represents the DCF interframe space and *SIFS* represents the short interframe space. *e_idle* is the average backoff time [12]. *phy_mac_hdr* is the time spent in transmitting the physical and MAC layer headers. *t_b* means the time used for transmitting a byte in the WiFi network. *delay* represents the maximum radio propagation between the terminal and the AP, which is 1μs [12]. *t_ack* is the time used to transmit an acknowledgement (ACK) packet. The calculation indicates that the WiFi network can support 11 VoIP users. This result complies with the results obtained by [16].

For file transfer service, $p\_on_{s\_t}$ can be derived as:

$$p\_on_{s\_t} = t\_p_{s\_t} / \tau_c \quad (15)$$

τ_c represents the time spent in transmitting a TCP ACK packet from the client to the server and the time spent in transmitting a TCP data segment from the server to the AP. Its value is random and exponentially distributed with a mean of 50 ms [14]. We take the mean value in our calculation. $t\_p_{s\_t}$ consists of the time used for transmitting the data segment payload, the TCP ACK payload, the headers and the overheads introduced by the CSMA/CA mechanism. $t\_p_{s\_t}$ can be calculated as:

$$\begin{aligned} t\_p_{s\_t} = & e\_idle + 2*delay + SIFS + t\_ack + DIFS + \\ & phy\_mac\_hdr + t\_b*data\_seg + \\ & e\_idle + 2*delay + SIFS + t\_ack + DIFS + \\ & phy\_mac\_hdr + t\_b*tcp\_ack \end{aligned} \quad (16)$$

data_seg is the payload of a data segment which is 1500 bytes [14]. *tcp_ack* represents a TCP ACK packet and is 40 bytes [14]. The definition of (16) is based on the assumption that the TCP advertised window size is equal to one. One data segment will not be transmitted until the TCP ACK has been received.

This model also can be used for the IEEE 802.11a network by replacing the values of *DIFS*, *SIFS*, *e_idle*, *phy_mac_hdr*, *t_b*, delay, and *t_ack*.

TABLE I
SERVICE PARAMETERS AND TYPICAL VALUES

SERVICE TYPE	VC	VS	AS	WB	SA	FT
NOMINAL BANDWIDTH (kbps)	64/128	128/256	32/64	32/64	32/64	64/128
TRANSPORT PROTOCOL	UDP	UDP	UDP	TCP	TCP	TCP

Simulations carried out in OPNET have confirmed the usability our model by generating similar results. Because of limited space, the simulation results are not presented here, but they can be found in [17].

IV. RADIO ACCESS NETWORK SELECTION ALGORITHM

We consider six different types of services: Video Call (VC), Video Streaming (VS), Audio Streaming (AS), Web Browsing (WB), File Transfer (FT) and HTTPS/TLS based Secure Applications (SA). 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 bandwidth, which provides the minimum quality constraint and threshold that the service should meet. The premium service class has a higher bandwidth, which provides better service quality when resources are available. The service parameters and typical values are listed in Table 1 [18]. For the UDP based real-time services, the bandwidth values are fixed. For the TCP based non-real-time services, the bandwidth values are the minimal requirements for each service class.

In this paper, we will present and compare two basic RAN selection algorithms that we called A1 and A2 for simplicity.

