Distributed radio access technology selection for adaptive networks in high-speed, B3G infrastructures

K. Tsagkaris*, †, G. Dimitrakopoulos, A. Saatsakis and P. Demestichas

University of Piraeus, 80 Karaoli Dimitriou str., Piraeus 18534, Greece

SUMMARY

Wireless systems migrate towards the era of ‘Beyond the 3rd Generation’ (B3G). A fundamental facilitator of this vision is the evolution of high speed, adaptive networks, needed for better handling the offered demand and improving resource utilization. Adaptive networks dynamically select their configuration, in order to optimally adapt to the changing environment requirements and conditions. This paper presents optimization functionality that can be used to support network adaptability (cognition-reconfigurability) in a B3G context. The paper starts from the business case that justifies the need for placing research onto adaptive networks and then continues with the management functionality for (re)configuration decisions, which is targeted to the dynamic selection of the appropriate radio access technologies (RATs). RAT selection is modelled through an optimization problem called (RAT, Demand and QoS-Assignment problem—RDQ-A), the solution of which assigns in a distributed manner the available RATs to adaptive Base Station transceivers and the demand (users) to these transceivers and to QoS levels, respectively. The RDQ-A optimization problem is decoupled in several sub-problems and is implemented in phases corresponding to the aforementioned assignments, while efficient custom greedy algorithms are mobilized in each phase for obtaining the optimum assignment. Finally, indicative results from the application of the proposed functionality to a simulated network are presented. Copyright © 2006 John Wiley & Sons, Ltd.

Received 1 December 2005; Revised 1 June 2006; Accepted 1 July 2006

KEY WORDS: high speed; B3G infrastructures; adaptive (cognitive-reconfigurable) networks; radio access technologies (RATs); optimization; greedy algorithms

1. INTRODUCTION

The gradual migration of today’s wireless communications (2G/2.5G/3G) towards the systems beyond 3G (B3G) reflects the most recent trend in the communications landscape. B3G wireless systems are recognized as such that can achieve high data rate transmissions and provide
adequate capacity, cost efficiency and highly sophisticated services, comparable to those offered by wired networks, for a variety of applications, such as interactive multimedia, VoIP, network games or videoconference. As long as the currently known radio access technologies (RATs) are not mature enough to satisfy the aforementioned criteria in a standalone manner [1], the idea of diverse RATs to be optimally combined and coordinated under a global infrastructure called ‘B3G wireless access infrastructure’ stands as a basic requirement for the consolidation of B3G systems. Major contributors towards this convergence are the cooperative networks concept [2,3] and the evolution of adaptive (cognitive-reconfigurable) networks [4]. In summary, the situation is depicted in Figure 1.

The cooperative networks concept assumes that diverse technologies such as cellular 2/2.5G/3G mobile networks and its evolutions (GSM/GPRS/UMTS/HSDPA), wireless local/metropolitan area networks (WLANs/WMANs), wireless personal area networks, (WPANs) and short range communications, as well as digital video/audio broadcasting (DVB/DAB) can be components of a heterogeneous wireless-access infrastructure and cooperate in an optimal way, in order to provide high speed and reliable connectivity anywhere and anytime [5]. The cooperation is materialized through the agreement for exchanging traffic or sharing spectrum among the cooperative Network Providers (NPs) and/or the joint configuration of network segments, providing assistance to each other in handling new traffic conditions or service management requests and maximizing the offered QoS levels.

Yet, the realization of the B3G concept based only on network cooperation may not be viable or efficient. First, there can be objections to a business model that requires extensive inter-NP cooperation. Additionally, the cooperative networks concept implies that the whole set of the alternate RATs should be deployed (installed and configured) a priori in both network segments and terminals, which would require constant, potentially risky, investments in software and hardware, whenever new technologies are introduced. Obviously, this can be inefficient, considering that all the technologies are not suitable for all the conditions.

Adaptability (cognition-reconfigurability) is seen as means to overcome the shortcomings identified above, mainly relying on the evolution and wide use of software defined radio (SDR)
Adaptive networks have the ability to dynamically adapt their behaviour (configuration) to the various conditions (e.g. hot-spot situations, traffic demand alterations, etc.) at different time zones and spatial regions, by exploiting deployments with much fewer pre-installed components. This process, in general, imposes software enabled (re)configuration actions which may affect all layers of the protocol stack. Such actions indicatively include RAT selection, spectrum allocation, algorithms selection and other software defined parameter configuration (at the PHY/MAC layer), TCP adaptation, IP QoS configuration, etc. (at the Network/Transport layer) or adaptation to appropriate QoS levels (at the Middleware/Application layer).

In this paper we focus on a distributed RAT selection process which in general aims at dynamically selecting and adapting to the most appropriate RAT for all terminals in the network, depending on the network conditions, system capabilities and service/demand aspects. RAT selection is modelled through an optimization problem called RDQ-A problem (RAT, Demand and QoS-Allocation problem). The solution of the RDQ-A problem consists in new reconfiguration actions, i.e. new allocations: (a) of available RATs to transceivers, (b) of demand (users) to transceivers and (c) of demand (users) to QoS levels. To this point, it should be noted that for the rest of this paper, the term transceiver refers to a reconfigurable transceiver located in any adaptive Base Station (Access Point) [6]. The solution method for obtaining the desired output takes part in four phases, each of which corresponds to one of the three aforementioned allocations (solution triplet) plus the final decision/selection phase. In general, the solution faced herein concerns only the network segment that faces the need for reconfiguration, i.e. it is focused on a specific problematic cell. Therefore, the problem has a distributed nature, since it can be solved at the level of the site, or set of sites that constitute a subset of the overall network.

