Service Configuration and User Profiling in 4G Terminals

Artemis A. Koutsorodi[*], Evgenia F. Adamopoulou[*], Konstantinos P. Demestichas[*],
Michael E. Theologou[*], *Member, IEEE*

[*]*National Technical University of Athens, School of Electrical and Computer Engineering*
9 Iroon Polytechniou St., Zographou 157 73, Athens, Greece, Tel. +30 210 7721478
e-mail: artemis@telecom.ntua.gr, jenny@cn.ntua.gr, cdemest@cn.ntua.gr, theolog@cs.ntua.gr

Abstract— **This paper presents a middleware platform for managing devices that operate in heterogeneous environments. The proposed management framework supports terminal-controlled, preference-based access network selection. Two separate problems are identified in this domain: one involving the computation of optimal allocations of services to access networks and quality levels (service configuration), and one concerning the dynamic inference of the user's preferences, according to the usage context (user profiling). This paper includes an approach to the definition, mathematical formulation and solution of both these problems. Indicative results of the proposed solution methods are presented in the context of a real-life scenario simulating a day in the life of an ordinary user.**

Index Terms— **Access network selection, Bayesian inference techniques, user profiling, preference modelling**

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]-[3]). These access technologies, Wireless Local Area Networks (WLANs, [4],[5]), third-generation cellular systems ([6],[7]) and broadcast networks, such as Digital Video Broadcasting (DVB, [8]), will thus constitute cooperating components of a composite radio infrastructure, interconnected by a backbone (e.g. an IP-based fixed 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 ([9]-[12]).

A central issue related to the above is the design of suitable management frameworks, both for the terminal and the network, which will allow for seamless service provision, independently of the network the user is attached to. Thus, the latter will benefit from being able to access his

personal services anywhere and anytime. 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 ([13]-[15]).

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 ([16]-[18]). 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 middleware platforms must also provide a way for services and terminal settings to adapt appropriately to this constantly changing environment ([19]-[21]).

Moreover, one of the main criteria that influence the terminal's access selection process, in an environment with multiple available access technologies, 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 ([22],[23]), thus influencing in different ways the process of selecting the most appropriate access network through which services should be obtained.

Hence, the goal in developing suitable middleware platforms for 4G systems proves to be twofold: not only does it lie in the seamless integration and the joint utilization of all available technologies for the provision of higher data rates, but also in the exploitation of a variety of contextual data for the development of innovative, personalized services.

Accordingly, in the context of effectively managing wireless terminal operation in composite radio environments, two problems are identified: *Intelligent Access Selection* (IAS), involving the computation of optimal allocations of services to access networks and to quality levels ([25]-[27]), given network availability and user preferences, and *Modelling an Adaptation to User Preferences* (MAUP), involving, firstly, the processing and 'decoding' of contextual information for 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. Their combined solution enables user terminals to be driven to allocations of services to access networks and quality levels that satisfy user preferences, in the order of priority specified by the usage context.

In this paper, we argue that appropriate functionality for handling the IAS and MAUP problems, must be incorporated in a middleware platform 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 work of this paper 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, Section 3 provides the high-level and the formal description of the IAS and MAUP problems, and Section 4 discusses solution approaches. Finally, Section 5 discusses implementation issues and indicative results of the proposed solution, and Section 6 concludes the paper.

2.   Requirements and design

2.1.   Requirements

The most salient feature of evolving systems will be the multiplicity of access technologies and terminals, which will allow users to benefit from the realization of the *always best connected* (ABC) concept: being connected at any point in time to the best available access network. The latter is defined as the network that, currently, best suits the user's needs, and may be a function of the user's personal preferences, the operators' policies and business agreements and the infrastructure's available network resources ([3]).

