Adaptive and Efficient Radio Access Selection and Optimisation 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 wireless network technologies have been developed and deployed. Future mobile communication system will have a packet switched core. The access to the mobile communication system will not be restricted to the mobile cellular networks but may be via other wireless or even wired technologies. Such hybrid access can enable service convergence, joint resource management and adaptive quality of service. The future mobile communication system still has many pending issues to solve. One of them is the selection of the most appropriate radio access network when receiving a service request. This paper addresses this issue by proposing a new adaptive and efficient algorithm. This algorithm facilitates radio access network selection and optimisation considering application requirements, user satisfaction, gains, as well as network resource availability and utilisation. The simulation results show that the proposed algorithm can improve the network performance and capacity.**

I. INTRODUCTION

Nowadays, multiple wireless network technologies are available as commercial wireless systems. They 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 (e.g. WiMAX). The overlapping of different Radio Access Networks (RANs) creates heterogeneous communication environments. Future moblie communication systems consider the heterogeneous communication environments and foresee universal access to realise service convergence, joint resource management and adaptive quality of service [1]. In an environment with multiple technlogies, it is a challenge to make the RANs cooperate with each other to achieve the above aims. Considering that today's wireless and fixed networks are increaingly packet-based, one solution is to introduce a packet switched (PS), IP-based core network, for example, the 3GPP's Long Term Evolution [1]. The advantages of having a PS core network can be: enabling intelligence at the network edge and supporting various business models [2]; facilitating the provision of diversified and flexible services which can fulfil different user requirements [3]; providing a flatter network architecture and simplifing network integration [4].

An internetworking architecture is presented in Fig. 1. It includes a PS core network, which is the backbone integrating different RANs, such as the evolved UTRAN, 2G/3G RAN, WiFi, and WiMAX. The infrastructure of the RANs can be maintained without modifications. The PS core can directly connect with the evolved UTRAN, the WiFi and the WiMAX networks. In some situations, such as for security reasons [1], a gateway may be introduced between the PS core and the

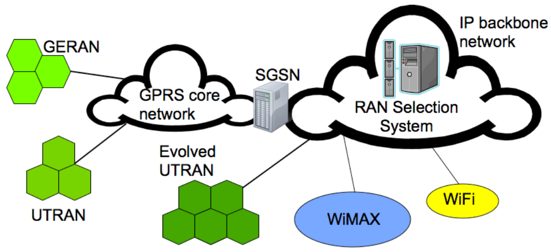

Fig. 1. Internetworking Architecture

WiFi/WiMAX networks. For RANs like the legacy circuit-switched mobile networks, the connection to the PS core can be made via an access router or a gateway, as for instance, a Serving GPRS Service Node.

In such a heterogeneous environment, an intelligent RAN selection system should be implemented to make the heterogeneous communication system function in an adaptive and efficient way. Zhu and McNair [5] investigate using a cost function to solve the handover issues in a heterogeneous environment. The cost function takes the attributes of a network and returns a value, which represents the cost of actual handover. The network that results in the lowest value is selected. Koundourakis et al. present a network-based access and interface selection system in [6]. Their research concentrates on the optimal usage of network resources and provision of acceptable QoS. In [7], we propose a network-centric RAN selection system, which resides in the PS core as shown in Fig. 1. The RAN selection system collects and updates user/terminal and network context information, and accepts user service requests. Based on the context information and a selection algorithm, it selects an appropriate network for the user.

This paper describes our new access selection and optimisation algorithm used to generate an optimal solution for our RAN selection system. The proposed algorithm not only selects an appropriate network, but also applies adjustments to the network resources utilisation and management if necessary. Section 2 describes the RAN selection and optimisation algorithm. Section 3 presents the performance analysis of the proposed algorithm, and section 4 concludes the paper.

II. RADIO ACCESS SELECTION AND OPTIMISATION ALGORITHM

Our RAN selection and optimisation 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 introduces complexity to the selection process. Furthermore, due to user handovers, network resource availability may vary frequently, even during the lifetime of an active service session. Such variability may affect QoS guarantees and user admission where adaptations to the service sessions and network resource utilisation would be required. It makes the selection process even more dynamic and complex. Considering the above challenges, we propose a *Policy-based Radio Access Selection and Optimisation* (P-RASO) algorithm.

