Modelling User Preferences and Configuring Services in B3G Devices

Konstantinos P. Demestichas^{*}, Artemis A. Koutsorodi^{*}, Evgenia F. Adamopoulou^{*}, Michael E. Theologou^{*}, *Member, IEEE*

^{*}Institute of Communication and Computer Systems School of Electrical and Computer Engineering, National Technical University of Athens e-mail: artemis.koutsorodi@gmail.com, eadam@cn.ntua.gr, cdemest@cn.ntua.gr, theolog@cs.ntua.gr

Abstract— This paper discusses a management architecture for devices operating in heterogeneous environments, that enables access network selection through terminal-controlled, preference-based mechanisms. In this domain two problems are identified, mathematically formulated and solved: Intelligent Access Selection (IAS) and Modelling and Adaptation to User Preferences (MAUP). Their objective is to compute the optimal allocation of services to access networks and quality levels, and to dynamically determine user preferences according to the usage context, respectively. A greedy algorithm is proposed for the IAS problem, while the MAUP problem is handled through the construction of a Bayesian network that allows inference and learning of profile and usage patterns. Extensive simulation results of the proposed methods and algorithms are also presented.

Keywords— Bayesian networks, user profiling, preference modelling, access selection, heterogeneous networks

1. Introduction

The evolution of wireless communication systems over the past years demonstrates a clear trend towards architectures that will support multiple access technologies and types of terminals, capable of alternatively operating in the diverse radio segments available in the infrastructure ([1],[2]). These access technologies, Wireless Local Area Networks (WLANs), third-generation cellular systems and broadcast networks, such as Digital Video Broadcasting (DVB), will thus constitute cooperating components of a composite radio infrastructure, interconnected by a

backbone (e.g. an IP-based core network) and jointly operated in an optimized fashion that will allow for an improved overall resource management. This trend is often referred to as 'systems beyond 3G' or '4G wireless systems' and continues to attract notable research attention ([3],[4]).

A central issue within this area, is the design of suitable management frameworks, both for the terminal and for the network, which will allow for seamless service provision, independently of the network the user is attached to. Another important aspect is the fact that the user will need to control the usage of the available networks, especially when this usage comes with a price. This involves a potentially complex decision-making process, which may be guided by policy management tools, with support from both the user devices and the networks, thus giving rise to the cardinal issue of optimally distributing intelligence between the network and the mobile terminal in order to support seamless mobility and service provisioning ([5],[6],[7]).

Another challenge emerging communication systems are faced with, involves the ability to trace and exploit a broad collection of data that characterize the so-called usage context. By context, we refer to any information related to user tasks and experience (including surrounding conditions, user preferences and profile, historical data and other) that being available to the system for further processing can assist it in fulfilling its goals ([8]). User mobility creates situations where the user's context is highly dynamic: interaction, execution and usage needs change frequently and rapidly. This implies that 4G management frameworks must also provide a way for services and terminal settings to adapt appropriately to this constantly changing environment ([9]).

Moreover, one of the main criteria that influence the terminal's access selection process, in an environment with multiple access technologies available, involves the user's preferences and profile, which may also be dependent on the usage context. For example, a user on a business trip may not consider cost to be an important issue in comparison to quality of service, but the same user may want to minimize cost in a leisure time situation ([10],[11]), thus influencing in

different ways the process of selecting the most appropriate access network. In this vein, user profiles related to user requirements for reconfigurable radio systems are introduced in [12].

Hence, the goal in managing 4G terminals proves to be twofold: not only does it lie in the joint utilization of all the available technologies for the provision of higher data rates, but also in the exploitation of a variety of contextual data for the development of personalized services. Accordingly, in the context of effectively managing wireless terminal operation in composite radio environments, two problems are identified:

- Intelligent Access Selection (IAS), which involves the computation of optimal allocations of services to access networks and different levels of Quality of Service (QoS), given network availability and user preferences, and
- *Modelling and Adaptation to User Preferences* (MAUP), which involves, firstly, the processing and the 'decoding' of contextual information for the purpose of determining the usage context, and, secondly, the automatic determination of the user's preferences, according to the usage context and the set of services under investigation.

This paper includes the formal definition and a potential solution to these problems. It extends the work presented in [13],[14], by unifying the approach to these two problems, and by providing a consolidated implementation approach for an appropriate terminal management framework. Their combined solution enables user terminals to be driven to allocations of services to access networks and quality levels which satisfy user preferences in any given usage context. We argue that appropriate functionality for handling the IAS and the MAUP problems must be in place at the mobile terminal, for enabling it to handle basic mobility management tasks, to support the applications in dealing with the dynamics and heterogeneity of the available access networks and to dynamically adapt to changing user preferences. The proposed approach is generic with respect to the wireless technologies included in the composite radio infrastructure, and also generic with respect to the modelling of the context space. This work can be the basis for

further research in the area of developing efficient algorithms for service configuration, as well as in the area of preference modelling and dynamic user profiling.

The paper is structured as follows: Section 2 discusses the requirements and some basic design concepts for a terminal management architecture, Sections 3 and 4 provide the high-level and formal descriptions of the IAS and MAUP problems as well as possible solution approaches. Section 5 discusses implementation issues and presents extensive simulation results, which provide valuable insight into the proposed solution approach and algorithms. Finally, Section 6 concludes the paper.

2. Requirements and design

2.1. Requirements

It is clear that the most salient feature of 4G systems will be the multiplicity of access technologies and terminals, which will allow users to benefit from the realization of the *always best connected* concept: being connected at any point in time to the best available access network. The latter is defined as the network that best suits the user's needs at that particular moment and may be a function of the user's personal preferences, of the different operator policies and business agreements and of the network resources available in the infrastructure ([2]).

Besides, the enhanced capabilities of evolving terminals have introduced the idea of extending service intelligence from the network to the terminal. As modern devices constitute small, but increasingly powerful operating systems that may also be equipped with multiple network interfaces to enable network connectivity using different types of access networks, the trend of assigning more and more complex tasks to the user terminal is gaining ground ([5],[6],[7]).