While the management problem may seem complex, the solution is simple in its structure and implementation. Moreover, the application of such a scheme is essential for simplifying the operation of a wireless network, in a context involving multiple user classes, applications, network technologies. These different elements necessitate adaptability. Accordingly, we state that the proposed technique is the basis, which can be enhanced with machine learning capabilities, in order to accelerate reconfigurations, and make the adaptation procedure even simpler.

Regarding the cooperative networks concept, relevant research attempts have been made in the recent past especially concentrating on the allocation of traffic to the different RATs and networks, as well as with the allocation of users to QoS levels [7,8]. In addition, previous work to handle integration issues between 3G wireless systems and Wireless LANs has been presented, e.g., in [9,10]. Moreover, interesting techniques for network selection in an integrated wireless LAN and UMTS environment has been published in [11], while Gelabert et al. [12] has presented a framework for developing policy-based initial RAT selection algorithms within a common RRM context in a heterogeneous network. The above-mentioned papers refer to the procedure through which a terminal selects a network to initially camp on or handover to, and mainly concern 3G and WLAN technologies. On the contrary, the method presented herein concentrates in the network side and specifically into the dynamic RAT selection by base stations and furthermore, it extends to the inclusion of any kind of RAT with the use of reconfigurable transceivers, so as to facilitate the adaptation process in support of high speed, B3G infrastructures. This work comprises an extended and significantly improved version of the preliminary study that has been presented in [13].
The rest of this paper is structured as follows. Section 2 presents some supporting information on adaptive networks, as well as a business case envisaged to occur in B3G systems, that constitutes the motivation for this work. Section 3 describes the RDQ-A problem in detail. Sections 4 and 5 give the overall solution method and the detailed phased approach for acquiring the solution, respectively. Finally, results from the application of the method in two different scenarios of a simulated network are contained in Section 6, while concluding remarks are drawn in Section 7.

2. MOTIVATION, BUSINESS CASE AND HIGH LEVEL PROBLEM DESCRIPTION

2.1. Motivation for research in adaptive networks

Adaptive networks have been introduced with the goal to offer to users, manufacturers, NPs, service providers and software providers the capability of deploying and using a bunch of innovative services, at high speeds and affordable cost, in different heterogeneous contexts, using diverse equipments and through several technologies. This is achieved by allowing their segments to dynamically select and configure the set of the most appropriate RATs, in order to better handle service area region or time variant requirements. This is graphically depicted in Figure 2, where the base stations of a composite radio access network (RAN) adapt to network-related requirements by means of changing their transceivers’ operating RAT, including the capability of using a completely new RAT.

In such context, it is envisaged that all actors of the wireless world will experience significant benefits: users will experience ubiquitous connectivity; manufacturers will benefit through the reduction in the risk from investments in new technologies, through simplified product evolution and reduced development costs; additionally, NPs will have more options for providing the required QoS and capacity levels through software enabled reconfigurations of their own infrastructure and will be able to introduce value-added services decreasing capital expenditure (CAPEX) and controlling operational costs (OPEX).

![Figure 2. Adaptive networks’ capabilities overview.](https://example.com/figure2.png)
In general, it is envisaged that B3G environments will comprise heterogeneous technologies, which may require constant investments in hardware. The use of equipment which is reconfigurable through software, promises to bring to the foreground the capabilities of selecting the most attractive network, of achieving seamless interoperability among different standards and of adapting to the optimized radio transmission characteristics, according to the environmental transformations. In this way it is believed that adaptive networks will facilitate the easy penetration of new technologies while also overcoming the constraint for costly hardware deployment schemes. Of course, adaptive networks require proper management functionality for determining the reconfigurations. This is the subject and contribution of the paper. Furthermore, we advocate the usefulness and practicality of such an approach relying on the flexibility of software and the great evolution of SDR technology.

2.2. Business case

This section aims at exemplifying the role of adaptive networks and raising the issue of management functionality for supporting their adaptability. The business case assumes that a NP, called target NP, operating a network as described and owning several reconfigurable transceivers (i.e. that are capable of operating with one or more RATs), faces a situation, which constitutes the trigger for reconfiguration, somewhere within their administrative domain. Typically, the situation faced would be that the target QoS and capacity levels cannot be achieved and thus the appropriate adaptation of the network is required, so as to overcome the relevant difficulties. A potential solution to the problem would be appropriate traffic redistribution among the available NPs [7]. Additionally, adaptive networks assume that the target NP can change the RATs in the transceivers and accordingly allocate resources in its own network. These changes result also to new allocations of demand to transceivers as well as to QoS levels.

A high-level scenario corresponding to the business case is shown in Figure 3. The scenario evolves in three main phases. During the first phase, the NP recognizes the need for adaptation to a new situation, considering system and demand aspects. At the same time, the co-operating NPs’ offers are acquired and managed accordingly, i.e. NPs are asked whether they can participate in the handling of the event [14]. The second phase is related to the reconfiguration decision process, by applying new management functionality (addressed by the problem discussed herein). Finally, in the third phase, the reconfiguration is implemented. This includes download, installation and validation of software components as well as notification of the cooperating NPs for the portion of the demand they will absorb, and potentially establishment of new agreements with the target NP.

2.3. High level problem description

The goal of this work is targeted at the functionality for reconfiguration actions in support of networks’ adaptability. Specifically, the problem addressed (RDQ-A problem) uses as input some system aspects, as well as service area and demand aspects and aims at optimally assigning the available RATs to transceivers, the demand to transceivers and demand to QoS levels, by respecting some meaningful constraints. The RDQ-A problem can be generally described as in Figure 4 and mathematically formulated as follows.
3. RDQ-A—PROBLEM FORMULATION

The RDQ-A problem can be generally described as aforementioned by a certain input and the desired output, acquired through respecting the relevant constraints (Figure 4).

3.1. Input

The input of the RDQ-A problem provides information on the system, as well as on the service area and demand as follows.