The P-RASO algorithm is adaptive and works with multiple policies. It can cope with different scenarios, network conditions and aims of operators by implementing different policies in a dynamic way. The policies measure the gains obtained from every candidate network covering a user request and select the network which can provide the greatest gains. The policies evaluate the obtained gains in different ways. Some may consider the gains as having a larger group of users. Some may regard the gains as having a higher level of overall user satisfaction. Some may view the gains as obtaining higher revenue. The gains are represented by the values in an objective function (OF). The aim of the policies is to maximise the objective function. The adaptivity of the P-RASO algorithm lies not only in its ability of implementing different policies according to different scenarios, network conditions and aims of operators, but also in the way the policies maximise the objective function when facing varying network resource availability.

As network resource availability changes, inside each policy, a data rate adaptation will be performed in order to optimise the network resource utilisation while considering two factors, the changed carried traffic and the OF value. For example, when the resources in network A are not sufficient to accept a user request, the data rate adaptation considers the new user's data rate requirement and takes into account the data rates allocated to the existing users. Then, it extracts a subset of the existing users who are capable of suffering data rate degradation without violating their QoS guarantees. The data rate adaptation decreases the data rates of certain users in the subset to obtain sufficient network resources for admitting the new user. The above degradation may result in a lower OF value. The guideline of this adaptation is to obtain as many network resources as possible from each degraded user, meanwhile maintaining the OF value as great as possible. On the other hand, when an existing user leaves network A and releases the resources formerly being consumed, the data rate adaptation only considers the data rates allocated to the existing users. It extracts a subset of the existing users who are capable of having data rate upgradation. The data rate adaptation increases the data rates of certain users in the subset to consume the extra network resources. The above upgradation may result in a greater OF value. The guideline of this adaptation is to let each upgraded user consume as little extra network resources as possible, meanwhile maximising the OF value as much as possible. The aim of the data rate adaptation inside each policy is to effectively consume the network resources and to maximise the gains. The data rate adaptation will be comprehensively explained in subsection D.

In order to better understand the use of multiple policies, let us take the following scenario as an example. We assume that the P-RASO algorithm is supplied with two policies: *Overall User Satisfaction Improvement* (OUSI) policy and *User Number Increase* (UNI) policy. The OUSI policy considers the gains as having a higher level of overall user satisfaction. The UNI policy regards the gains as having a larger group of users. Suppose that a new user arrives in the heterogeneous communication system. If the OUSI policy is used, firstly it finds out the network which can provide the greatest OF value. Then, the OF value, which represents overall user satisfaction, is compared with the previous OF value generated before the new user is admitted. If the new value is greater, the network found by the OUSI policy will be selected to admit the user request. Otherwise, the user request will be rejected. The OUSI policy considers the overall user satisfaction and only admits user requests which can improve the overall user satisfaction, despite of possibly having higher blocking probability. In contrary, if the UNI policy is used, it will directly select the network providing *currently* the greatest OF value. For the UNI policy, when users' QoS guarantees are not violated, it will try to accept as many user

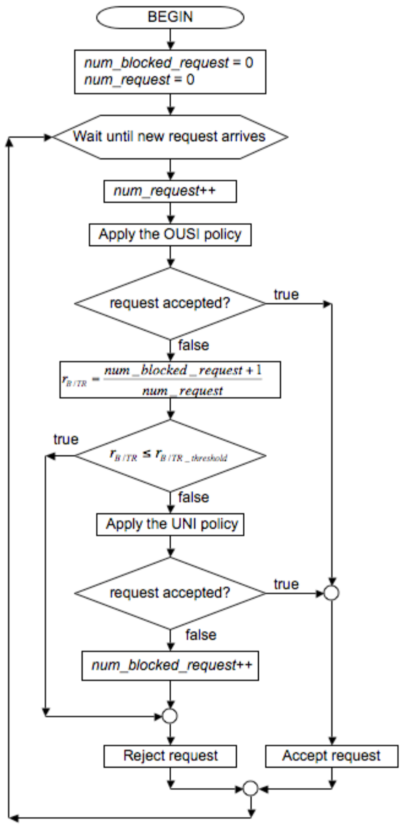

Fig. 2. An Example of the P-RASO Algorithm

requests as possible despite of possibly having lower overall user satisfaction.