When a service request arrives, A1 will consider the type of

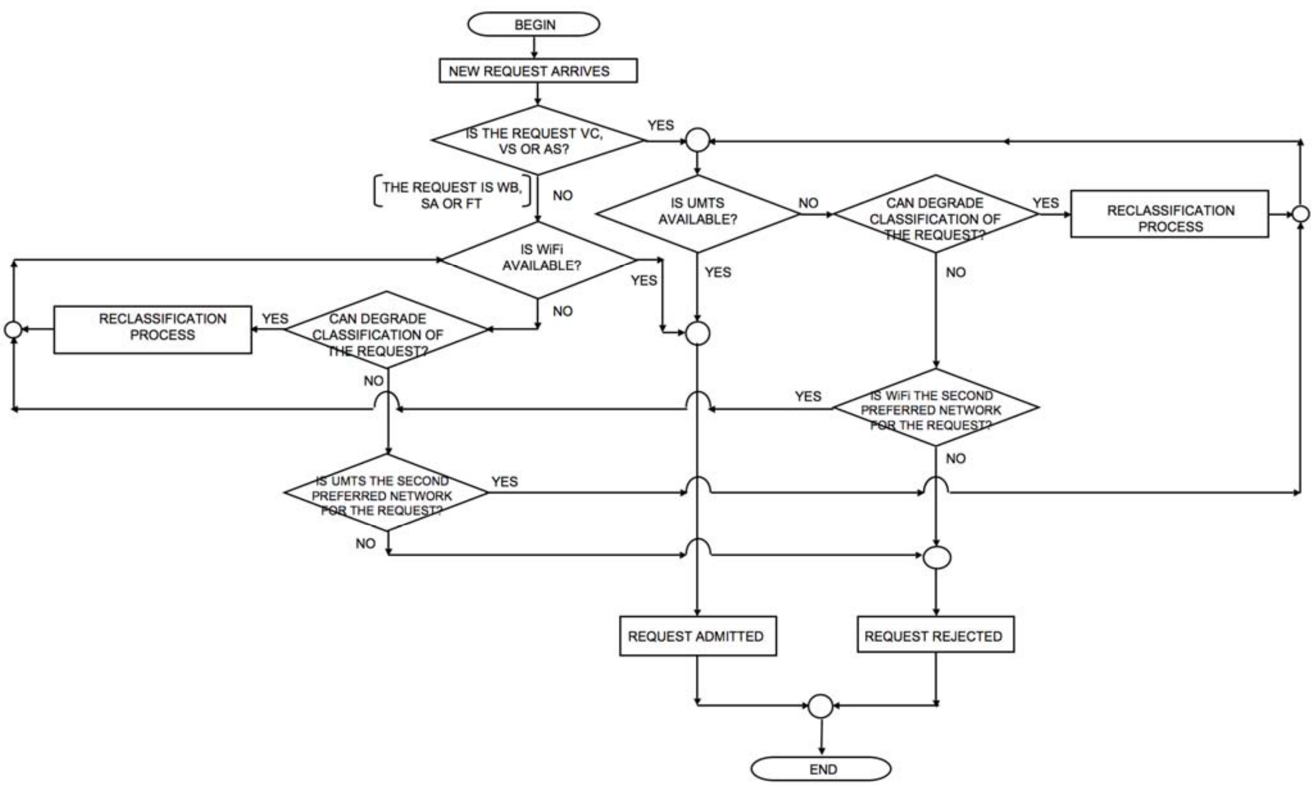

Fig. 3. Flowchart of Algorithm A1

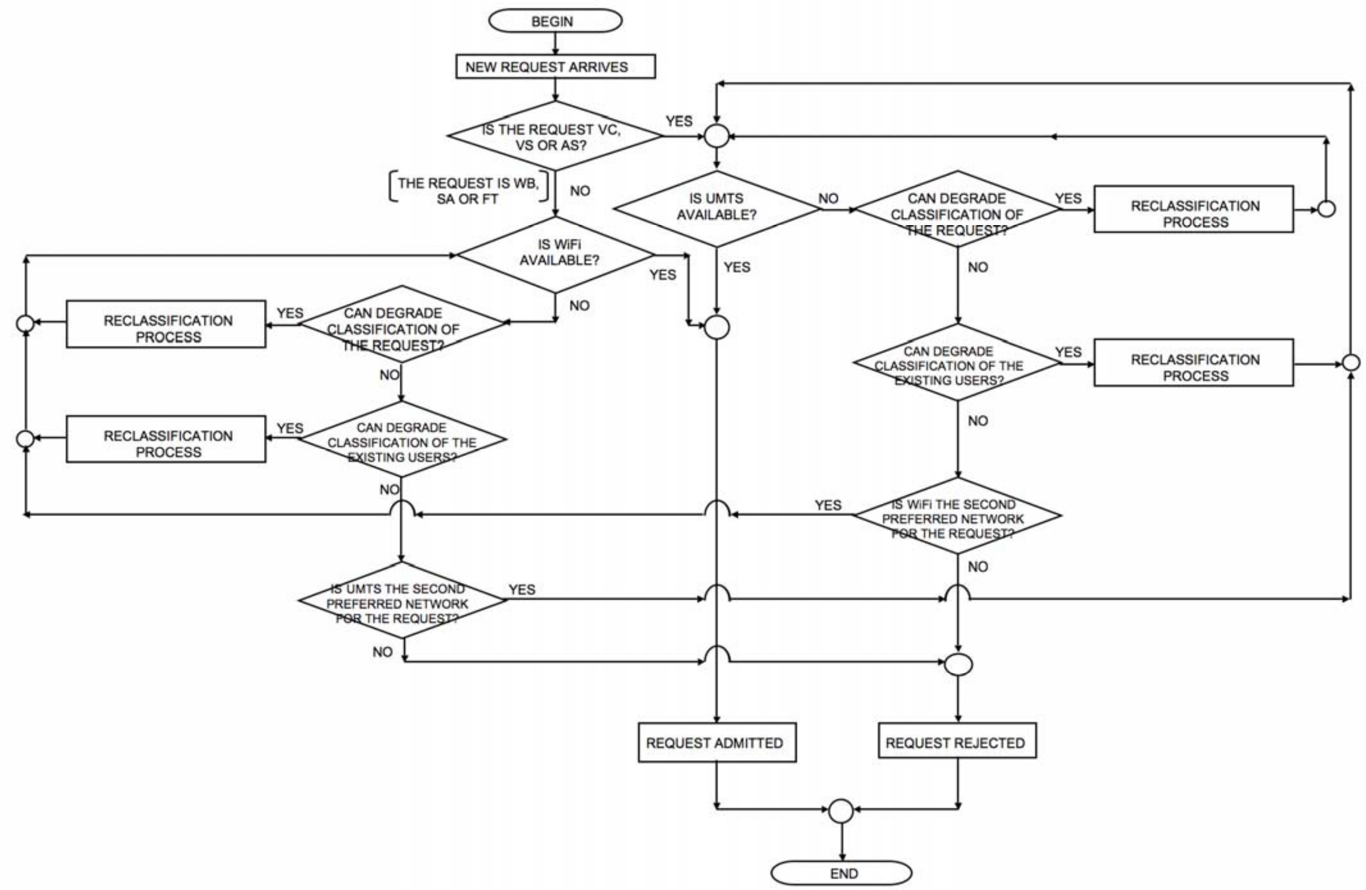

Fig. 4. Flowchart of Algorithm A2

the requested service and determine the preferred network. UMTS is the preferred network for real-time services. WiFi is the preferred network for non-real-time services. Then, A1 will firstly investigate whether the network resource are sufficient to admit the service request with its premium class, based on the network models presented in section 3. If insufficient, A1 will degrade the service class of the request to basic and investigate the network resources again. If the resources still are not sufficient, a second preferred network would be selected and investigated. Fig. 3 depicts the flowchart of A1.

The A2 algorithm considers the existing service types and classes admitted by all the networks. When there are insufficient network resources to admit a new request, A2 will try to degrade the service classes of the existing users in order to optimize the network resource usage and reduce the call blocking rate. The degradation is priority-based. Starting from the left, the services are classified from low to high priorities as FT, WB, SA, AS, VS, and VC. The flowchart of A2 is shown in Fig. 4.

V. Radio Access Network Simulations and Results

In order to compare the network performance obtained from the different RAN selection algorithms, we have implemented call level simulations. The simulations implement six types of services presented above. The proportion of users of each service in relation to the total number of users is: 25% of WB users, 5% of SA users, 20% of FT users, 15% of VC users, 30% of VS users and 5% of AS users [18]. No background traffic is assumed [18]. The simulations consider a group of 30 users. In the first 10 seconds, each user starts one service. In the next 10 seconds, 14 users request one more service each. Finally, from 30 to 60 seconds, all users stop their services gradually. The service starting time and termination time are uniformly distributed. This configuration is based on [18].