*System aspects:* The system aspects provide information on the available transceivers and the potential operating RATs. Set $R$ represents the set of available RATs in the system. The distributed approach of the problem imposes that we focus on one site, equipped with a set of
reconfigurable transceivers, denoted as set $T$. Finally, set $R_t, (t \in T)$ represents the set of RATs with which transceiver $t$ can work. In general, $R_t \subseteq R$.

Service area and demand aspects: We consider $N$ active users that comprise the demand volume in the service area. The location of such a user $i \in \{0, 1, \ldots, N\}$ is denoted as $l_i$. The services (applications) offered in the service area are represented through set $S = \{s : s = 0, 1, \ldots, |S|\}$. Each user $i$ requires a service $s_i (s_i \in S)$. By assumption, a physical user requiring more than one services is modelled by more than one indexes ($i$) in the set.

Moreover, set $Q_s = \{q : q = 0, 1, \ldots, |Q_s|, s \in S\}$ represents the sets of various quality levels at which service $s$ should be offered. Similarly, set $R_s = \{r : r = 0, 1, \ldots, |R_s|, s \in S\}$ represents the RATs through which service $s$ can be offered. In general, $R_s \subseteq R$.

Additional requirements are the utility volume and the resource consumption, when a service is offered at a certain quality level. Specifically, the utility volume for user $i$ that renders service $s_i$ at quality level $q_i$ is denoted as $u_{i,q_i} ((s_i,q_i) \in (S \times Q_s))$. The term ‘utility’, borrowed from economics [15] presents the degree of user satisfaction gained from the consumption of network resources.

The required network resources (e.g. loading factor, bandwidth) for user $i$ which uses service $S_i$ at quality level $q_i (q \in Q_s), are$ denoted as $b_{i,q_i} ((s_i,q_i) \in (S \times Q_s))$.

3.2. Output

As already stated, the objective (output) of the RDQ-A problem is to determine new configurations, i.e. new allocations (a) of RATs to transceivers, (b) of demand (users) to transceiver/RATs, and (c) of demand (users) to QoS levels. The output allocations can be described as follows:

(a) The transceiver configuration (allocation of RATs to transceivers) is described through the set $A_{RT} = \{r_t \forall t \in T\}$. Each element $r_t (r_t \in R_t)$ denotes the RAT assigned to transceiver $t$.

(b) The allocation of demand (users) to transceivers is denoted as $A_{DT} = \{t_i \forall i \in [1,N]\}$. Each element $t_i$ represents the transceiver to which user $i$ is bound.

(c) Finally, the allocation of demand (users) to QoS levels can be described as $A_{DQ} = \{q_i \forall i \in [1,N]\}$. Each element $q_i (q \in Q_s)$ is the quality level that will be offered to user $i$ (that utilizes service $s_i$ from transceiver $t_i$).

The three allocations constitute the solution triplet for the RDQ-A problem and should maximize the utility which is associated with the assigned QoS levels, while minimizing the resulting reconfiguration cost related to the assignment.

4. OVERALL SOLUTION METHOD

The proposed solution method for the RDQ-A problem is implemented in four phases highly collaborating with each other, as depicted in Figure 5. Each phase exploits the output of its preceding phase, whilst it produces appropriate output to be used in the next phase, and so on.

Specifically, during the first phase, all candidate allocations of RATs to transceivers ($A_{RT}$) are found, considering the aforementioned system aspects. Each $A_{RT}$ allocation corresponds to the launch of a sub-problem. The resulting sub-problems obtained are then processed in parallel.
In the second phase, for each of the sub-problems (with fixed ART allocation), the quality levels offered to users are set to their lowest possible values (basic ADQ) and the allocation of demand to transceivers (ADT) is extracted following the algorithm described in Section 5.2.

Then, in the third phase, the optimum allocation of demand (users) to QoS levels (ADQ) is explored by retaining the ADT allocation from the previous phase and continuously increasing the offered quality levels according to the algorithm presented in Section 5.3.

Finally, the fourth phase contains the selection of the best allocations’ triplet, in terms of maximization of an objective function related to utility, i.e. the level of users’ satisfaction and the minimization of the associated cost, as described in Section 5.4. Eventually, the optimum, selected reconfiguration decisions are translated into proper reconfiguration actions, which must be applied in order to drive the network elements to efficiently adapt to the new conditions.

The individual input, output and processing procedures in each of the four phases that comprise the solution method and were briefly described above, are thoroughly presented in Section 5.

5. DETAILED PHASED APPROACH

5.1. Phase 1— allocation A_Rt

Given the sets |R| and |T| that denote the number of available RATs and transceivers respectively, this phase results in the identification of the total amount of allocations A_Rt, i.e. |R|^|T| sub-problems to be launched in parallel. Under specific conditions though, proper
reduction of the problem size can be achieved by exploiting certain technological aspects, such as the assumption that all transceivers have the same capabilities when operating in the same RAT. Additionally, allocations are excluded from the searching space if for a user \(i\) there is not a candidate set \(T_i (T_i = \emptyset)\), probably because none of the available transceivers can either cover user \(i\) or provide the required service to him. Eventually, the above lead to \(m \leq |R|^{|T|}\) allocations \(A_{RT}\).

Moreover, as long as a specific \(A_{RT}\) allocation is isolated, the RAT with which a transceiver \(t\) is assigned is known \(= r_t\) and therefore, RAT-specific characteristics like the coverage \((\text{cov}_t, t \in T)\) and peak capacity \((\text{cap}_t, t \in T)\) of that transceiver can be deduced. Additionally, set \(S_t (t \in T)\) represents the services that are offered through transceiver \(t\).