The implementation and dynamic change of the OUSI and the UNI policies in P-RASO is presented in Fig. 2. When a user request arrives, the OUSI policy is firstly applied for evaluation. If the OUSI policy determines that the user request cannot be admitted, the ratio of blocked requests to the total number of requests, $r_{B/TR}$, will be calculated and compared with a predefined threshold value $r_{B/TR\_threshold}$, which can be the greatest blocking probability that the operator can accept. If $r_{B/TR}$ is smaller than or equal to $r_{B/TR\_threshold}$, the decision made by the OUSI policy will be carried out and the request will be rejected. This is because $r_{B/TR}$ is not greater than the threshold value and, in order to maintain a high overall user satisfaction, such rejection is acceptable by the operator. However, if $r_{B/TR}$ is greater than $r_{B/TR\_threshold}$, in order to keep $r_{B/TR}$ smaller than or equal to the threshold value and admit more users, the P-RASO algorithm will use the UNI policy. Then, the user request will be evaluated by the UNI policy. Finally the radio access selection and optimisation decision made by the UNI policy will be carried out.

The next subsection explains the objective function being used in the P-RASO algorithm.

A. P-RASO objective function

Assuming network A is being evaluated for a service request made by a new user, the P-RASO objective function is calculated as:

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

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 represents the *i* th existing service session in network A, X_j represents the *j* th 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 *k* th existing service session in network X_j. The objective function includes $G(RS, X_A)$, which is the measure of the gains that can be obtained when selecting network A for the new requested service *RS*, $\sum_{i=1}^{N_{X_A}} G(S_i)$ is the impact upon the gains of serving the existing users in the network A if *RS* is accepted, and $\sum_{X_j, j\neq A} \sum_{k=1}^{N_{X_j}} G(S_k)$ is the collective gains obtained from serving the rest of the existing users in the networks which are not being evaluated.

The function *G* is calculated as:

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

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. In heterogeneous environments, the RAN selection and optimisation are context aware and consider various types of context information such as data rate, service cost, etc. The types of context information may be different between policies considering their configurations or aims. This can be shown in $NORM\left(Attr_{S,i}^X\right)$, which is a normalised value presenting part of the gains obtained from network *X* if it accepts service *S*. $Attr_{S,i}^X$ represents the attribute or characteristic of the network *X*, such as available data rate, service price, etc., corresponding to a specific type of context information *i* which is being considered in the provision of service *S*. The value of $NORM\left(Attr_{S,i}^X\right)$ is supplied by a normalisation function and ranges from 0 to 1. The greater value, the more gains can be obtained. The normalisation function will be better explained in section 3. $W_{S,i}$ 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 $W_{S,i}$ ranges from 0 to 1. Before presenting the data rate adaptation scheme, we will present the methods for evaluating network resource availability.

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_o (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_o as [8]:

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

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 is calculated as:

$$P_n = \frac{(E_b/N_o)_n \times \upsilon_n \times R_n \times [(1-\alpha) \times L_{BS.n} \times P_{total} + I_n]}{W \times L_{BS,n}} \qquad (4)$$

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:

$$P_{total} = \sum_{n=1}^{N} \frac{(E_b/N_o)_n \times \upsilon_n \times R_n \times [(1-\alpha) \times L_{BS.n} \times P_{total} + I_n]}{W \times L_{BS,n}} \qquad (5)$$

Isolating P_{total} we have:

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

If P_{total} is greater than zero and smaller than or equal to the maximum base station transmission power P_{max}, the UTRAN downlink network has sufficient resources to accept the new user request. Otherwise, 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 propose 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]. Here, we summarise the main information about the model.