The efficient exploitation of the composite radio infrastructure calls for the development of an appropriate, sophisticated Terminal Management System (TMS) that should be capable of detecting available access networks and processing measurements regarding the capabilities of

available access points (e.g. supported technology, network operator, cost per service and per quality level, etc), handling the user profile and incorporating a mechanism for adapting automatically to the user's preferences according to the identified usage context, and, finally, optimally selecting allocations of services to access networks and to quality levels (let it be noted that by 'access point' we shall refer to any available network attachment point available in the infrastructure). The TMS must be modelled for a modern, general-purpose operating system, running on a mobile terminal that may be continuously switched on. Power management and power conservation issues are out of scope.

2.2. Design

Figure 1 illustrates the main components of the TMS architecture, namely the Network Interface Adaptation Module (NIAM), the Mobility Management Module (MMM), and the User Preferences Module (UPM).

(Figure 1)

The NIAM is responsible for providing the terminal with a level of abstraction from the different network drivers, the UPM for accessing and processing the user profile, and the MMM for handling mobility events, for performing service configuration and for optimally computing handover targets. The following paragraphs discuss these modules in more detail.

2.2.1. Adaptation to network drivers

The NIAM serves two purposes: (a) the connection and disconnection of the appropriate interface during power-up of the mobile terminal or during a handover and (b) the retrieval of layer-2 measurements in the network interface. This means that the NIAM is able to provide the terminal with measurements retrieved from the different network drivers, reflecting the signal quality or connectivity status in each of these interfaces in an abstracted way (e.g. good, average or poor signal strength). In all, the NIAM is able to provide the terminal with a list for each

access point in the terminal neighbourhood, each list comprising information about the access point signal strength and bandwidth availability, type of technology, network operator and about the cost at which it is able to provide each available service at each of the service's permissible quality levels.

Therefore, the NIAM should incorporate a mechanism for retrieving such measurements from access points in its range (such as the Candidate Access Router Discovery protocol [15]) and for processing this information in order to detect if a new access point has appeared in the terminal's neighbourhood, if an old one has disappeared, or if the perceived signal strength from an already selected access point has severely deteriorated. In all of these cases, the MMM is notified accordingly for the purpose of triggering the process of selecting handover targets for all running applications.

2.2.2. Mobility management

The MMM depicted in Figure 1 incorporates functionality for managing mobility both at the link layer (computation of optimal allocations to access points and quality levels, control of connection establishment and release at the network interface level) and at higher levels (IP layer handoff transparency, application control). The first aspect of mobility management is provided through the use of an algorithm that solves the IAS problem, while the second aspect may be provided through the incorporation of appropriate protocol solutions, such as Mobile IP ([16],[17]) and SIP ([18]). That is, the assumption is made that the terminal implements the multihoming concept, the ability to simultaneously use multiple network interfaces, using different network addresses over each interface ([19]). These functions are presented in Figure 2.

(Figure 2)

Mobility management and service configuration in 4G architectures has been addressed in a number of papers ([4],[20]). In most approaches, the computational load that involves the terminal vertical handovers is assigned to the network central management unit, entailing

potential overload and delay problems, while the applications running on each terminal are not individually considered as the objects of a handover, but are rather handled collectively.

2.2.3. User profiling

The UPM is responsible for storing, accessing and editing the user's profile, and for addressing the MAUP problem. A system that considers usage context and service requirements shall provide appropriate prioritization of the parameters that may influence the access network selection process. This prioritization is equivalent to the specification of values for the different coefficients w_q , w_o , w_t and w_c , which correspond to parameters 'quality', preferred 'network operator', preferred 'technology type' and 'cost' respectively, and represent the measure by which each one of these parameters is weighted in the access network selection algorithm. For example, if the user chooses to specify that, at a given moment, 'quality' is for him the most important factor in access network selection, 'technology type' comes second, 'cost' comes third and last comes 'network operator', then the respective coefficients will be assigned values $w_q > w_t > w_c > w_o$.

In this paper, we propose an approach to the MAUP problem based on a system that models user preferences and the causal relationships between them using Bayesian networks – directed acyclic graphs with network structures that encode conditional independence assertions about a set of variables ([21],[22]). The goal of such a system is the prediction of the user's preferences. This function presupposes the development of appropriate mechanisms through which:

- The usage context will be inferred, given contextual information (user location, time and date, application type being requested, terminal type being used, content type being accessed e.g. web page with business data, phone call to a family member etc.).
- The user preferences will be inferred, given usage context and service requests.

These functions are depicted in Figure 3.

(Figure 3)

Bayesian networks have proven to be valuable tools for encoding, learning and reasoning about probabilistic relationships, and once they have 'learned' the correct structure and parameters, they may support probabilistic inference of the user's preferences ([10]).

3. Intelligent Access Selection

In the following subsections, the IAS problem is described in high-level and formal manner and a computationally efficient solution is proposed.

3.1. Problem description

The input to the IAS problem falls into three categories:

- A set of measurements for each available access point in the terminal neighbourhood: (i) its availability in bandwidth, (ii) the signal strength perceived at the terminal from that point, (iii) the type of the technology that point supports, (iv) the network operator it belongs to, (v) its identifier, as well as the identifier of the terminal network interface which communicates with that point.
- The set of applications already running on the mobile terminal, the corresponding quality levels at which these applications are being provided as well as the set of applications that the user is requesting to use.
- The set of user preferences, according to which the parameters 'quality', 'network operator', 'technology type' and 'cost' are prioritized.

The solution of the IAS problem is the result of an optimization process that should compute allocations of services belonging to the two aforementioned sets to quality levels and to access points. These allocations should optimize an objective function associated with the quality level and cost at which each service is provided, with the type of technology and the network operator through which this provision is effectuated, and with the importance the user allocates each of these factors, at a given time.

The constraints that the computed allocations have to satisfy fall in the following categories:

- Each service must be allocated to a permissible quality level.
- The maximum quality level that the user defines for each service must not be violated.
- The user's preferences must be taken into consideration in the order of priority that is specified in his profile.

The algorithm for solving the IAS should be triggered in the following cases:

- Whenever the user issues a service request.
- Whenever the user modifies his profile.
- Whenever the received signal strength for any access point the terminal is connected to, falls below a threshold (for a certain time window).
- Whenever a new available access point 'appears' in the terminal neighbourhood, or an already available one 'disappears'.