Besides, the enhanced capabilities of today's terminals have introduced the idea of extending service intelligence from the network to the terminal. The trend of assigning more and more complex tasks to the user terminal is gaining ground ([14],[15],[24],[28]), since modern mobile phones constitute small – but increasingly powerful – terminals that are equipped with advanced operating systems (Windows Mobile, Symbian), that can run applications from different sources, that are configured with a standard IP stack to support the communication with other users and application servers over the Internet, that also incorporate a powerful CPU (typically, a 32-bit RISC CPU based on ARM-9 series, with clock speed varying from 104 MHz to 220 MHz), a relatively big colour screen (e.g., 240x320 screen resolution, with 18-bit colour depth, supporting

262,144 colours) and sufficient memory (internal dynamic memory of around 25 MB which can be extended by additional storage cards) to store and run additionally installed software, and that may, finally, be equipped with multiple network interfaces to enable network connectivity using different types of access networks (GSM, UMTS, WLAN).

There are several arguments − both from a technical and a business viewpoint − for the terminal to incorporate functionality for performing access selection and user profiling: *(i)* The terminal is aware of the different access technologies in its neighbourhood: it knows which hardware interfaces it is equipped with and it can detect the availability of access networks in its physical surroundings. *(ii)* The terminal is aware of the applications that are already running on it, and can also specify which additional services it can support based on its capabilities (i.e., screen resolution, memory, etc.). *(iii)* In a heterogeneous environment, with multiple available access technologies, information about network availability can be highly dynamic. The mobile terminal is responsible for constantly monitoring access network availability and unavailability − as these are perceived by its network interfaces − and may thus acquire direct access to a wide range of network characteristics (signal strength, available bandwidth in nearby access points, etc.) Let it be noted that by '*access point*' (AP) we shall refer to any available network attachment point available in the infrastructure. *(iv)* The user is in direct contact with his terminal, therefore his preferences can be stored in the mobile device easily and promptly. *(v)* The user can autonomously make decisions in his own best interest.

In all, the efficient exploitation of the composite radio infrastructure calls for the development of an appropriate, sophisticated *Terminal Management System* (TMS), capable of: *(i)* Detecting available access networks; *(ii)* Receiving and processing measurements regarding the capabilities of available APs (e.g., supported technology, network operator, cost at which a service may be

provided at each quality level, etc); *(iii)* Accessing, modifying and storing the user profile; *(iv)* Allowing the user to dynamically redefine his preferences. That is, providing the user with a practical mechanism through which he will be able to specify which factor he considers to be the most important one in access network selection, which factor he considers to be second most important, and so on. Such factors may include the use of a preferred type of technology or a preferred network operator, the cost at which services at various quality levels are offered and the quality the user may demand to obtain for a specific service; *(v)* Alternatively, the terminal may incorporate an appropriate mechanism for adapting automatically to the user's preferences, according to the identified usage context (e.g., business or leisure context). This implies the development of suitable functions for the detection and processing of various parameters that determine the usage context (e.g., time, location, type of service content, communicating counterpart, etc), and for automatically deducing the user's preferences based on this contextual information; *(vi)* Optimally selecting allocations of services to access networks and to quality levels ([25]-[27]), by taking into consideration network availability, user preferences and requested and running services' requirements, as well as activating and configuring the connections to appropriate APs; *(vii)* Supporting the applications in handing over the existing connections from one access network to another in a seamless manner.

The TMS must be modelled for a modern, general-purpose operating system, running on a mobile terminal that may be continuously switched on ([28],[29]). Power management and power conservation issues are out of scope.

2.2. Design

Figure 1 illustrates the main components of the proposed TMS architecture, namely the *Network Interface Adaptation Module* (NIAM), the *Mobility Management Module* (MMM), and

6

the *User Preferences Module* (UPM), in the context of a generic B3G network infrastructure. These modules are examined in detail in the following paragraphs. The detailed description of network-side (as opposed to terminal-side) B3G components is out of scope. However, Section 2.2.1 includes a brief discussion of requirements for external (i.e., network-side) entities.