5.2. Phase 2—allocation \(A_{DT}\) (basic \(A_{DQ}\) allocation)

Given the output of the previous phase \((A_{RT}, \text{cov}_t, \text{cap}_t, S_t, T_i)\) the second phase aims at finding the optimum \(A_{DT}\) allocation assuming that the QoS levels offered are the lowest acceptable ones, denoted as basic \(A_{DQ}\) allocation i.e. \(q_i = 0, \forall i \in [1 \ N]\). For this purpose, a custom, computationally efficient, greedy algorithm has been designed. To this point, some new magnitudes need to be introduced:

\(U^c_t = \{i : t \in T_i\}\): The set of users that the transceiver \(t\) belongs to their candidate set \(T_i\).
\(pd_t\): The potential demand for transceiver \(t\), i.e. the network resources that users belonging to \(U^c_t\) would require from that transceiver. It can be expressed through \(pd_t = \sum_{i \in U^c_t} b_{si,qi}\).
\(rc_t\): The remaining capacity for transceiver \(t\).

The algorithm for obtaining allocation \(A_{DT}\) takes place in steps (Figure 6) as follows:

**Step 1**: Sorting users.

First, users are sorted beginning from the ones with the smaller candidate set \(T_k\) towards the ones with the biggest \(T_k\). In case that two or more users have equal candidate set size, the user that will precede the others in the sorted list is selected in a random way. To continue, we pick the first user from the sorted list e.g., user \(k\).

**Step 2**: Allocating user to a transceiver.

The transceiver that will serve user \(k\) is selected among the \(|T_k|\) candidate ones, according to the following relations which must be simultaneously satisfied:

\[
\sum_{\{i: i \neq k\}} b_{si,qi} + b_{sk,qk} \leq \text{cap}_t
\]

(1)

\[
t_k = \arg\min_{i \in T_k} (pd_i/rc_i)
\]

(2)

Relation (1) ensures that the capacity constraint for the selected transceiver is not violated. According to relation (2) the algorithm gives priority to those transceivers which have either low potential demand or high remaining capacity.\(^1\) If none of the transceivers satisfies the above relations (1) and (2), the algorithm fails, but only for this specific sub-problem.

\(^1\)When \((pd_i/rc_i) = (pd_{i'}/rc_{i'})\) for two (or more) transceivers, the selection is done as follows: If \(pd_i \neq pd_{i'}\) (or equivalently \(rc_i \neq rc_{i'}\)) for two transceivers \(t \neq t'\), the transceiver with the greater potential demand is selected. If \(pd_i = pd_{i'}\) and \(rc_i = rc_{i'}\), the selection takes place with the same probability.
Step 3: Update values.

For every transceiver $t \in T_k$ quantities $p_{dt}$ are updated. Furthermore, quantity $r_{c_k}$ of the
selected transceiver $t_k$ is also updated. If there are more users to allocate, pick up the next one
and return to Step 2, else finish the algorithm with Step 4.

Step 4: END.

By the end of the algorithm, an allocation of users to the available transceivers is proposed.
The set of users bound to transceiver $t$ is $U_t = \{i : t_i = t\}$.

5.3. Phase 3—allocation $A_{DQ}$ (QoS levels improvement)

In the beginning of phase 3 we have knowledge of the basic $A_{DT}$ allocation, as well as of each
transceiver’s remaining capacity($r_c$). In this phase, the QoS levels to be offered by each
transceiver to its serving users are gradually incremented in a greedy fashion, considering the
criterion of the maximization of the following objective function:

$$OF_t = \sum_{i \in U_t} \sum_{i \in U_t} [u_{s,i} - c_{s,qi}(l_i, r_{c_i})]$$  \hspace{1cm} (3)
where the quantity of \( i \) expresses each user’s contribution to their transceiver’s total objective function (OF) value. The OF to be maximized is associated with the utility deriving from the QoS levels allocation in the service area and the associated cost, depending on the quality level assigned to service \( s_i \), on the user location \( l_i \) and on the transceiver’s operating RAT \( r_{ti} \), extracted from phase 1 \((A_{RT})\). The related algorithm (Figure 7) aims at obtaining the final \( A_{DQ} \) allocation by giving priority to users that tend to increase the objective function and can be analysed as follows.

For every transceiver \( t \in T \)

**Step 1:** Create List.

**Step 1.1:** Compute the each user’s tentative contribution \((o^{t}_i)\) to \( O_{F_t} \). The term tentative refers to the computation of value \( o_i \) assuming that the QoS level is incremented by one (i.e. from level 0 to level 1). Users are excluded from the list in case they cannot be improved further (e.g. pure voice users).

**Step 1.2:** The rest are marked as ‘not examined’ and sorted in descending order according to the computed contributions.

![Figure 7. Phase 3—solution algorithm for allocation \( A_{DQ} \).](image-url)
Step 1.3: If the list is empty, the algorithm ends (Step 4), else select the first user from the sorted list e.g. \( k \) and continue with next step.

Step 2: Capacity check.
Step 2.1: Increment the QoS level by one\( (q_k := q_k + 1) \).
Step 2.2: Check whether the capacity of the serving transceiver is violated, i.e.

\[
\left( \sum_{\{i : t_i = t, i \neq k\}} b_{s_i,d_i} \right) + b_{s_k,d_k} > \text{cap}_t
\]

- If true, return to the previously allocated QoS level \( (q_k := q_k) \) and remove that user from the list. If the list is not empty pick next user and continue with Step 2, else transfer to Step 4.
- If false, continue with Step 3.

Step 3: QoS Improvement.
Step 3.1: Increment the selected user to the next higher QoS level\( (q_k := q_k + 1) \).
Step 3.2: Update the of\( k \) value for user \( k \). Mark the user as ‘examined’.
Step 3.3: If there are more ‘not examined’ users in the list, pick up next user and continue with Step 2. Otherwise, return to Step 1 and create the new list with the users that can be further improved.

Step 4: END.