Let $S = \{s_1, s_2, ..., s_k\}, k \in \square$ be the set of both running and requested services at the mobile terminal, let $P = \{p_1, p_2, ..., p_n\}, n \in \square$ be the set of access points that the terminal perceives and let $Q_s(p) = \{q_1, q_2, ..., q_m\}, m \in \square$ be the set of quality levels at which access point $p, p \in P$, may offer service $s, s \in S$.

We aim to compute allocations of services to quality levels $(A_{S-QL} = \{ql(s) \mid \forall s \in S\})$ and to access points $(A_{S-AP} = \{ap(s) \mid \forall s \in S\})$ that optimize the objective function

$$OF(A_{S-QL}, A_{S-AP}) = \sum_{s \in S} x_{s}(q) \cdot y_{s}(p) \cdot \left\{ F(p,q,s) + \sum_{s' \neq s} F'(p,q,s') \right\}$$
(1)

where $F(p,q,s) = w_q \cdot F_Q(p,s) + w_t \cdot F_T(p) + w_o \cdot F_O(p) - w_c \cdot F_C(p,q,s)$,

 $F'(p,q,s') = w_q \cdot F'_{\mathcal{Q}}(p,s') + w_t \cdot F_T(p) + w_o \cdot F_O(p) - w_c \cdot F_C(p,q,s') \text{ and}$

 $F' \varrho(p,s') = f(BW_{Av}(p) - BW(s,q))$, subject to:

 $\sum_{q \in \mathcal{Q}(p)} x_s(q) = 1, \ \forall s \in S$ (2)

$$\sum_{p \in P} y_s(p) = 1, \forall s \in S$$
(3)

$$ql(s) \le ql_{s, \max}, \forall s \in S$$
(4)
and $OF(\emptyset, P) = 0.$

Relation (1) expresses the objective of finding allocations A_{s-QL} and A_{s-AP} that optimize an objective function, associated with the satisfaction of user preferences, in the order that these are specified in his profile. Variables $x_s(q)$ and $y_s(p)$ take the value 1 in case service s is assigned to quality level q and to AP p, correspondingly.

Coefficients w_q , w_o , w_t , w_c correspond to the user's preference for (high) quality, a specific network operator, a specific type of technology and (low) cost respectively.

Factor $F_Q(p,s)$ expresses the quality level at which AP p offers service s, and is a function both of the available bandwidth of p ($BW_{Av}(p)$), and also of the perceived signal strength from point p (SS(p)):

$$F_{\mathcal{Q}}(p,s) = q \cdot \sigma \tag{5}$$

where q is the maximum quality level supported by $BW_{A\nu}(p)$, and σ is a multiplication factor depending on SS(p). An approach for the possible values of σ is given in Table I of Section 5.1.

Factors $F_T(p)$ and $F_O(p)$ correspond to the network operator and the type of technology that characterize point p. In case the candidate AP p belongs to the preferred operator and/or supports the preferred technology, it is granted a 'bonus', whose value is proportional to the difference $F_Q(p,s) - F_C(p,q,s)$. If not, the value of this 'bonus' is zero. Symbol b_f in relation (6) represents the proportionality factor and may take values between 0 and 1.

$$F_{O}(p), F_{T}(p) = \begin{cases} b_{f} \cdot [F_{Q}(p,q) - F_{C}(p,q,s)], & \text{if } p \text{ matches preference} \\ 0, & \text{if } p \text{ does not match preference, or if } F_{Q}(p,q) < F_{C}(p,q,s) \end{cases}$$
(6)

Factor Fc(p,q,s) expresses a normalized indication of the cost at which AP p offers service s at quality level q, and is computed as follows: Let c(p,q) be the nominal cost (per unit of time or volume) for a service offered by an AP p at quality level q. Let c_{max} be the maximum cost for this service (among all available APs and quality levels) and $n_{q,lev}$ the number of the different quality levels at which this service can be offered. We define:

$$F_{c}(p,q,s) = c(p,q) \cdot \frac{n_{q.lev.}}{c_{max}} \quad (7)$$

Factor F'(p,q,s') expresses the fact that every allocation of another service s' must take under consideration the reduced (by BW(s,q), which corresponds to the bandwidth allocated to service s at quality level q) bandwidth that characterizes factor $F'\varrho(p,s')$.

Constraints (2) and (3) guarantee that every service s is allocated to one quality level q and to one AP p. Finally, relation (4) ensures that the user's constraint regarding the maximum allowed quality level for each service will be respected.

3.2. Problem solution

Since the determination of the optimal problem solution requires a significant amount of computations, the design of an efficient algorithm that may provide near-optimal solutions in reasonable time is necessary. Several approaches towards this issue have been studied ([23]). In this paper, we rely on the greedy algorithm paradigm ([24]) for the development of a faster approach.

The algorithm runs in three main phases:

- Input Retrieval and Initial Processing, where the complete list P of access points is constructed and the user's preferences are automatically resolved.
- Computation of Service Allocations, where all possible permutations of set S are generated and evaluated.

 Selection of Aggregate Solution, where the permutation that corresponds to the maximum aggregate 'grade' is determined.

The complexity of this algorithm is $O(n \times \ell \times ((m+1)!))$, in the worst case, where *m* is the number of services that are running or have been requested at the mobile terminal (size of set *S*), *n* is the number of access points in the terminal's neighborhood (size of set *P*), and ℓ is the overall maximum number of quality levels, regardless of the service and access point in question. Although this complexity may seem significant, due to the presence of the factorial (m+1)!, it should be noted that, on one hand, in small portable devices, such as mobile phones or even smart-phones, there are usually no more than 3 or 4 services running simultaneously, while, on the other hand, more advanced devices, such as notebooks, can easily handle the computational load.

4. Modelling and Adaptation to User Preferences

In the following subsections, the MAUP problem is described in high-level and formal manner and a solution is proposed.

4.1. Problem description

We propose the construction of a two-level Bayesian meta-network (based on an approach presented in [10]) as depicted in Figure 4, where the division of the MAUP problem in – initially – two separate stages is apparent: first, inference of usage context and, second, inference of user profile parameters.

(Figure 4)

This structure assumes that probability distributions associated with nodes on the predictive level depend on probability distributions associated with nodes on the contextual level.