2.2.1. Adaptation to network drivers

The NIAM is responsible for identifying the different interfaces present in the terminal, for monitoring their status, for collecting measurements from each interface and for executing the selection and de-selection of the appropriate interface. More specifically, the NIAM serves two purposes: *(a)* the connection and de-connection of the appropriate interface during the terminal's power-up or during a handover, and *(b)* the retrieval of layer-2 measurements in the network interface. In all, the NIAM is able to provide the terminal with a list for each AP in its range, each list comprising information about the AP's: *(i)* identifier, *(ii)* signal strength, *(iii)* type of technology, *(iv)* network operator, *(v)* available bandwidth, and *(vi)* cost at which it is able to provide each available service at each of the service's permissible quality levels. This information is subsequently used by the MMM module, for actually selecting the most appropriate AP for each service.

Although the first four types of information can be obtained in a straightforward manner, the retrieval of information concerning the available bandwidth and cost (types *(v)* and *(vi)*) requires the presence of additional entities at the network side. The two external entities needed are the Bandwidth Broker and the Cost Broker. The information flow between the NIAM and these two entities is depicted in Figure 2.

The Bandwidth Broker is responsible for querying the APs in its proximity, in order to obtain estimations about the bandwidth availability of each one. The functionality of this entity follows

the paradigm of the Candidate Access Router Discovery protocol ([30]). Although the latter is intended solely for supporting the selection of (target) access routers during handoffs, the same mechanism can potentially be used for the purpose of IAS. The presence of network entities equipped with such functionality has already been proposed by the research community ([31],[32]).

The Cost Broker is responsible for providing information about the cost at which each network operator provides each specific service. This cost is expressed in monetary units per time or volume unit (i.e., seconds or KBs, respectively). Cost-related values depend not only on the network operator and the type of service, but also on the type of access technology used and the quality level offered. The Cost Broker's role is to hold all this information in a single XML document, which is regularly updated by communicating with the network operators' billing systems. Thus, the Cost Broker acts a central entity, which gathers cost-related information. At the terminal side, the NIAM is responsible for downloading the most recent version of the XML Costs File from the Cost Broker. As prices are not likely to change on a daily basis, this function needs not be performed frequently (e.g., only once per day).

2.2.2. Service configuration

Service configuration and access selection issues in 4G terminals may be considered to fall under the broad umbrella of 'mobility management' functions. Mobility management encompasses functions such as session continuity, which involves the ability to maintain a session when the ABC terminal moves between different access networks, session transfer, which involves the ability to maintain a session when the user moves between different devices, and user reachability, which involves the ability to reach a user at his current access network and device. Out of this group of problems, one of the most significant, both from a theoretical and a

technical perspective, is related to the processing of the terminal's various connectivity options and the determination of the option that guarantees optimal service provision.

This constitutes the short- to mid-term optimization problem previously identified as IAS, whose solution enables the selection of the appropriate access technology through which the terminal will receive each service efficiently in terms of quality and cost, in near real time. The solution to the IAS problem, therefore, consists of an optimal allocation of services to access networks and to quality levels.

The MMM depicted in Figure 1 incorporates functionality for managing mobility both at the link layer (computation of optimal allocations to APs 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 ([33]-[36]) and SIP ([37]). 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 ([38]-[40]). These functions are presented in Figure 3.

4G mobility management and service configuration has been addressed in a number of papers. Most address the problem of aggregate traffic (re)distribution in composite radio environments ([11],[41]), while others discuss the development of management platforms for handling the mobile terminal's access selection problem in a macroscopic scale ([12],[42],[43]). In most approaches, the computational load related to the terminals' vertical handovers is assigned to the network's central management unit, entailing potential overload and delay problems, while the

applications running on each terminal are not considered as individual handover objects, 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. By 'preferred network operator' we refer to a network operator, which the user specifies, and through which he would 'prefer' (by comparison to other operators) to be served, e.g. due to a prior favourable experience in dealing with this operator. 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 ([44]-[46]). The goal of such a system is the prediction of the values of user preferences, every time the IAS function is initiated. This presupposes the development of appropriate mechanisms through which (Figure 4(a)): *(i)* The usage context is inferred, given contextual information (user location, time and date, application type being requested, terminal

10

type being used, content type being accessed – e.g., web page with business data, phone call to family, etc.); *(ii)* The user preferences are inferred, given usage context and service requests.

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 ([22],[53]).