To this point, we must state that there exist alternative mechanisms to the network adaptive approach; for example, terminals can adapt by accepting lower QoS levels for services such as voice or video. Our approach acts in a complementary manner, providing users with constant QoS, which is as high as possible (but also required) considering also profile information.

5.4. Phase 4—selection of the best triplet

After the completion of the 3 first phases of the solution method, the optimum triplet \((A_{RT}, A_{DT}, A_{DQ})\) has to be selected among the \( m (\ll |R|^{|T|}) \) possible combinations. A reasonable criterion for that selection is the maximization of the total objective function, which can be deduced by summing values in (3) over all the transceivers:

\[
\text{OF}_{\text{tot}}(A_{RT}, A_{DT}, A_{DQ}) = \sum_{t \in T} \text{OF}_t \tag{4}
\]

Furthermore, it is very important to embody into the selection process the cost from the reconfigurations, \( C_{\text{reconf}} \), imposed by the application of the solution triplet, \((A_{RT}, A_{DT}, A_{DQ})\). Specifically, when we are to decide among two solutions giving the same or near the same performance in terms of OF values, we prefer the one that requires the minimum number of reconfigurations. Thereby, we give priority to the solution that leads to less firmware downloads and installations or less signalling overhead for the involved network elements. Obviously, one should be aware of the previous state of the network in question, so as to decide for the solution with the minimum reconfigurations.

We distinguish three different types of possible costs reflecting to messages that have to be sent for a particular reconfiguration action that causes a transition from network state \( n \) to
network state $n + 1$. Each of the assumed types of cost contributes additively to the total reconfiguration cost, and can be classified either into base station side or into user terminal side:

**Cost associated with base station side**

- $c_T$ associated with the number of base station transceivers, that change their operating RAT during the transition from state $n$ to state $n + 1$.

**Cost associated with user terminal side**

- $c_R$ associated with the number of user terminals that change their serving RAT during the transition from state $n$ to state $n + 1$.
- $c_Q$ associated with the number of increase/decrease steps in the QoS levels of a user terminal during the transition from state $n$ to state $n + 1$.

Accordingly, relation (5) describes the total reconfiguration cost as a weighted sum of the three contributing cost components presented above.

$$C_{\text{reconf}}(n \rightarrow n + 1) = w_1 \cdot c_T + w_2 \cdot c_R + w_3 \cdot c_Q$$

where $w_1$, $w_2$ and $w_3$ are real value weights that could be appropriately adjusted according to the importance that any NP would have planned to give to each distinct type of cost. The exact determination of these quantities will be part of our future research. For the results section that follows, they will be set equal to 1.

6. RESULTS

This section makes use of 2 different scenarios, in order to validate the method's effectiveness, by applying it to a simulated network. The scenarios are differentiated with respect to the transceivers' capabilities, i.e. the first one deals with an adaptive network capable of operating in UMTS/WLAN, while the second one considers the utilization of HSDPA/WiMAX, as the transceivers' operating RATs.

6.1. Adaptive network with UMTS–WLAN capability

The first scenario considers the use of UMTS/WLAN as being available for the reconfigurable transceivers and aims at providing evidence on the advantages that derive from the capability of a network to adapt to environment-originated requirements, i.e. when it should cope with a certain (high) load situation. To do so, let us consider a simple service area layout and structure as the one depicted on Figure 8. The whole area consists of a number of cells/sites. We assume that a random, shady cell in the service area layout faces a situation that creates the need for reconfiguration, e.g. in order to handle the load more efficiently. Thus, the target NP which controls the problematic site accepts a reconfiguration trigger. We concentrate on this specific cell and apply the solution method to the RAT selection problem (in a distributed manner). The latter statement implies that no action is taken regarding the resource configuration, e.g. spectrum or interference management and furthermore, we are not aware of the offers and agreements between different NPs.

The cell is equipped with 3 reconfigurable transceivers that may operate in one of the UMTS and WLAN (802.11b) RATs. Additionally, a maximum allowed value of 0.6 is allowed for
UMTS loading factor, while the maximum capacity value for IEEE 802.11b is taken almost equal to 5.5 Mbps.

The set of available services consists of 2 services. A voice service ($s_1$) and a data service ($s_2$). The voice service can be offered only through UMTS. For the data service a set of quality levels is provided. Table I contains the bandwidth requirements per service, the QoS levels offered, as well as the utility volume whenever a service is offered at a specified QoS level.

To accurately model the demand in the service area, an appropriate configuration build upon average values is needed. Specifically, we consider a demand in the problematic cell initially consisting of 260 active voice and data sessions and in the sequel, the method is invoked for 9 different cases. Sessions are randomly created and uniformly distributed within the cell. Each test case corresponds to a different combination of sessions per service, as depicted in Table II (after the appropriate rounding). The number of sessions per service $s$, denoted as $X_s$, is a random variable, which is distributed in the interval defined by $M_s \pm dM_s$ bounds, where $M_s$ denotes a mean value and $dM_s$ corresponds to 3 sessions. Each test case is executed several times in order to achieve a satisfactory level of fairness and the average result is obtained.

**Allocation $A_{RT}$:** As analysed above, in the first phase of the solution method the problem is split into 4 sub-problems corresponding to different $A_{RT}$ allocations. As a result, notation UUU equals to the case where all the 3 transceivers are configured to operate in UMTS. Similarly, notation UUW means that 2 out of the 3 available transceivers are configured to operate in UMTS and the remaining one in WLAN technology and so on. Actually, the total number of resultant sub-problems would be $|R|^T = 8$ (as defined in Section 5.1). However, assuming that all transceivers have the same capabilities when operating in the same technology, there is no need to separately examine configurations like UUW and WUU. So, this number reduces to 4 (UUU, UUW, UWW and WWW). Additionally, allocation WWW is excluded from the

![Figure 8. UMTS–WLAN interworking: service area layout and structure.](image-url)
searching space by the assumption that the existence of voice service imposes at least one of the transceivers to operate in UMTS. As long as the sub-problems have been determined we may continue with the rest phases.