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. If the value of e_{ncp} 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 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} will be viewed as the number of '*long-live*' *TCP sessions* explained in [9], and the analysis method proposed by Bruno et al. in [9] is implemented to calculate the effective transmission rate (excluding traffic and protocol overheads) of each packet generated by the requested connection and the existing users. The TCP based requested service and the existing services will be delivered inside the constraints of the calculated effective transmission rate. If their minimum QoS requirements are still complied with, the new request can be accepted. Otherwise, the request will be rejected.

D. Data Rate Adaptation Scheme

A data rate adaptation scheme is performed when the resources in a network are insufficient to accept a user request, or an existing user leaves the network. The data rate adaptation scheme is presented as follows.

In the case of having insufficient resources to accept the new request, 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 service, and the data rate vector V_X is updated with the decreased data rate. This process stops when one of the following conditions is reached:

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

In the case of having extra resources (caused by the departure of users), the data rates of the existing users served by the same network are grouped together into a vector V_Y. In each iteration, the scheme selectively increases the data rate of one service, and the data rate vector V_Y is updated with the increased data rate. This process stops when no data rate that remains in the vector V_Y can be increased within the network constraints.

In both cases, the data rate adaptation scheme aims to effectively consume the network resources and maximise the gains obtained by each policy. In [11], Chatterjee et al. demonstrate the rate allocation problem in CDMA network is NP-Complete. The proposed data rate adaptation scheme is a

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

Negative overload phase:

Take V_X and calculate D_{total} according to (6)
Extract the subset $DS_X = [R_1, R_2, \dots, R_i, \dots, R_m, \dots]$
While the UTRAN is in the negative 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 $V_X' = [R_1, R_2, \dots, R_i', \dots, R_{N_X+1}]$, 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 data rate (R_{max}) supplies the greatest *RT* value: $R_{max} = argmaxRT(R_i)$
  Decrease the data rate R_{max} to a lower level as R_{max}'
  Delete R_{max} from DS_X
  Form a new $V_X = [R_1, R_2, \dots, R_{max}', \dots, R_{N_X+1}]$
  Take the new V_X and calculate P_{total} according to (6)
End While

Positive overload phase:

Take V and calculate P_{total}
Extract the subset $DS = [R_1, R_2, \dots, R_i, \dots, R_m, \dots]$
While $P_{total} > P_{max}$
  For each user *i* in DS
    Decrease the data rate R_i to a lower level as R_i'
    Form a vector $V' = [R_1, R_2, \dots, R_i', \dots, R_{N_X+1}]$, 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 data rate (R_{max}) supplies the greatest *RT* value: $R_{max} = argmaxRT(R_i)$
  Decrease the data rate R_{max} to a lower level as R_{max}'
  Delete R_{max} from DS
  Form a new $V = [R_1, R_2, \dots, R_{max}', \dots, R_{N_X+1}]$
  Take the new V and calculate P_{total}
End While

Fig. 3. Pseudo Code of the Data Rate Adaptation Scheme for UTRAN (in the case of having insufficient resources)

heuristic method.

1) Adaptation in a UTRAN network (in the case of having insufficient resources):

Given a data rate vector V_X, the required base station power P_{total} can be obtained according to (6). When there are not sufficient resources 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), $\frac{1-\alpha}{W} \times \sum_{n=1}^{N}(E_b/N_o)_n \times \upsilon_n \times R_n$ (denoted as D_{total} for simplicity) will be greater than 1 and P_{total} becomes negative. This is 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 are not sufficient resources, P_{total} becomes greater than P_{max}.

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. 3.

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 of suffering degradation but still comply with minimum QoS requirements. They are 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 $D_{total} - D'_{total}$ 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 $RT(R_i)$, where R_i is the data rate of service *i*. The service whose hypothetical data rate degradation results in the greatest ratio will be selected for an actual degradation. The selected data rate is denoted as R_{max}. After degradation, a new value, R'_{max}, is obtained. Then, R_{max} will be deleted from DS_X and the data rate vector V_X will be updated with R'_{max}. 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 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, $OF - OF'$. The candidate service whose hypothetical data rate (R_{max}) degradation results in the greatest ratio will be selected for an actual rate degradation. Then, R_{max} will be deleted from DS_X and the data rate vector V_X is updated with R'_{max} 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.