The MAUP solution proposed herein modifies and extends the approach introduced in [10] according to the following points: *(i)* It employs a modified predictive level, which is better-suited for a generic access selection optimization process; *(ii)* An actual terminal-side Java-based implementation is provided (which is crucial for validating the feasibility of such a solution on resource-constrained mobile terminals); *(iii)* A probabilistic neural network is used for learning the parameters of the network. This approach fully automates the learning process – only some initial training input is required – since no user confirmation / intervention is needed thereafter.

The following subsections discuss the three separate stages in which the MAUP problem is dealt with.

4.1.1. Stage 1: Inference of usage context

The Bayesian network of the second level (contextual level - Figure 4) consists of four nodes: one for the user's location (L), one for the time of day (T), one for the session's communicating counterpart (CP) (predictive attributes), and one for the usage context (C) (target attribute). Each node represents a random variable. The knowledge of the conditional probability distribution for all four variables and the application of basic Bayesian inference rules lead to the determination of the most probable value of variable C, given the values of variables T, L and CP.

Let t, l, and cp be the known values of random variables T, L and CP. We seek to determine the value c of random variable C for which the following probability is maximized: P(C = c | T = t, L = l, CP = cp).

If we assume that random variable T has n_{time} mutually exclusive and discrete states, variable L has $n_{loc.}$ and variable CP has $n_{c/p}$ discrete states, then a 4-dimensional table is required, where for every context state C_i the value of the aforementioned conditional probability will be given for every one of the $n_{time} \cdot n_{loc.} \cdot n_{c/p}$ possible combinations of the other variables.

4.1.2. Stage 2: Inference of profile parameters

The Bayesian network of the first level (predictive level - Figure 4) consists of five nodes: one for service and four more, one for each of the coefficients w_q , w_o , w_t and w_c .

The value of variable C, derived following reasoning on the second level of the Bayesian meta-network, influences the way in which the values of w_q , w_o , w_t , w_c depend on the value of variable S. Knowledge of the respective conditional probability distributions and the application of basic Bayesian inference rules lead to the determination of the most probable values of w_i , given the values of variables C and S.

Let *s* and *c* be the known values of variables *S* and *C*. We seek to determine the values of coefficients w_q , w_o , w_t kay w_c – let *a*, *b*, *c*, *d* be these values – for which the following probability is maximized: $P(w_q = a, w_o = b, w_t = c, w_c = d | S = s)$

It holds that: $P(w_q, w_o, w_t, w_c | S) = \frac{P(w_q, w_o, w_t, w_c, S)}{P(S)}$, and according to the independence

relationships encoded in the structure of Figure 4 (predictive level):

$$P(w_q, w_o, w_t, w_c, S) = P(S) \cdot P(w_q \mid S) \cdot P(w_o \mid S) \cdot P(w_t \mid S) \cdot P(w_c \mid S)$$

Hence: $P(w_q = a, w_o = b, w_t = c, w_c = d \mid S = s) = P(w_q = a \mid S = s) \cdot P(w_o = b \mid S = s) \cdot P(w_t = c \mid S = s) \cdot P(w_c = d \mid S = s)$

and the computation of $P(w_q, w_o, w_t, w_c | S)$ reduces to the computation of the product of partial conditional probabilities, which result from their respective conditional probability tables.

4.1.3. Stage 3: Learning and adaptation

The values of w_q , w_o , w_t and w_c that are computed, following reasoning on the first level of the Bayesian meta-network, are used in turn as input to the IAS problem. The decision that is reached based on this input may, or may not satisfy the user's needs, given the fact that his preferences were automatically defined, without his interference. Therefore, the ultimate goal is the training of the proposed network, in order to render it capable of accurately and independently predicting the user preferences. This implies the development of an appropriate algorithm for the adjustment of the CPTs, according to the system's initial success or failure in predicting the values of w_q , w_o , w_t and w_c .

This process involves the use of an appropriate training set, provided to the system by the user and comprising examples of his priorities in some situations: the user is given some configurations of his profile parameters, for different combinations of service and context, and answers whether each configuration satisfies him ('yes' answer) or not ('no' answer). Our goal, therefore, is to extract knowledge about user preferences in any context, given his preferences in some instances of context. An underlying assumption we make, throughout our discussion of the MAUP problem, is the fact that the user's behaviour and preferences are not randomly decided upon, but rather comply with some implicit logic, which we are trying to approximate.

In general, learning Bayesian networks (BNs) from data involves two aspects: learning network structure and learning network parameters. The problem of identifying a Bayesian network (among those where each node has at most a given number of parents) that has a relative posterior probability greater than a given constant is NP-complete ([25]). In this paper, we make a reasonable assumption about the structure of the Bayesian network that best describes our model, and seek efficient algorithms for learning network parameters.

4.2. Problem solution

The training, or 'learning', of the Bayesian network can be achieved using a variety of different methods, such as linear regression, probabilistic neural networks and probabilistic decision trees ([26],[27],[28]). Most approaches to learning Bayesian networks involve exact data. In our case, the training data that is provided to the system by the user only specifies a 'region' of positive or negative preference. Therefore, we need a mechanism for identifying the similarity of a new

configuration of the profile's parameters to these regions and for appropriately upgrading the respective conditional dependencies, according to this degree of similarity.

We propose the construction of a probabilistic neural network that uses the training set as a reference for the appropriate adjustment of the values stored in the Bayesian network's CPTs. This process is depicted in Figure 5.

(Figure 5)

The PNN receives as input a new allocation element computed by the system: given C = c and S = s and for $w_q = a$, $w_o = b$, $w_t = c$, $w_c = d$ designated by the solution of the Bayesian network, it performs the following tasks:

- It assesses the similarity of the new instance to the patterns stored in the training set.
- It produces a probability vector, comprising classification probabilities to each class:
 k₁ = 'accept', k₂ = 'reject'.
- It selects the maximum of these probabilities, using the Bayes-optimal decision rule.

The PNN's output is subsequently used for the reinforcement or the diminishment of the conditional probabilities $P(w_q | S)$, $P(w_o | S)$, $P(w_t | S)$ and $P(w_c | S)$ in the corresponding CPTs (by simple application of Bayes' theorem), in the case of classification of the system's decision into class k_1 or k_2 respectively.