**Allocation ADT (basic ADQ allocation):** In this phase, as the QoS levels are set to their lowest possible values for each sub-problem (each one with fixed ART allocation), the basic ADT allocation is obtained for each configuration. Figure 9 shows the number of sessions allocated to each transceiver for the 9 cases considered. It should be noted though that the results reflect the consideration of uniform distribution of sessions within the cell. We can observe that configuration UWW is conceptually inappropriate for cases 1–6, due to the large number of voice sessions that cannot be served by 1 UMTS transceiver.

**Allocation ADQ (QoS levels improvement) and selection:** After the gradual augmentation of the QoS levels, a solution to our problem is attainable, by obtaining the final, average OF values for all acceptable configurations (ART). Figure 10 shows the performance of the three possible configurations: UUU, UUW and UWW. Performance is expressed in terms of the average aggregate utility volume that derives from the QoS levels offered.

In general, the demand in case 1, since it contains only voice sessions, cannot utilize any configuration with WLAN technology. Thus, only configuration UUU makes sense. So, the value of the aggregate utility volume, indicatively for case 1, derives from the fact that there are 260 sessions at QoS level 1 (with utility 1). Carefully examining configuration UUU at a first

---

**Table I. Bandwidth requirements, QoS levels and utility volume.**

<table>
<thead>
<tr>
<th>QoS Levels</th>
<th>Voice</th>
<th>Data</th>
<th>Voice</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bit Rate (kbps)</td>
<td>Utility volume</td>
<td>Bit Rate (kbps)</td>
<td>Utility volume</td>
</tr>
<tr>
<td>0</td>
<td>16</td>
<td>1</td>
<td>32</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>64</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>128</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table II. Adaptive network with UMTS/WLAN capability: load cases.**

<table>
<thead>
<tr>
<th></th>
<th>Data</th>
<th>Voice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>Sessions</td>
</tr>
<tr>
<td>Case1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Case2</td>
<td>6.5</td>
<td>16</td>
</tr>
<tr>
<td>Case3</td>
<td>14</td>
<td>32</td>
</tr>
<tr>
<td>Case4</td>
<td>23</td>
<td>48</td>
</tr>
<tr>
<td>Case5</td>
<td>33</td>
<td>64</td>
</tr>
<tr>
<td>Case6</td>
<td>45</td>
<td>80</td>
</tr>
<tr>
<td>Case7</td>
<td>59</td>
<td>96</td>
</tr>
<tr>
<td>Case8</td>
<td>76</td>
<td>112</td>
</tr>
<tr>
<td>Case9</td>
<td>97</td>
<td>128</td>
</tr>
</tbody>
</table>

---

DISTRIBUTED RADIO ACCESS TECHNOLOGY SELECTION

Copyright © 2006 John Wiley & Sons, Ltd.

Int. J. Commun. Syst. 2007; 20:969–992
DOI: 10.1002/dac
stage, we note that as data load increases, the value of OF increases accordingly. This happens for cases 1–5, because more data sessions can be offered higher QoS levels. After case 5, no more capacity is available for increasing the QoS levels offered to ‘coming’ data sessions. Consequently, the value of OF initially (cases 5–8) remains the same, since the increase due to coming data sessions is equivalent to the decrease imposed by the decreasing voice sessions. Finally, at case 9, the data sessions have become so many, that the QoS levels offered need to be decreased, compared to case 8, and thus, the value of OF decreases.

Regarding UUW configuration, the initial load in the system exceeds the capacity of the 2 UMTS transceivers and thus, the value of OF is zero in case 1. Beginning from case 2, the voice

Figure 9. Adaptive network with UMTS/WLAN capability: basic $A_{DT}$ allocation: (a) UUU; (b) UUW; and (c) UWW.
sessions have decreased so that they can be served by 2 UMTS transceivers. Consequently, a value for OF is obtained. This value increases as the data sessions increase (along with a total increase in the system’s load), as more and more sessions can be offered higher QoS levels, by both WLAN and UMTS transceivers. e.g. in case 5 the aggregate value is achieved by offering QoS level-3 (utility volume 8) to half of the data sessions, while the rest of the data sessions obtain at least the basic QoS level. So 132 voice sessions contribute utility volume 1 each, half of the data sessions contribute utility volume 8 each, and the rest contribute at least 2 each.

Finally, configuration UWW in cases 1–6 leads to a value of zero for the OF, as the available UMTS transceiver cannot serve the voice sessions. After case 6, the ‘U’ transceiver is not anymore problematic, as it can serve the voice sessions, as well as the data sessions that fall in its service area, due to the uniform distribution of demand. So, we obtain an OF value also for UWW, which increases along with the decrease in the voice sessions, e.g. in case 8, voice is served by the ‘U’ transceiver. The rest of the capacity offers at least QoS levels-1 and 2 to half of the data sessions (aggregate utility value is approximately 290).