Given data rate vector $V_Y = [R_1, R_2, \dots, R_{N_Y}]$, calculate the Objective Function *OF* and P_{total}
Extract the subset $DS_Y = [R_1, R_2, \dots, R_i, \dots, R_m, \dots]$
While DS_Y is not empty
  For each user *i* in DS_Y
    Increase the data rate R_i to a upper level as R'_i
    Form a vector $V'_Y = [R_1, R_2, \dots, R'_i, \dots, R_{N_Y}]$, calculate P'_{total}
    If $P'_{total} \le P_{max}$
      Calculate *OF'* and the ratio $RT(R_i)$:
$$RT(R_i) = \frac{P'_{total} - P_{total}}{OF' - OF}$$
    Else
      Delete R_i from DS_Y
    End If
  End For
  Among the calculated *RT*s, select the service whose data rate (R_{min}) supplies the smallest *RT* value: $R_{min} = argminRT(R_i)$
  Increase the data rate R_{min} to a upper level as R'_{min}
  Delete R_{min} from DS_Y
  Form a new $V_Y = [R_1, R_2, \dots, R'_{min}, \dots, R_{N_Y}]$
  Take the new V_Y and calculate P_{total} according to (6)
End While

Fig. 4. Pseudo Code of the Data Rate Adaptation Scheme for UTRAN (in the case of having extra resources)

2) Adaptation in a UTRAN network (in the case of having extra resources):

Given a data rate vector V_Y, the required base station power P_{total} can be obtained according to (6). If V_Y leads to a P_{total} value which is smaller than P_{max}, the data rate adaptation scheme will increase the data rates of certain users so as to effectively consume the extra network resources and improve user satisfaction. The pseudo code of the scheme is presented in Fig. 4. The scheme extracts a subset DS_Y from V_Y. The DS_Y subset includes the data rates of the services which are prone for an increase in data rate. They are candidates for the improvement. In each round, the scheme will hypothetically increase the data rate of each candidate service belonging to DS_Y by one level and calculate the base station power P'_{total}. If P'_{total} is infeasible, the former data rate increase process will be cancelled and this candidate service will be deleted from DS_Y. If P'_{total} is feasible, the ratio of the difference of the base station power, $P'_{total} - P_{total}$, to the outcome difference of the objective function, *OF'–OF*, is calculated. The candidate service whose hypothetical data rate (R_{min}) upgradation results in the smallest ratio of the difference of the base station power to the outcome difference of the objective function will be selected for an actual rate upgradation. Then, R_{min} will be deleted from DS_Y and the data rate vector V_Y is updated with R'_{min} and the next round of adaptation will proceed until no more data rates can be increased without producing an infeasible P_{total}. The adaptation scheme aims to minimise the increase of power consumption and maximise the

```
Given 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 characteristic, calculate e_ncp
Extract subset DS_X = [R_1, R_2, ..., R_i, ..., R_m, ...]
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, ..., R'_i, ..., R_{N_X+1}], calculate e'_ncp and
    OF'
    Calculate the ratio RT(R_i):
              RT(R_i) = (e_ncp - e'_ncp) / (OF - OF')
  End For
  Among the calculated RTs, select the service whose data rate
  (R_max) supplies the greatest RT value: R_max = argmaxRT(R_i)
  Decrease the data rate R_max to a lower level as R'_max
  Delete R_max from DS_X
  Form a new V_X = [R_1, R_2, ..., R'_max, ..., R_{N_X+1}]
  Take the new V_X and service characteristics, calculate e_ncp
End While
```

Fig. 5. Pseudo Code of the Data Rate Adaptation Scheme for WiFi (in the case of having insufficient resurces)

improvement of user satisfaction.