3. Problem description

3.1. Intelligent Access Selection

3.1.1. High-level description

The input to the IAS problem falls into three categories: *(i)* A set of measurements, reflecting the capabilities of available APs, as perceived by the mobile terminal. The information contained in these measurements has been specified in Section 2.2.1; *(ii)* The set of applications that are 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; *(iii)* The set of user preferences, specifying the priority of the parameters 'quality', 'network operator', 'technology type' and 'cost'.

The solution of the IAS problem is the result of an optimization process that should produce the following results: *(i)* An allocation of services to quality levels; *(ii)* An allocation of services to APs. 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 into the following categories: *(i)* Each service must be allocated to a permissible quality level; *(ii)* The maximum quality level that the user defines for each service must not be violated; *(iii)* The user's preferences must be taken under 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: *(i)* Whenever the user issues a service request; *(ii)* Whenever the user modifies his profile; *(iii)* Whenever serious signal degradation, or complete signal loss, from an AP that the terminal is connected to is detected. *(iv)* Whenever a new available AP 'appears' in the terminal's neighbourhood.

3.1.2. Formal description

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 APs that the terminal perceives and let $Q_{p,s}(p) = \{q_1, q_2, ... q_m\}, m \in \square$ be the set of quality levels at which AP p, $p \in P$, may offer service s, $s \in S$. The size of set $Q_{p,s}$ depends on both the bandwidth capabilities of AP p and the type (specific characteristics) of service s.

We aim to compute allocations of services

- To quality levels: we denote this allocation as $A_{S-QL} = \{ql(s) \mid \forall s \in S\}$

- To APs: we denote this allocation as $A_{S-AP} = \{ap(s) \mid \forall s \in S\}$

These allocations should optimize the objective function

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

where $F(p,q,s) = w_q \cdot F_Q(p,s) + w_t \cdot F_T(p) + w_o \cdot F_O(p) - w_c \cdot F_C(p,q,s)$,
$F'(p,q,s') = w_q \cdot F'_Q(p,s') + w_t \cdot F_T(p) + w_o \cdot F_O(p) - w_c \cdot F_C(p,q,s')$ and
$F'_Q(p,s') = f(BW_{Av}(p) - BW(s,q))$, subject to:

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

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

$$ql(s) \le ql_{s, max}, \ \forall s \in S \qquad (4)$$

and $OF(\varnothing, P) = 0$.

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 ($SS(p)$). More specifically, $F_Q(p,s)$ takes into account the maximum quality level supported by $BW_{Av}(p)$ and a multiplication factor which is a function of $SS(p)$. Thus, high bandwidth availability and strong signal strength are jointly evaluated for determining $F_Q(p,s)$.

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: $F_Q(p,s) - F_C(p,q,s)$.

Factor $F_C(p,q,s)$ expresses the cost at which AP p offers service s at quality level q.

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 $F'_Q(p,s')$.

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.

We also note that the proposed objective function is scalable with respect to the factors that may be taken into account for selecting the most appropriate allocations A_{S-QL} and A_{S-AP}. An interesting extension of the function defined in (1) involves the incorporation of an additional coefficient - factor pair corresponding to the security level provided by each AP under consideration: $w_S \cdot F_S(p)$. Each AP's security level may be classified in a predefined scale, e.g. poor-average-good, and be made available to the terminal through similar procedures as those supporting the availability of the other elements in (1).

3.2. Modelling and Adaptation to User Preferences

3.2.1. High-level description

We propose the construction of a two-level Bayesian meta-network (based on an approach presented in [22]) (Figure 4(b)), 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. This structure assumes that the interoperability between the component networks (on the contextual and predictive levels) can also be modelled by another Bayesian network. Thus,

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 [22] 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.

3.2.1.1. Stage 1: Inference of usage context

The Bayesian network of the second level (contextual level - Figure 4(b)) 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.

3.2.1.2. Stage 2: Inference of profile parameters

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

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.

3.2.1.3. 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 (allocation) 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's 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 accurately 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. 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.

3.2.2. Formal description

The following subsections provide the formal description of the MAUP problem.