Comparing now the available configurations, we find that, at the very initial demand patterns, where voice sessions dominate, UUU configuration is the only feasible one. However, from case 2 up to case 3, a note-worthy superiority of UUW configuration is observed. The two ‘U’ transceivers are adequate for serving voice, as well as data sessions outside the range of the ‘W’ transceiver. The ‘W’ transceiver can be dedicated to offering higher QoS levels, to data sessions in its coverage range. This is also depicted on Figure 11, that contains the basic (Figure 11(a)) and final (Figure 11(b)) $A_{DQ}$ allocation in case 2 for data sessions. It can be easily observed that the number of data sessions that are served with higher QoS levels is bigger in WUU, than in UUU. However, in case 4, UUU is the most appropriate configuration, since its OF value, having been increasing since case 1, has reached a higher level than the WUU OF value. This is depicted on Figure 12, that contains the basic (Figure 12(a)) and final (Figure 12(b)) $A_{DQ}$ allocation for data sessions in case 4, of which their number offered higher QoS levels is bigger in UUU configuration and leads accordingly to a bigger OF value. Moreover, it should be noted that, if we were supposed to face a case where the OF values obtained for UUU and UUW were equivalent, the final selection would depend on the network’s previous state. If we suppose, e.g. that the network’s previous configuration was UUU, we would result in the selection of UUU again, in order to minimize the number of necessary RAT changes.
Figure 11. Adaptive network with UMTS/WLAN capability: $A_{DO}$ allocation for case 2: (a) basic; and (b) improved.

Figure 12. Adaptive network with UMTS/WLAN capability: $A_{DO}$ allocation for case 4: (a) basic; and (b) improved.
6.2. Adaptive network with HSDPA/WiMAX capability

This scenario aims at proving the effectiveness of the proposed method in adaptive networks that witness the emergence of innovative standards, such as High-Speed Downlink Packet access (HSDPA) which has been recently standardized for UMTS [16], as well as WiMAX (Worldwide Interoperability of Microwave Access) [17], which is a wireless technology that provides high-throughput broadband connections over long distances.

For this purpose, the service area previously described has been appropriately restructured, in order to comply with these two emerging standards capabilities (Figure 13), and assuming that the transceivers may now operate in either UMTS/HSDPA or WiMAX. This implies that, with respect to UMTS/HSDPA, voice sessions are served by UMTS with a maximum loading factor value of 0.6 and data sessions are bound to HSDPA with its peak capacity taken equal to 6 Mbps. WiMAX capacity, on the contrary, is considered to reach 15 Mbps, but supporting only limited mobility. So, only a percentage of the total number of voice sessions can be served by WiMAX, while the rest ones can be served exclusively by UMTS/HSDPA, due to their increased mobility characteristics. For our scenario, this percentage follows a uniform distribution (with a mean value of 25%), along 9 test cases that comprise data and voice sessions, as shown in Table III.

Allocation $A_{RT}$: In the first phase of the solution algorithm, the same approach is followed, i.e. the overall problem is split into 3 different sub-problems that are launched in parallel, each of which has a fixed $A_{RT}$ allocation. The sub-problems obtained are again 3, i.e. UUU, UUW and UWW. $^8$ To this point, it should be noted that we conceptually exclude WWW configuration, $^8$The notations used are (for legibility purposes) equivalent to the previous scenarios, i.e. U = HSDPA and W = WiMAX.
since according to today’s WiMAX capabilities, the probability of not having even one high-mobility user in a cell, which cannot be served by WiMAX, is negligible.

Allocation $A_{DF}$ (with basic $A_{DF}$ allocation), QoS levels improvement and selection: The results can be summarized in Figure 14, where we see the evolution of the OF value for the three acceptable configurations. The intermediate results are omitted for brevity reasons.

Examining configuration UUU, we note the OF value augmentation along with the increase in the oncoming data sessions, until the system has reached its peak capacity (case 5). This is followed by a reduction in the OF value invoked by the large number of data sessions, which cannot be handled efficiently by 3 UMTS/HSDPA transceivers and consequently the QoS levels offered to existing sessions need to be decreased. Moreover, UUW’s OF value increases in a similar manner, noting that the capacity now is much higher. On the contrary, UWW is considered inappropriate when no data sessions are existent, while its OF value follows the same as above rule, providing though the highest results, due to its large total capacity.

Comparing now the acceptable configurations (phase 4 of the solution method), the existence of data sessions results, in general, in a surveillance of WiMAX transceivers. As data sessions increase, switching a second transceiver to WiMAX equals to increasing users’ satisfaction. This constitutes a strong argument for the deployment of transceivers supporting WiMAX, which

<table>
<thead>
<tr>
<th>Case</th>
<th>Data %</th>
<th>Sessions</th>
<th>Voice %</th>
<th>Sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>300</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>20</td>
<td>93</td>
<td>260</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>40</td>
<td>85</td>
<td>220</td>
</tr>
<tr>
<td>4</td>
<td>25</td>
<td>60</td>
<td>75</td>
<td>180</td>
</tr>
<tr>
<td>5</td>
<td>36</td>
<td>80</td>
<td>64</td>
<td>140</td>
</tr>
<tr>
<td>6</td>
<td>50</td>
<td>100</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>66</td>
<td>120</td>
<td>34</td>
<td>60</td>
</tr>
<tr>
<td>8</td>
<td>87.5</td>
<td>140</td>
<td>12.5</td>
<td>20</td>
</tr>
<tr>
<td>9</td>
<td>100</td>
<td>160</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 14. Adaptive network with HSDPA/WiMAX capability: OF evolution.

Table III. Adaptive network with HSDPA/WiMAX capability: load cases.
however should contradict to the associated cost, mostly in cases (such as cases 1–3), where the OF values obtained from all configurations are equivalent. The selection, thus, of the most suitable configuration depends on the overall networks’ previous state, so that to minimize the associated cost. Considering now that 3G networks should only need a slight software upgrade so as to support HSDPA, one would consider UUU as ‘the previous state’, assuming that the cost for the deployment of HSDPA is lower than the respective cost for introducing a completely new technology (WiMAX). This results in a higher ranking for UUU configuration in such cases, while UUW and UWW follow.