5. Implementation and Evaluation

5.1. Implementation aspects and simulation parameters

Two Java-based implementations of the proposed TMS have been realized: the first one relies on the Java 2 Standard Edition (J2SE) v1.2 platform, while the second one is based upon the Java 2 Micro Edition (J2ME) platform. The J2SE implementation is suitable for laptops, while the J2ME implementation is targeted for mobile phones and (entry-level or high-level) PDAs that support the Connected Limited Device Configuration (CLDC) v1.1 and the Mobile Information Device Profile (MIDP) v2.0 (these specifications are supported by the majority of mobile terminals available in the market).

Simulations are executed in the framework of customizable scenarios. These scenarios are developed using a custom-built Java tool, which features an easy-to-use GUI. Through this GUI, one can specify, save, load and edit the parameters of a scenario (i.e., the access points in the terminal's range and their characteristics, the services that will run on the terminal, context-related parameters, the training set that will be used, etc.). The scenario's parameters are stored in XML files, which are then provided as input to the TMS, in order to test and evaluate its functionality.

Cost-related parameters are also stored in an XML file and serve as input when simulating a scenario. Cost is expressed in monetary units per time or data volume unit (i.e., per sec or KB, respectively). Cost values depend on the service type, the quality level, the network operator and the type of technology. A qualitative approach has been employed for the construction of the aforementioned XML file. Although the cost values have been partially randomized, the following guidelines have been followed: *(i)* Higher quality levels are more expensive; *(ii)* WLAN is normally more cost-effective compared to the other technologies; *(iii)* GSM is less expensive for voice calls than UMTS; *(iv)* DVB is very cost-effective for video-related services (e.g., video streaming). A sample of the employed cost file is given below.

```
<costs>

<service code = "1" maxCost="0.032">

<QoS level = "1">

<provider code = "1">

<technology code = "1">0.003</technology>

<technology code = "2">0.005</technology>

<technology code = "3">0.002</technology>

<technology code = "3">0.002</technology>

<technology code = "4">0.006</technology>

</provider>

<technology code = "2">

<technology code = "1">0.0032</technology>

<technology code = "2">

<technology code = "3">0.0016</technology>

<technology code = "4"><technology </provider>
```

</costs>

Table I summarizes the values of the multiplication factor σ of relation (5) that are used throughout the simulation of the present section. The proportionality factor b_f of relation (6) has been set equal to 0.5.

SS(p)	σ
1	0.40
2	0.60
3	0.80
4	0.95
5	1.00

Table I. Evaluation of the received signal's strength

In the simulation that will follow, we consider an ordinary user X and validate the system's functionality through representative scenarios. We consider two usage contexts (business and leisure) and four services (voice call, video call, video streaming and web browsing). For each service, there are five permissible quality levels (Q1 through Q5), the bandwidth requirements of which are given in Table II.

Service	Q1	Q2	Q3	Q4	Q5
Voice Call	4	8	16	20	25
Video Call	32	64	128	256	384
Video Streaming	64	128	512	1024	2048
Web Browsing	32	64	128	384	1024

Table II - Bandwidth requirements (in Kbps) for scenario services

A high-level description of user X's preferences is given in Table III. The user has also specified 'Op1' as his preferred operator, and WLAN as his preferred technology for video streaming and voice call.

	Leisure context	Business context
General Preferences	cost is more important than quality	quality is more important than cost

Table III - User preferences (implicit)

Voice Call, Video Call	-	_
Video Streaming	technology is more important than operator	operator is more important than technology
Web Browsing	technology is more important than operator	operator is more important than technology

On the basis of these preferences, the user provides 'accept' or 'reject' answers to sample configurations of the factors quality, cost, operator and technology (Table IV). More specifically, the user answers one pair of questions for each available service. The first question (configuration) of each pair is randomly generated, while the second one is appropriately selected so as to significantly differ from the first one. In Table IV, 'Q' stands for quality, 'C' for cost, 'O' for network operator and 'T' for technology. Coefficients w_q , w_o , w_t and w_c are assigned values 0.8, 0.6, 0.4 and 0.2, according to their prioritization as first, second, third and fourth. To avoid confusion, a factor's priority level is denoted as w_t ', where $i \in \{q, c, o, t\}$, while a factor's actual value is denoted as w_i . Hence, $w_i' \in \{1, 2, 3, 4\}$, while $w_i \in \{0.8, 0.6, 0.4, 0.2\}$. For example, if 'quality' is regarded as the most important factor, then $w_q'=1$ and $w_q=0.8$.

Contort	Somioo		Acc	ept		Reject				
Context	Service	Q	С	0	Т	Q	С	0	Т	
	Voice Call	4	1	3	2	1	4	2	3	
Lainuna	Video Call	4	1	2	3	2	4	1	3	
Leisure	Video Streaming	4	1	2	3	2	4	1	3	
	Web Browsing	4	2	3	1	1	3	4	2	
	Voice Call	1	4	2	3	4	1	2	3	
Dusinasa	Video Call	2	4	1	3	3	1	4	2	
Business	Video Streaming	2	4	1	3	3	1	4	2	
	Web Browsing	3	4	1	2	4	1	2	3	

Table IV - User answers to randomly selected questions

These answers of Table IV constitute the training set, based on which the system is trained to accurately and independently predict the user's preferences. Let it be noted that we assume X to

be an average user, without any specific knowledge of the factors he is trying to prioritize. Thus, it is possible that X answers a question incorrectly (with respect to the underlying preferences initially considered), or gives contradictory answers. The system tries to smooth out these contradictions and determine what the user actually has in mind, and is still able to predict his preferences, even if he accepts or rejects both of the proposed configurations, or even if he provides answers for less than two questions per service and per context.