3.2.2.1. Inference of usage context

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:

$$P(C = c \mid T = t, L = l, CP = cp)$$

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.

3.2.2.2. Inference of profile parameters

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 και w_c – let a, b, c, d be these values – for which the following probability is maximized: $P(w_q = a, w_c = b, w_t = c, w_o = d \mid S = s)$

It holds that: $P(w_q, w_c, w_t, w_o \mid S) = \dfrac{P(w_q, w_c, w_t, w_o, S)}{P(S)}$, and according to the independence relationships encoded in the structure of Figure 4(b) (predictive level):

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

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

and the computation of $P(w_q, w_c, w_t, w_o \mid S)$ reduces to the computation of the product of partial conditional probabilities, which result from tables of the format of Table 1 (for $i = q, c, t, o$, and for every possible value of variable C).

3.2.2.3.   Model learning and adaptation

The estimate of values for w_q, w_o, w_t and w_c that is produced by the system is subsequently used as input to the IAS optimization problem. The algorithm that solves this problem computes some allocations (of services to quality levels and APs) which are subjected to the user for his approval or rejection, with the following interpretations: *(i)* User approval: the system was successful in predicting his preferences, for the given context and service. *(ii)* User rejection: the system was not successful in predicting his preferences, for the given context and service.

This process results in the formation of a training set through which the Bayesian network will be 'trained' to accurately predict the user's preferences. In general, learning Bayesian networks (BNs) from data involves two aspects: learning network structure and learning network parameters. There are different ways to view a BN, each suggesting a particular approach to learning ([47]): For example, a BN may be considered to be a structure that encodes the joint distribution of attributes. This suggests that the best BN is the one that best fits the data, and leads to the *scoring-based* learning algorithms, which seek a structure that maximizes some kind of scoring function ([46],[48]). Alternatively, the BN structure may be considered to encode a group of conditional independence relationships among the nodes, according to the concept of *d-separation* ([44]). This suggests learning the BN structure by identifying the conditional independence relationships among the nodes. These algorithms are referred to as *CI-based* algorithms or constraint-based algorithms ([49],[50]).

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 ([51],[52]). 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 the network's parameters (values of CPTs).

The training, or 'learning', of the Bayesian network can be achieved using a variety of different methods, such as linear regression ([54]-[57]), probabilistic neural networks ([58]-[61]) and probabilistic decision trees ([62],[63]).

4. Problem solution

4.1. IAS

This section discusses computationally efficient solutions for the IAS problem. 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 ([64],[65]). In this paper, we rely on the greedy algorithm paradigm ([66]) for the development of a faster approach.

The algorithm runs in three main phases. The first phase (steps 0 and 1), called Input Retrieval and Initial Processing (IRIP), generates the list of candidate APs and detects the user's preferences for each service. The second phase (step 2), called Computation of Service Allocations (CSA), generates different feasible allocations of services and evaluates each of them. The third phase (step 3), called Selection of Aggregate Solution (SAS), finds the aggregate solution that has the highest objective function value.

- Step 0: The measurements reflecting the current network conditions, i.e. APs' availability and bandwidth, signal strength, etc., are retrieved through the NIAM, and processed, so as to construct the complete list P of APs operating in the terminal's neighbourhood.

19

- Step 1: The list of all running/requested services is used, in combination with contextual information available, in order to automatically resolve the user's preferences. Thus, for each service, the corresponding context type is decided and, thereby, the coefficients w_q, w_o, w_t and w_c, are calculated.

- Step 2: All possible permutations of set S are determined, and for each permutation the following steps are executed:

  - Step 2(a): The permutation's first element is retrieved.

  - Step 2(b): Let s be the current service element. For each AP $p \in P$, the quality levels at which AP p is able to provide service s are calculated. This is accomplished by comparing $BW_{Av}(p)$ with the bandwidth requirements of each permissible quality level of service s, and then adjusting for $SS(p)$. Thus, a list of allocation choices is produced for service s, each allocation choice consisting of two parameters: an AP and a quality level at which service s may be provided.