3) Adaptation in a WiFi network (in the case of having insufficient resources):

Given a data rate vector V_X and service characteristics, 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 to 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 packets contention to access the network channel. The effective packet transmission rate and the end-to-end bandwidth calculated according to 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 shown in Fig. 5.

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 [9] and no further adjustment is available.

```
Given data rate vector V_Y = [R_1, R_2, ..., R_{N_Y}], calculate the Objective
Function OF and e_ncp
Extract subset DS_Y = [R_1, R_2, ..., R_i, ..., R_m, ...]
While DS_Y is not empty
  For each user i in DS_Y
    Increase the data rate R_i to a upper level as R'_i
    Form a vector V'_Y = [R_1, R_2, ..., R'_i, ..., R_{N_Y}], calculate e'_ncp
    If e'_ncp ≤ 1
      Calculate OF' and the ratio RT(R_i):
              RT(R_i) = (e'_ncp - e_ncp) / (OF' - OF)
    Else
      Delete R_i from DS_Y
    End If
  End For
  Among the calculated RTs, select the service whose data rate
  (R_min) supplies the smallest RT value: R_min = argminRT(R_i)
  Increase the data rate R_min to a upper level as R'_min
  Delete R_min from DS_Y
  Form a new V_Y = [R_1, R_2, ..., R'_min, ..., R_{N_Y}]
  Take V_Y and service characteristics, calculate e_ncp
End While
```

Fig. 6. Pseudo Code of the Data Rate Adaptation Scheme for WiFi (in the case of having extra resources)

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

4) Adaptation in a WiFi network (in the case of having extra resources):

Given a data rate vector V_Y and service characteristics, the expected number of contending packets e_{ncp} can be obtained. When the service types are hybrid, if V_Y leads to a e_{ncp} value which is smaller than 1, the data rate adaptation scheme will selectively increase the data rates of certain users so as to effectively consume the extra network resources and improve user satisfaction. The pseudo code of the scheme is presented in Fig. 6

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 have simulated a heterogeneous communication environment with two RANs, UTRAN and WiFi. The objective is to gauge the performance of the P-RASO algorithm and compare with the algorithm MUSE-VDA presented in [5]. In the simulation, the P-RASO algorithm implements two policies, the OUSI policy and the UNI policy. The P-RASO algorithm changes to use different policies according to the example as shown in Fig. 2.

However, the P-RASO algorithm is not confined to UTRAN and WiFi networks, but it can be performed in more complex situations. The way in which the objective function is adopted to measure and evaluate the UTRAN and the WiFi CSMA/CA network also can be applied to other wireless networks and technologies. Furthermore, the policies used by the P-RASO algorithm can be changed according to the aims of operators, different scenarios and network conditions (such as emergency situations).

In the simulations, six types of services are considered:

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 minimum requirements for each service class. The parameters of the objective function of each service types are presented in Table 2.

Considering the OUSI and UNI policies are used, the value of the objective function represents the level of user satisfaction. 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 video call service in an UTRAN network as *high*, and in the WiFi network as *very low*. During the selection process, these linguistic terms are converted to crisp numbers. For the linguistic term '*high*' and '*very low*', the crisp numbers are *0.717* and *0.091*, respectively. For a detailed explanation on this conversion process, please refer to [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) [11]. As

TABLE I. PARAMETERS AND TYPICAL VALUES

Service Type	SP	VC	VS	AS	WB	FT
Required Data Rate (kbps)	6.7/12.2	64/128	64/128	32/64	128/256	64/128
Transport Protocol	UDP	UDP	UDP	UDP	TCP	TCP
Proportion of Users	12.5%	12.5%	12.5%	12.5 %	25%	25%

the data rate increases, the user satisfaction also increases. A Sigmoid curve has a convex and a concave characteristic. When the data rates are quite low or very high for these types of service, an increase in data rate will not significantly improve the user satisfaction. 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 be noticed by 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. Therefore, for these service types, the *data rate* attribute is normalised by an Exponential function [12]. 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. Table 2 also shows examples for defining the weighting values.