5.2. User preferences prediction

In this subsection, a more detailed explanation of the user preferences prediction process is given, along with an indicative, enlightening example. Based on the answers of Table IV, the system must be trained to predict the user's preferences. Initially, all conditional probability values are uniformly distributed, i.e.: Given context C = c and service S = s,

$$P(w_i'=1 \mid S=s) = P(w_i'=2 \mid S=s) = P(w_i'=3 \mid S=s) = P(w_i'=4 \mid S=s), \forall i \in \{q, c, o, t\}$$

This is depicted in Table V(a), which represents the initial CPT for leisure context (the analysis hereafter is equivalent for business context). When the user submits his answers, an initial training of the CPT takes place. During this initial training process, each node of the training set is given as input to the PNN, following the procedure described in Section 4.2. This results in some of the values of the CPT being reinforced and some being diminished. When the initial training is over, the CPT has the form of Table V(b).

								LEIS	URE							
(a)	Voice Call			Video Call			Video Stream			Web Browsing						
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Quality	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
Cost	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
Operator	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
Technology	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25

Table V - (a) Initial CPT; (b) CPT after initial training (leisure context)

		LEISURE											
(b)	Voice Call Video Call Web Browsin					owsing							
	1	2	3	4	1	2	3	4		1	2	3	4

Quality	0.012	0.306	0.306	0.374	0.295	0.048	0.295	0.360	 0.071	0.288	0.288	0.352
Cost	0.374	0.306	0.306	0.012	0.360	0.295	0.295	0.048	 0.288	0.352	0.071	0.288
Operator	0.306	0.012	0.374	0.306	0.048	0.360	0.295	0.295	 0.288	0.288	0.352	0.071
Technology	0.306	0.374	0.012	0.306	0.312	0.312	0.063	0.312	 0.352	0.071	0.288	0.288

Let us now assume that we need to predict the user's preferences (priority levels w_q ', w_c ', w_o ', and w_t ') for the video call service and for leisure context. The TMS performs the following tasks:

First, it searches through Table V(b), in order to detect the most probable values for the set $\{w_q', w_c', w_o', w_t'\}$. In this example, $w_q'=3$, $w_c'=1$, $w_o'=2$, and $w_t'=4$, i.e. $\{3, 1, 2, 4\}$, seems to be the most probable prioritization.

Subsequently, this instance ('candidate prioritization') is given as input to the PNN, as described in Section 4.2. The TMS calculates the distance (degree of similarity) of this instance from the nodes of the training set (i.e., Table IV). The 'degree of similarity' metric takes values between 0 and 1, with '1' indicating perfect match, and '0' reflecting no match. For example:

The degree of similarity of the candidate instance compared to node {Leisure, Video Call, {4,

1, 2, 3}, Accept} is
$$1 - \frac{\sqrt{(4-3)^2 + (1-1)^2 + (2-2)^2 + (3-4)^2}}{\sqrt{20}} = 0.6838$$
, where $\sqrt{20}$ is used

for normalization, since it is the maximum possible distance. Similarly, the degree of similarity of the candidate instance compared to node {Leisure, Video Call, {2, 4, 1, 3}, Reject} is

$$1 - \frac{\sqrt{(2-3)^2 + (4-1)^2 + (1-2)^2 + (3-4)^2}}{\sqrt{20}} = 0.2254.$$
 However, the degree of similarity

compared to node {Business, Video Call, {2, 4, 1, 3}, Accept} is 0 (due to different context). Also, the degree of similarity compared to node {Leisure, Voice Call, {4, 1, 3, 2}, Accept} is

$$1 - 0.5 \times \frac{\sqrt{(2-3)^2 + (4-1)^2 + (1-2)^2 + (3-4)^2}}{\sqrt{20}} = 0.4523 \text{ (same context, different service), i.e.}$$

in this case we multiply the normalized distance by a factor of 0.5. The value of this factor is a design choice. Different implementations may select different values (between 0 and 1).

In the next step, degrees of similarity that correspond to the same class ($k_1 = 'accept'$, $k_2 = 'reject'$) are aggregated (summed). The aggregate degrees are then normalized (so that we obtain the instance's acceptance-probability and rejection-probability) and compared. For the candidate instance under consideration, we obtain the following results: Acceptance Probability = 0.7423, Rejection Probability = 0.2572. Thus, in this case, the candidate instance represents a valid set of user preferences. From the system's viewpoint, these are the actual user preferences for the given context and service. This result is cached (stored) by the TMS for future use. In case a candidate instance is classified as rejected, the system weakens the corresponding conditional probabilities by a factor that is proportional to the rejection probability, and repeats the entire process until it reaches an acceptable instance.

Since the system takes into account more than one patterns when evaluating a candidate instance, it manages to handle contradictory answers rather efficiently. In particular, when the user correctly answers the majority of the generated questions, then the contribution (weight) of the correct answers in the evaluation process prevails against the contribution of the contradictory answers.

Table VI summarizes the user's preferences as predicted by the TMS.

Contort	S	Prioritization							
Context	Service	Q	С	0	Т				
Leisure	Voice Call	4	1	3	2				
	Video Call	3	1	2	4				
	Video Streaming	3	1	2	4				
	Web Browsing	4	2	3	1				
	Voice Call	1	4	3	2				
Business	Video Call	2	4	1	3				
	Video Streaming	2	4	1	3				

Table VI - Predicted user preferences

Web Brows	ing 3	4	1	2
-----------	-------	---	---	---

5.3. Scenario results

Four scenarios are simulated in the following. Table VII illustrates network coverage in each scenario. For every available access point, its access technology, network operator, signal strength (SS) and available bandwidth are specified. The indicated values of the access points' available bandwidth are examples of the total bandwidth available (at a specific time) in each access point and depend both on the capabilities of each access technology and on access point traffic load. Three out of the four abovementioned services have been considered: voice call, video streaming and web browsing.

	GSM APs	UMTS APs	WLAN APs	DVB APs
		1. Op1, SS-5, 120Kbps		
Scenario 1		2. Op2, SS-4, 380Kbps		
		3. Op3, SS-4, 760Kbps		
Scenario 2	Op1, SS-5, 44Kbps	Op2, SS-5, 380Kbps	Op3, SS-3, 2100Kbps	
Scenario 3	Op1, SS-5, 44Kbps	Op2, SS-5, 760Kbps	Op3, SS-3, 500Kbps	
Scenario 4	Op1, SS-5, 44Kbps	Op2, SS-5, 760Kbps	Op3, SS-3, 1000Kbps	Op4, SS-4, 2400Kbps

Table VII - Network coverage in each scenario (AP: access point)

(a) Scenario 1

In this scenario, we assume that three UMTS access points are in the terminal's neighbourhood. We first consider a business usage context and simulate the voice call and the video streaming service. Figure 6 depicts detailed results, illustrating the values of each of the factors 'quality', 'cost', 'operator' and 'technology', as well as the overall grade (objective function value), for every potential allocation in the framework of this scenario. The 'cost' factor is represented as a negative quantity, while all the other factors are positive. The overall grade may be either negative or positive, depending on which factor prevails. The allocation that features the highest overall grade is considered as optimum.