  - Step 2(c): For each allocation choice specified in Step 2(b), the value of $F(p,q,s)$ (relation (2)) is calculated. The allocation choice that produces the maximum value is selected, and the maximum value is added to the overall 'grade' of the permutation under consideration.

  - Step 2(d): The bandwidth of the AP that was selected as part of the optimum allocation choice is modified: it is reduced by the bandwidth required for the allocation of service s determined in Step 2(c).

  - Step 2(e): The next element of the current permutation is retrieved and the solution process jumps back to Step 2(b).

- Step 3: The optimum solution of allocating the set of running and requested services to APs and quality levels is selected, by determining the permutation that corresponds to the maximum aggregate 'grade'.

- Step 4: End.

Let m be the number of services that are running or have been requested at the mobile terminal (equivalent to size of set S), n the number of APs in the terminal's neighborhood (equivalent to size of set P), and ℓ the overall maximum number of quality levels, regardless of the service and AP in question. Then, the complexity of the above-presented algorithm is $O\big(n \times \ell \times \big((m+1)!\big)\big)$, in the worst case.

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 smart-phones, there are usually no more than three or four services running simultaneously, while, on the other hand, more advanced devices, such as notebooks, can easily handle the computational load.

It is worth mentioning that the efficiency of this algorithm can be further enhanced by the parallelized execution of Step 2 for all possible permutations of the service list. It is also possible to achieve a more balanced approach by executing the aforementioned algorithm only in the case of the appearance or disappearance of APs, whereas requests for new services may trigger an even 'greedier' algorithm, which would compute the best allocation for the newly requested service alone, without trying to reallocate the already running services. Such an approach would significantly reduce the computational load.

4.2. MAUP

In this paper, we propose the construction of a probabilistic neural network (PNN) 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(a).

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: *(i)* It assesses the similarity of the new instance ($C = c$, $S = s$, $w_q = a$, $w_o = b$, $w_t = c$, $w_c = d$) to the patterns stored in the training set (pattern layer neurons functionality). *(ii)* It produces a probability vector, comprising classification probabilities to each class: $k_1 = {'}accept{'}$, $k_2 = {'}reject{'}$ (summation layer neurons functionality). *(iii)* It selects the maximum of these probabilities, using the Bayes-optimal decision rule (output layer neuron functionality). The architecture of the employed PNN is drawn in Figure 5(b).

The PNN's output is subsequently used for the reinforcement or the diminishment of the conditional probabilities $P(w_i | S)_C$ in the corresponding CPTs, in the case of classification of the system's decision into class k_1 or k_2 respectively. This correction of the $P(w_i | S)_C$ values is realized by applying Bayes' theorem, which allows the expression of the posterior probability that variable w_i may obtain a specific value given the newly performed classification, using prior probability $P(w_i | S)_C$ and the probability of performing this classification given the value of w_i (class-conditional probability):

$$P((w_i | S)_C | k_j) = \frac{P(k_j | (w_i | S)_C) \cdot P(w_i | S)_C}{P(k_j)}$$

where $P(k_j) = P(k_j | (w_i | S)_C) \cdot P(w_i | S)_C + P(k_j | \neg(w_i | S)_C) \cdot P(\neg(w_i | S)_C)$ is a normalizing factor, which can be computed by requiring that $P((w_i | S)_C | k_j)$ and $P(\neg(w_i | S)_C | k_j)$ sum to 1.

5. Implementation and evaluation

5.1. Implementation

A Java-based implementation of the TMS discussed in this paper has been completed. The Java-based implementation relies on the Java 2 Micro Edition (J2ME) platform, and is targeted for terminal devices supporting the Connected Limited Device Configuration (CLDC) v1.1, and the Mobile Information Device Profile (MIDP) v2.0 (these devices constitute the large majority of Java-enabled terminal devices). J2ME is ideal for terminal devices that are resource-constrained, such as mobile phones, and is supported by the mass-market.

The implementation of the external entities, namely the Bandwidth Broker and the Cost Broker, is based on the Java 2 Enterprise Edition (J2EE) v1.4 platform. For simulation purposes, the information provided by the Bandwidth Broker does not originate from actual interactions with APs, but is rather randomly generated following some scenario guidelines. The communication between the TMS and the Brokers is realized through Web Services, using the Simple Object Access Protocol (SOAP) over HTTP.