In the simulations, we evaluate the performances of different algorithms in three aspects, joint (UTRAN and WiFi) network throughputs, objective function values and blocking probabilities, with the increasing network load (the ratio of user arrival rate to user departure rate, which is denoted as ρ).

Fig. 7 depicts the joint (UTRAN and WiFi) network

TABLE II. Parameters of OBJECTIVE 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***	$NORM(dr) = \begin{cases} 0, & dr < 6.7 \\ 0.5 + \frac{1}{2 + 2 \times e^{-2\times(dr-9.45)}}, & dr \geq 6.7 \end{cases}$ $NORM(m, WiFi) = 0.091$ $NORM(m, UTRAN) = 1$	$Weight_{dr} = 0.717$ $Weight_m = 0.5$
VC	*SNCL(VC, UTRAN)* = 0.717 *SNCL(VC, WiFi)* =0.091	Supplied date rate ***dr***; mobility support ***m***	$NORM(dr) = \begin{cases} 0, & dr < 64 \\ 0.5 + \frac{1}{2 + 2 \times e^{-2\times(dr-9.6)}}, & dr \geq 64 \end{cases}$ $NORM(m, WiFi) = 0.091$ $NORM(m, UTRAN) = 1$	$Weight_{dr} = 0.717$ $Weight_m = 0.5$
VS	*SNCL(VS, UTRAN)* = 0.717 *SNCL(VS, WiFi)* =0.091	Supplied date rate ***dr***; mobility support ***m***	$NORM(dr) = \begin{cases} 0, & dr < 16 \\ 1/\left[1 + e^{-0.06\times(dr-48.5)}\right], & dr \geq 16 \end{cases}$ $NORM(m, WiFi) = 0.091$ $NORM(m, UTRAN) = 1$	$Weight_{dr} = 0.909$ $Weight_m = 0.5$
AS	*SNCL(AS, UTRAN)* = 0.717 *SNCL(AS, WiFi)* = 0.283	Supplied date rate ***dr***; mobility support ***m***	$NORM(dr) = \begin{cases} 0, & dr < 32 \\ 1/\left[1 + e^{-0.1743\times(dr-24.25)}\right], & dr \geq 32 \end{cases}$ $NORM(m, WiFi) = 0.091$ $NORM(m, UTRAN) = 1$	$Weight_{dr} = 0.909$ $Weight_m = 0.283$
WB	*SNCL(WB, UTRAN)*=0.5 *SNCL(WB, WiFi)* =1	Supplied date rate ***dr***; mobility support ***m***	$NORM(dr) = \begin{cases} 0, & dr < 32 \\ 1 - e^{-5.85\times dr/384}, & dr \geq 32 \end{cases}$ $NORM(m, WiFi) = 0.091$ $NORM(m, UTRAN) = 1$	$Weight_{dr} = 0.717$ $Weight_m = 0.091$
FT	*SNCL(FT, UTRAN)* =0.283 *SNCL(FT, WiFi)* =0.909	Supplied date rate ***dr***; mobility support ***m***	$NORM(dr) = 1 - e^{-11.4\times dr/384}$ $NORM(m, WiFi) = 0.091$ $NORM(m, UTRAN) = 1$	$Weight_{dr} = 0.717$ $Weight_m = 0.091$