As may be observed from Figure 6, for voice call, the UMTS#1 access point is selected, which features better signal strength and belongs to the preferred operator. Quality level 5 is selected,

since the context is 'business'. However, video streaming is allocated to UMTS#3, as this is the only access point that is capable of offering quality level 3, thanks to its relatively higher available bandwidth.

(b) Scenario 2

In this scenario, we assume that one GSM, one UMTS and one WLAN access point exist in the terminal's neighbourhood. All three services are simulated. Figure 7 illustrates the overall grade per service and per potential allocation, for business context. Notice that some access points may not support all possible quality levels, e.g. the GSM access point does not support web browsing at quality level 2 and above (due to bandwidth constraints). Voice call is allocated to the GSM access point, since the latter features higher signal strength and belongs to the preferred network operator. However, the WLAN access point is chosen for video streaming and web browsing, as it is the only available access point that is able to support high quality levels.

(c) Scenario 3

In this scenario, we have the same access points but with different bandwidth characteristics. Concerning the business context (Figure 8(a)), the allocation chosen for the voice call service is the same, in comparison with Scenario 2. Web browsing is allocated to the WLAN access point, at a lower quality level, since the access point no longer supports the maximum quality level, due to limited bandwidth. The UMTS access point is selected for video streaming, since it is able to offer a higher quality level compared to WLAN. Regarding the leisure context (Figure 8(b)), the WLAN access point is selected, since it represents the most cost-effective solution. Quality level 1, i.e. the minimum one, is chosen for video streaming. Quality level 3 is chosen for voice call and web browsing. Referring back to Table VI, we may notice that 'technology' is considered as an important factor for voice call and web browsing. This results in the significant reinforcement of quality levels that offer a better combination of quality and cost (rather than just low cost), as in the case of quality level 3.

(d) Scenario 4

In this scenario, one additional access point, supporting the DVB technology, is available. Table VIII summarizes the allocation decisions that have been reached by the TMS. As may be observed, the appearance of the DVB access point has affected the access selection process. In business context, the video streaming service is allocated to the DVB access point, the available bandwidth of which allows the provision of this service at the maximum possible quality level. In leisure context, DVB is again chosen for video streaming, as it is inherently more cost-effective for video-related services (also, see Section 5.1).

Contoxt	Somulao	Allocation					
Context	Service	Tech.	Oper.	QoS			
	Voice Call	GSM	#1	5			
Business	Video Streaming	DVB	#4	5			
	Web Browsing	WLAN	#3	4			
	Voice Call	WLAN	#3	3			
Leisure	Video Streaming	DVB	#4	1			
	Web Browsing	WLAN	#3	3			

Table VIII - Allocations selected by the TMS in Scenario 4

5.4. Comparison with a rule-based approach

An interesting point involves the comparison of the proposed 'Bayesian' approach for determining user preferences with a simple rule-based system, which encodes user preferences in different contexts as rules. For this purpose, let us assume that the terminal manufacturer has integrated the following rules in the TMS:

- When in 'leisure' context, let $w_c = 0.8$ and let all other factors be weighted equally, i.e. $w_q = w_o = w_t = 0.4$.
- When in 'business' context, let $w_q = 0.8$ and let all other factors be weighted equally, i.e. $w_c = w_o = w_t = 0.4$.

Simulating part of Scenario 4 using the above-mentioned rules would result in the allocations given in Table IX.

Context	Service	Allocation		
		Tech.	Oper.	QoS
Business	Video streaming	DVB	#4	5
Leisure	Voice call	GSM	#2	5
Leisure	Web browsing	WLAN	#3	4

Table IX – Allocations selected by a rule-based system for phase (b)

As can be observed, only the first allocation is the same compared to the 'Bayesian' approach. The other two are different, and apparently do not truly satisfy the user's preferences. Voice call is allocated to a GSM access point and to quality level 5, although a more cost-effective solution is appropriate. The same applies for the allocation of web browsing to a WLAN access point at quality level 4, as a lower quality level could have been chosen.

In general, the proposed approach to the MAUP problem is more 'sensitive' and effective than a simple rule-based solution, without entailing severe computational load (also, see Section 5.5). Furthermore, the use of a rule-based system presents the following significant drawbacks:

- It is not as adaptable and extendible: if rules are integrated in the TMS by the manufacturers it would be impossible for users to edit them according to their own needs and preferences.
- A rule-based approach cannot entirely satisfy the needs of a non-average user; a complex personal profile would require a very complicated set of rules.
- A rule-based approach lacks the ability to smooth out contradictory decisions stemming from false or poorly expressed rules.

5.5. Performance evaluation

The time needed for the algorithm to reach a single allocation decision is depicted in Table X. As can be deduced, the computational load of the J2SE implementation is negligible, and is easily handled by laptop computers. The execution time of the J2ME implementation is, as expected, bigger. Nonetheless, a delay in the order of 1.5 sec can be considered as tolerable. Indeed, by executing the IAS function in a separate process (thread), the user will, most probably, notice no pause.

Development Platform	Hardware	Execution Time	
J2SE v1.2	Pentium IV notebook	50-80 msec	
J2ME (CLDC v1.1, MIDP v2.0)	Nokia 6630 S60, 2 nd Ed.	1200-1600 msec	

Table X - Algorithm's execution time per allocation decision

Since modern mobile phones are equipped with faster processors and larger memory, the execution time of the J2ME implementation is expected to be even smaller on such hardware. Furthermore, our intention is to further fine-tune the system's performance by: *(i)* employing a more efficient and lightweight XML parser; and *(ii)* de-activating several debug messages, which consume significant time when being printed.