Figure 6 illustrates the typical sequence diagram that is followed whenever the MMM is triggered to perform the IAS function (the triggering cases have been discussed in Section 3.1.1). The components of the TMS have been implemented as separate Java-class objects. NIAM also runs in a separate thread. Signalling between the components of the TMS follows the familiar *listener* model, which is widely employed in several Java APIs, in order to enable objects to listen for specific events.

Initially, as illustrated in Figure 6, the MMM makes a request for measurements to the NIAM. The latter forms the list of the APs that are currently in the terminal's neighbourhood, then communicates with the Bandwidth Broker, in order to retrieve an estimate of the bandwidth

availability of each AP in the list. The composed set of measurements is then relayed to the MMM. The next step is the retrieval of the running services, which can normally be accomplished through a system call to the terminal's native OS. For each of the running services, the corresponding profile parameters must be obtained. For this purpose, for each running service an appropriate request to the UPM is formed, the response to which is generated after the completion of three stages (Inference of usage context, Inference of profile parameters, and PNN). After this step, all required input is available so as to execute the IAS function.

5.2. Evaluation

The scenario presented herein simulates a typical day in the life of an ordinary user X. We consider two usage contexts (business and leisure) and three services (voice call, video streaming and web browsing). A high-level description of user X's preferences is given in

Table **2**. The user has specified 'Oper#1' as his preferred operator and WLAN as his preferred technology for video streaming and voice call. The lowest quality level for each of the above-mentioned services is level '1', while the highest one is level '5'.

On the basis of these preferences, the user provides some 'accept' or 'reject' answers to sample configurations of w_q, w_o, w_t, w_c (

Table 3). 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. We also note that 'Q' stands for quality, 'C' for cost, 'O' for network operator and 'T' for technology, and that 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.

These answers constitute the training set, based on which the system is trained to accurately and independently predict the user's preferences. It should 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.

In the following, every step of the scenario, along with the corresponding allocations of services to networks and quality levels, are described in more detail. Figure 7 illustrates network coverage in each step of the scenario. For every available AP, its access technology, network operator, signal strength and available bandwidth are indicated.

(a) Morning, at home: X initiates a web browsing service and visits a sport news web site, while at home (Table 4). At first, the TMS determines the usage context. Based on current location (home), time (morning) and communicating counterpart (sports web page), context is defined as 'leisure'. Then, the system successfully predicts X's preferences for the specified context and service. 'Cost' is deemed more important than 'quality', and 'technology' more

25

important than 'operator'. Thus, the WLAN AP is selected in this case, in combination with the lowest quality level, as this is the most cost-effective solution. This decision is also reinforced by the fact that WLAN is the preferred technology type.

(b) Morning, en route to work: X initiates a voice call to a colleague, while en route. The web browsing session initiated in step (a) is still ongoing (Table 4). For the voice call service, the usage context type is defined as 'business' and thus 'quality' is more important than 'cost' and, consequently, a UMTS AP and quality level '5' are selected. For the web browsing service, the same AP and quality level '4' are selected. The selected AP (UMTS #1) also belongs to the preferred operator.

(c) Noon, at work: X initiates a video streaming service, while at work. After terminating this service, he makes a voice call and initiates a web browsing session (Table 4). For video streaming, the DVB AP is selected, as it is able to provide the maximum quality level, due to its high available bandwidth. This AP is also characterized by strong signal. In contrast, voice call and web browsing are allocated to the WLAN AP and delivered at the minimum quality level, as their usage context is defined as 'leisure'.

It should be noted at this point that the user's preferences where web browsing in leisure context is concerned, have already been determined and cached in phase (a), so there is no need for re-evaluation. The system's ability to cache already determined user preferences for a particular service and context further reduces the computational load.