Fig. 5, 6, and 7 depict the joint (UMTS and WiFi), UMTS, and WiFi network utilities obtained from the different RAN selection algorithms. Although the joint network utility based on A2 is greater than A1, the UMTS network utility obtained from A2 is smaller than from A1, while the WiFi network

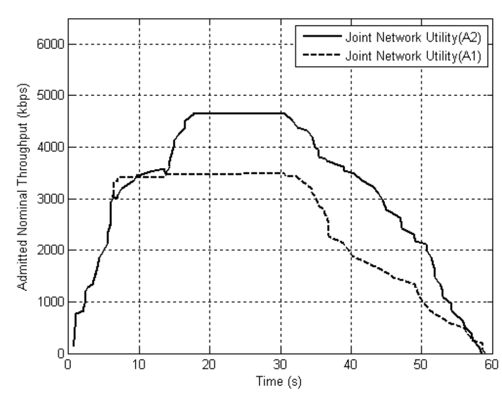

Fig. 5. Joint Network Utilities of Different Algorithms

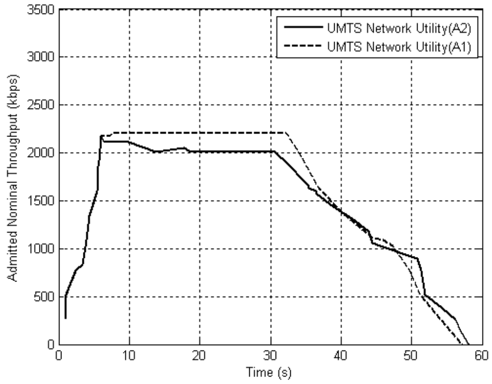

Fig. 6. UMTS Network Utilities of Different Algorithms

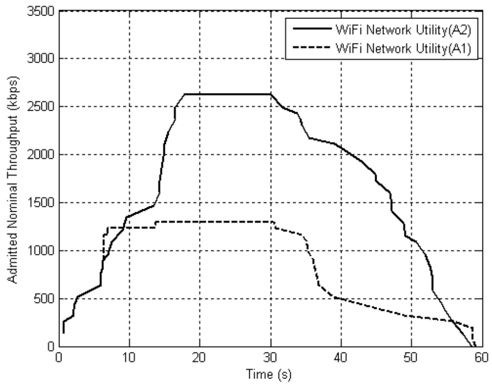

Fig. 7. WiFi Network Utilities of Different Algorithms

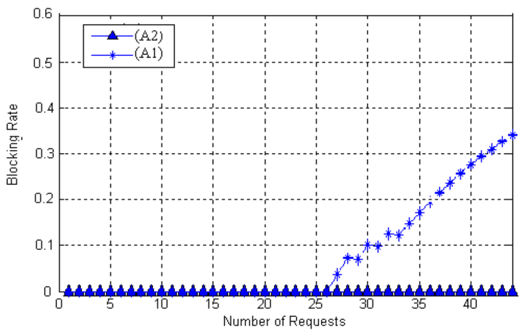

Fig. 8. Ratio of blocked requests to total number of requests of Different Algorithms

utility expresses the opposite effect. This is because, different from A2, A1 will only attempt to degrade the service class of the request. If this attempt fails, it will select the second preferred network for the request and perform the resource availability investigation on that network. For example, if A1 determines that UMTS is incapable to admit the new real-time service request, the WiFi network will be attempted. Because the service classes of the existing users in UMTS are not adjusted when using A1, the network utility may be greater than the value when using A2. However, once the WiFi network admits the UDP based real-time service request, its capacity will be dramatically impaired. This is because, if the existing service types in a WiFi network are hybrid, e_{ncp} must remain not greater than 1 so as to ensure the quality requirements of the existing UDP based services. Nevertheless, WiFi primarily provides users with TCP based data services. For TCP based data services, e_{ncp} can be greater than 1. But the existence of UDP based services limits the value of e_{ncp} and prevents the WiFi network from admitting further TCP based data service requests, which makes the WiFi utility smaller when using the algorithm A1.