In general, the solution method brings into view the benefits of adaptability that future systems will incorporate, while today, the allocation of the demand to RATs is much limited by the fixed configuration and thus, the optimality of the allocation for the specific demand cannot be ensured. On the contrary, adaptive networks are based on the optimality by searching among a number of possible configurations for the 3 transceivers. Furthermore, 3 reconfigurable transceivers could be lodged in the same site extensively reducing hardware requirements and the necessary signalling information between network segments. Finally, one could claim that the algorithm presented herein takes into account the users’ requirements for high bit rates, through the incorporation of the utility function. At the same time, NPs consider their cost for each service provision pattern. The success of the algorithm thus lies in the identification of a suitable balance among these approaches, in order to guarantee for the maximum reliability of an adaptive network.

7. CONCLUSIONS

This paper presented the major challenges that adaptive networks meet; it outlined the migration of network management based on reconfiguration and flexible reallocation of networking resources and described the major reasons that necessitate certain reconfiguration actions within a multi-RAT environment, so as to adapt to environmental requirements. Moreover, it described novel management functionality targeted at adaptive networks, aiming at the selection of the most appropriate RATs in order to cope with excessive load situations. The relevant problem was mathematically formulated and results were presented, for validation and consolidation of the functionality’s alleged benefits.

The solutions presented herein are based on a set of case studies. Specifically, the work reveals the advantages of adaptive networks compared to the conventional ones. Furthermore, the procedure of modelling methods and solutions to management problems raised up in this paper is not only suitable for those systems studied in this paper, but also provides a generalised framework for any situation involving cooperative radio networks and frequency bands. For such a radio environment encountered more and more often in future wireless systems, a certain co-ordination between the RATs as well as the NPs seems a basic prerequisite, so that the global optimisation can be obtained. However, more advanced mechanisms that consider all possible reconfigurable elements (terminals, protocol stacks, etc.) are yet to be developed. This will include part of our future activities, along with the development of, and experimentation with, alternate versions of the management functionality applicable in adaptive networks environments. One such approach will be the joint handling and solution of the overall problem, i.e. the solution will include a proper configuration of resources in terms of spectrum management.
Impacts of the reconfigurable terminals being able to support multiple air interfaces simultaneously will be further investigated for this study.

In conclusion, with the ceaseless advance of wireless communications and the gradual introduction of adaptive networks, more flexible network architecture can be achieved and programmable network management can be carried out. In the future, network management functions should not only consider the features and capabilities of the actual network segments, but should also include traffic demand, resource and traffic scalability, as well as the cooperation between previously competitors to efficiently allocate the overall available resources. It is anticipated that, due to such approach, system performance can be significantly improved. This in turn will help to reduce the deployment and operational cost of networks.

ACKNOWLEDGEMENTS

This work has been performed in the framework of the EU funded projects E 2R and E 2R II. The authors would like to acknowledge the contributions of their colleagues from E 2R and E 2R II consortia.

REFERENCES

16. 3GPP TR 25.858 V1.0.4, January 2002.
AUTHORS' BIOGRAPHIES

**Konstantinos Tsagkaris** was born in Lamia, Greece, in 1976. He received his diploma (in 2000) and his PhD degree (in 2004) from the School of Electrical Engineering and Computer Science of the National Technical University of Athens (NTUA). His PhD thesis was awarded in 2005 ‘Ericsson’s awards of excellence in Telecommunications’. He has been involved in many international and national research projects, especially working on the area of wireless networks resource management and optimization. He is now involved with the FP6 End-to-End Reconfigurability (E²R) project. Since January 2004 he is working as a senior research engineer at the Department of Technology Education and Digital Systems of the University of Piraeus. From September 2005 he is an adjunct Lecturer of the Department of Technology Education and Digital Systems of the University of Piraeus. His current interests are in the design and management of wireless reconfigurable networks, optimization algorithms, learning techniques, software engineering, context-aware and pervasive computing. Dr Tsagkaris is a member of the IEEE and the Technical Chamber of Greece.

**George Dimitrakopoulos** received his degree in electrical and computer engineering from National Technical University of Athens in 2002. He is currently working as a Senior Research Engineer with the Telecommunications Networks and Services Laboratory of the department of Digital Systems of University of Piraeus, where he is also working towards finalizing his PhD. Currently, he is involved with the FP6 End-to-End Reconfigurability (E²R) project. His research interests include the performance evaluation, modeling and optimization of B3G communication networks, autonomic networking theory and reconfigurable systems. He is a member of the IEEE and the Technical Chamber of Greece.

**Aggelos Saatsakis** was born in Athens, Greece, in 1982. He has received the Diploma degree in Technology Education and Digital Systems from the University of Piraeus. Currently he is a research engineer at the University of Piraeus, pursuing an MSc degree in the area of networked information systems, and working on broadband networks, and B3G reconfigurable-radio systems.

**Panagiotis Demestichas** was born in Athens, Greece, in 1967. He received the Diploma and the PhD degrees in Electrical and Computer Engineering from the National Technical University of Athens (NTUA). From September 2002 he is an Assistant Professor at the University of Piraeus Research Center, in the department of Technology Education and Digital Systems. From 1993 to 2002 he has been with the Telecommunications Laboratory in NTUA. From February 2001 until August 2002 he was a lecturer at NTUA, in the department of Applied Mathematics and Physics. From September 2000 until August 2002 he taught telecommunication courses, in the department of Electronics of the Technological Education Institute of Piraeus. At the international level he has actively participated in the projects IST.
MONASIDRE, where he was the project manager, CREDO, SHUFFLE, MOEBIUS. He has also participated in projects of the ACTS, RACE II, BRITE/EURAM and EURET frameworks. He has extensive collaborations with Greek companies of the IT and telecommunications sectors. He is now involved with the FP6 End-to-End Reconfigurability (E²R) project. His research interests include the design, management and performance evaluation of mobile and broadband networks, service and software engineering, algorithms and complexity theory, and queuing theory. He is a member of the IEEE, ACM and the Technical Chamber of Greece.