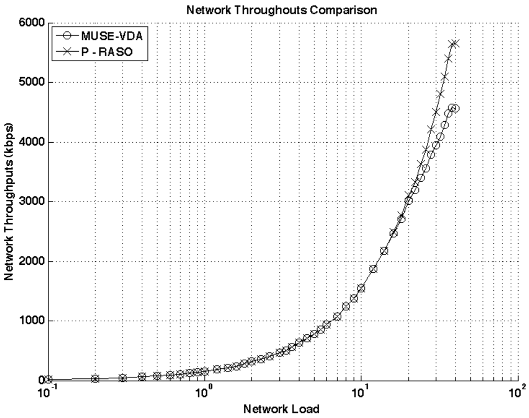

Fig. 7. Network Throughputs of Different Algorithms

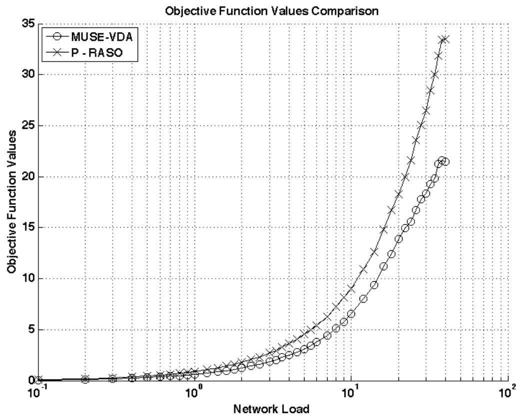

Fig. 8. Objective Function Values of Different Algorithms

throughputs for the MUSE-VDA and the P-RASO algorithms. When ρ is greater than 20, the network throughput introduced by P-RASO exceeds the one of MUSE-VDA. This value reaches about 5700 kbps when ρ is 40, while the throughput of MUSE-VDA is about 4600 kbps.

Fig. 8 presents the values of the objective function. When P-RASO is implemented, throughout the simulation, its value is always greater than the value of MUSE-VDA. The greater value of the objective function, the higher level of user satisfaction can be obtained.

Fig. 9 compares the blocking probabilities for different algorithms. In the simulation, the value of $r_{B/TR\_threshold}$ is assumed as 2%. When MUSE-VDA is used, the blocking probability starts to surge at ρ = 16 and gets over 2% at ρ = 20. When P-RASO is implemented, the blocking probability starts to grow at ρ = 26 and exceeds 2% at ρ = 40. At the end, the blocking probability of MUSE-VDA reaches 24% and the blocking probability of P-RASO grow to 4%.

These simulation results show that the P-RASO algorithm outperforms the MUSE-VDA algorithm. The use of the P-RASO algorithm provides greater throughputs, a higher level of user satisfaction and a lower blocking probability. By implementing the P-RASO algorithm, the overall network resources are effectively used to carry more traffic and the user satisfaction is improved. The use of P-RASO algorithm

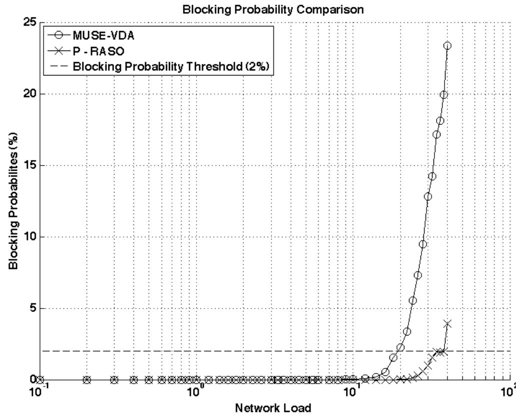

Fig. 9. Blocking Probabilities of Different Algorithms

provides a good performance, because it can dynamically and appropriately adjust the service classes of the existing users, and allow more requests to be admitted.

IV. CONCLUSION

In this paper, we also proposed a new adaptive and efficient RAN selection and optimisation algorithm, P-RASO. The P-RASO algorithm can be supplied with multiple policies and adaptively change to use different policies. It considers to effectively consume network resources and to maximise the gains obtained from network and optimisation. In the future, we will further study the performance of different policies and policy adaptations in more complex situations.

REFERENCES

[1] 3GPP. UTRA-UTRAN Long Term Evolution (LTE) and 3GPP System Architecture Evolution (SAE) [Online]. Available: ftp://ftp.3gpp.org/Inbox/2008_web_files/LTA_Paper.pdf

[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, WCNC 2007, pp. 3861-3866.

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

[6] 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

[7] Weizhi Luo and Eliane Bodanese, "Optimizing Access in an Integrated Wireless Network Environment", Proceeding of 2008 Internatianl Symposium on Performance Evaluation of Computer and Ttelecommunication Systems..

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

[9] 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.

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

[11] 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

[12] 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