Fig. 8 depicts the ratio of blocked requests to total number of requests suffered by the networks when using the different RAN selection algorithms. Before the 27th service request comes forth, the ratios of both algorithms are zero. Then, when A1 is used, the ratio begins to increase and reaches 35% in the end of simulation time. In contrast, when A2 is used, the ratio remains zero. A2 outperforms A1 because A2 can dynamically adjust the service classes of the existing users and allow more requests to be admitted.

VI. Conclusion

In this paper, we firstly proposed an internetworking architecture in a heterogeneous communication environment. A network based and context aware RAN selection system, the DM, currently resides in the backbone network. However, this is an on-going research. The position and specific functionalities of the DM is still under study. We intend to extend the system architecture and the DM to cope with multi-provider scenarios, where the DM may act as a broker.

Network resource availability is an important attribute of context information. This paper presented the mathematical models for evaluating network resource availability of two RANs: UMTS and WiFi. The models allow the investigation of more realistic scenarios of network and service types. Our work provides a foundation for future research in network modelling and RAN selection algorithm development.

References

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

[2] 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

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

[4] 3GPP. TR 22.978 v7.1.0 (2005-06) Technical Report, “All-IP Network (AIPN) feasibility study (release 7)”.

[5] W. Li. (2003). “A Service Oriented SIP Infrastructure for Adaptive and Context-Aware Wireless Services.” in the 2nd international conference on Mobile and ubiquitous multimedia Norkkoping, Sweden [Online]. Available: http://www.ep.liu.se/ecp/011/014/ecp011014.pdf

[6] M. Inoue, et al., “MIRAI: A Solution to Seamless Access in Heterogeneous Wireless Network.” Communications, 2003 IEEE International Conference, Volume 2, pp. 1033 - 1037

[7] P. Demestichas, et al., “Evolution in Wireless Systems Management Concept: From Composite Radio Environment to Reconfigurability.” in the Communications Magazine, IEEE, Volume 42, Issue 5, 90- 98, 2004

[8] H. Holma, et al., *WCDMA for UMTS – HSPA Evolution and LTE, Fourth Edition*. John Wiley & Sons, LTD, 2007, Chapter 8.

[9] Y. Chen, “Soft Handover Issues in Radio Resource Management for 3G WCDMA Networks”, Ph.D. dissertation, Department of Electronic Engineering, Queen Mary, University of London. 2003.

[10] ETSI. TR 101 112 v.3.2.0 Technical Report. “Universal Mobile Telecommunication System (UMTS); Selection procedures for the choice of radio transmission technologies of UMTS (UMTS30.03 version 3.2.0,)”

[11] G. Bianchi, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function." Selected Areas in Communications, IEEE Journal on, Volume 18, Issue 3, pp. 535-547, 2000.

[12] F. Cali, M. Conti and E. Gregori, "Dynamic Tuning of the IEEE 802.11 Protocol to Achieve a Theoretical Throughput Limit". Networking, IEEE/ACM Transactions on, Volume 8, Issue 6, pp. 785 – 799, 2000

[13] 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.

[14] Tony Dean, et al., "HRPD-A Downlink Voice/IP and Data Throughput and Delay Analysis". Proceedings of the 2006 International Symposium on a World of Wireless, Mobile and Multimedia Networks.

[15] IEEE 802.11 Wireless Local Area Networks. 2007. "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications"

[16] W. Wang, S. Liew and O. Li, "Solutions to Performance Problems in VoIP Over a 802.11 Wireless LAN." Vehicular Technology, IEEE Transaction on, Vol. 54, Issue 1, pp. 366 – 384, 2005.

[17] W. Luo and E. Bodanese, "Intelligent Radio Access Network Selection System for Heterogeneous Wireless Network Environments." In the Proceeding of Networking and Electronic Commerce Research Conference. 2007.

[18] G. Koundourakis, D. Axiotis and M. Theologou. "Network-based Access Selection in Composite Radio Environment." in the proceeding of 2007 IEEE Wireless Communications & Networking Conference, WCNC 2007, pp. 3877-3883.