6. Conclusions and Future Work

This paper identified and elaborated on two separate problems, in the domain of wireless terminal operation within fourth-generation networks. These problems, the IAS and the MAUP, were defined, mathematically formulated and solved. Results were presented, in the context of a real-life scenario simulating the day of an ordinary user. These results were also compared with an alternative approach to the MAUP problem, involving the use of a simple rule-based system that emulates the user's reasoning.

Directions for extending our work can be the following: (a) the development of alternative heuristic algorithms for the solution of the IAS problem, (b) the expansion of features taken into account by both problems, and (c) the development of an efficient algorithm for learning the structure of the Bayesian network used for modelling user preferences.

References

- [1] U. Varshney, "Recent advances in wireless networking", IEEE Computer, vol. 33, no. 6, pp. 100-103, June 2000.
- E. Gustafsson and A. Jonsson, "Always Best Connected", IEEE Wireless Communications, vol. 10, [2] no. 1, pp. 49-55, Feb. 2003.
- "Fourth generation wireless networks and interconnecting standards", Special issue in the IEEE [3] Pers. Commun., vol.8, no. 5, Oct. 2001.
- P. Demestichas, V. Stavroulaki, L. Papadopoulou, A. Vasilakos, and M. Theologou, "Service [4] configuration and traffic distribution in composite radio environments", IEEE Transactions on Systems, Man and Cybernetics Journal, vol. 33, no. 4, pp. 69-81, Nov. 2003.
- D. Chantrain, K. Handekyn, and H. Vanderstraeten, "The Soft Terminal: Extending Service [5] Intelligence from the Network to the Terminal", Alcatel Telecommunications Review, 2nd Quarter 2000.
- B. Busropan, J. van Loon, F. Vervuurt, and R. van Eijk, "Access network selection in heterogeneous [6] networks and the role of the operator", presented at 9th WWRF Meeting in Zurich, Switzerland, July 2003.
- IST Project E2R2, official web site, http://e2r2.motlabs.com. [7]
- M. Weiser, "The computer for the 21st century", Scientific American, vol. 265(3), pp. 94-104, [8] September 1991; reprinted in: IEEE Pervasive Computing, Jan.-Mar. 2002, pp. 19-25.
- J. Pascoe, N. Ryan, and D. Morse, "Issues in developing context-aware computing", in H.-W. [9] Gellersen, editor, Handheld and Ubiquitous Computing, no. 1707 in Lecture Notes in Computer Science, pp. 208-221, Springer-Verlag, 1999.
- E. Homayounvala, S. A. Ghorashi, and A. H. Aghvami, "A Bayesian approach to modelling user [10] preferences for reconfiguration", presented at E2R Workshop on Reconfigurable Mobile Systems and Networks Beyond 3G, Barcelona, Spain, 2004.
- [11] M. Dillinger, K. Madani, and N. Alonistioti, Software Defined Radio: Architectures, Systems and Functions, John Wiley & Sons, 2003.
- [12] IST Project SCOUT, official web site, http://www.ist-scout.org.
- E. Adamopoulou, K. Demestichas, A. Koutsorodi, and M. Theologou, "Intelligent access selection in [13] heterogeneous networks - simulation results", in: Proc. 2nd International Symposium on Wireless Communication Systems 2005 (ISWCS 2005), Sienna, Italy, pp. 279-283, Sept. 2005.
- A. Koutsorodi, E. Adamopoulou, K. Demestichas, and M. Theologou, "User profiling and preference modeling in 4G terminals", in: Proc. 17th IEEE International Symposium on Personal, [14] Indoor and Mobile Radio Communications, Helsinki, Finland, pp. 1-6, Sept. 2006.
- [15] M. Liebsch, A. Singh (editors), H. Chaskar, D. Funato, and E. Shim, "The Candidate Access Router Discovery Protocol", IETF RFC 4066, July 2005.
- [16]
- C. Perkins, "IP Mobility Support for IPv4", IETF RFC 3220, Jan. 2002. D. B. Johnson, "Mobility Support in IPv6", IETF RFC 3775, June 2004. [17]
- [18] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, "SIP: Session Initiation Protocol", IETF RFC 3261, June 2002.
- C. Ahlund and A. Zaslavsky, "Multihoming with Mobile IP", in: Proc. 6th IEEE International [19] Conference on High Speed Networks and Multimedia Communications, Springer-Verlag, July 2003.
- [20] P. Demestichas, L. Papadopoulou, V. Stavroulaki, M. Theologou, G. Vivier, G. Martinez, and F. Galliano, "Wireless beyond 3G: managing services and network resources", IEEE Computer Magazine, vol. 35, no.8, pp. 80-82, Aug. 2002.
- [21] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kauffman, 1988.
- V. Terziyan, "A Bayesian Metanetwork", International Journal on Artificial Intelligence Tools, vol. [22] 14, 2004.
- D. T. Pham and D. Karaboga, Intelligent optimisation techniques: genetic algorithms, simulated [23] annealing and neural networks, Springer, 2000.
- C. Papadimitriou and K. Steiglitz, Combinatorial optimization: algorithms and complexity, Prentice [24] Hall, 1982.

- [25] D. M. Chickering, "Learning Bayesian networks is NP-complete", in: D. Fischer and H.-J. Lenz, editors, Learning from Data: Lecture Notes in Statistics, pp. 121–130, Springer-Verlag, 1996.
- [26] W. Buntine, "A guide to the literature on learning probabilistic networks from data", *IEEE Transactions on Knowledge and Data Engineering*, vol. 8, pp.195-210, 1996.
- [27] D. MacKay, "Bayesian interpolation", *Neural Computation*, vol. 4, pp. 415-447, 1992.
 [28] D. Specht, "Probabilistic neural networks", Neural Networks, vol. 3, pp. 109-118, 1990.





Figure 2. Mobility management at the mobile terminal



Figure 3. Profile handling at the mobile terminal



Figure 4. A Bayesian meta-network for modelling user preferences



Figure 5. Use of a neural network for the training of the system

Business Context - Voice Call Service



Figure 6. Scenario 1 – Business Context: Potential allocations for (a) voice call, (b) video streaming



Figure 7. Scenario 2 – Business Context: Potential allocations per service



Figure 8. Scenario 3: Potential allocations per service in (a) business context, (b) leisure context