(d) Evening, at home: X initiates a web browsing session, while at home, visiting his corporate web site (Table 4). Web browsing is allocated to the WLAN AP and to the maximum quality level, as its usage context is 'business'. The WLAN AP is able to provide the maximum quality level. The decision is also reinforced by the fact that WLAN is the preferred technology.

An interesting point involves the comparison of the proposed IAS algorithm to the simplest available and widely adopted scheme, according to which the selection of a new AP whose signal strength indication surpasses the currently selected AP's signal strength indication, by a certain threshold for a certain amount of time, is favored. This method shall be referred to as *Best Signal Strength* (BSS).

Table **5** depicts the service allocations (to APs and quality levels) computed by the IAS algorithm and by the BSS method, in each scenario step. In the chart depicted in Figure 8, the y-axis represents the aggregate utility volume U of each step allocation, computed using both the IAS algorithm and the BSS method. Utility is a term in economics ([67]), defined as the aggregate sum of satisfaction or benefit that the user gains from the consumption of goods or services. Utility is an abstract concept rather than a concrete, observable quantity, but is nevertheless a very useful tool for analyzing consumer choices and behavior ([68],[69]). In fact, the objective function $OF(p,q)$ defined in Section 3 may be interpreted as a measure of the user's benefit from each service allocation, and is used for computing utility volumes in each scenario step. The comparison depicted graphically herein renders apparent the benefits of the IAS algorithm in service configuration and access selection: there is a clear trend in maximizing the overall level of quality delivered to the user and in raising the utility associated with service usage.

Another 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. This comparison is at the core of a broader academic discussion involving the most appropriate method for representing knowledge in machine learning

problems. In general, it has been suggested that the main advantage of rule-based representations is their modularity, and their relatively simple inference procedures ([70],[71]). On the other hand, it is argued that rule-based methodologies impose strong restrictions on the kinds of dependencies that can be adequately represented, and that the implicit nature of the assumptions they incorporate have contributed to a 'myth' regarding their scalability ([72],[73]). In the simple scenario that we are investigating in this paper, where the user's profile comprises only four parameters, we cannot provide a definitive answer as to the clear advantage of using the proposed probabilistic model instead of a well-formulated rule-based system. We conjecture, however, that even the best rule-based system's ability to accurately predict the user's preferences is exhausted in similar, simple cases of causal relationships between variables, and the extension of the set of rules so as to include more dependencies is disproportionately difficult. In contrast, our proposed approach is easily extendible to any number of profile parameters, without entailing a severe computational load.

6. Conclusions and future work

This paper analyzed two main problems identified in the domain of wireless terminal operation within fourth-generation, heterogeneous networks: one involving the optimal configuration of services received at the mobile terminal and the management of user requests for new services; and one involving the exploitation of contextual information for the purpose of determining the user's profile. These two problems, the IAS and the MAUP respectively, were defined, mathematically formulated and solved. More specifically, the IAS problem was formally described and solved using a greedy algorithm for computing optimal (or near-optimal) service allocations in real-time. The MAUP problem was approached by constructing an appropriate

probabilistic model. In this context, the formal description of the parameter-learning problem in this model was given, and a solution that uses a probabilistic neural network was proposed.

Furthermore, implementation issues concerning the integration of the proposed solutions into an appropriate management platform were also discussed: the architectural 'blueprints' of the implemented terminal management system were provided, including details on the programming language and development platform used. The signalling and information flow, both internal (among the system's components) and external (communication with external entities) was also described. Finally, the proposed solutions were tested in the context of a simple service-usage and mobility scenario simulating the activities of an ordinary user.

The work presented in this paper is being extended in the following direction: we are exploring approaches to the problem of learning the *structure* of the Bayesian network in the predictive level, i.e. using the training set provided by the user to learn the causal relationships between the random variables T, L, CP and S, for different contexts. More specifically, this problem involves the evaluation of the goodness-of-fit of different network structures to the available data, and the use of a search algorithm which identifies network structures with high evaluations.

We believe this work may provide the basis for more complex context-aware service configuration procedures, which may take into account a variety of factors possibly influencing access selection, and which may consequently exploit a very sophisticated user profile. We believe that this growing complexity may be efficiently tackled using probabilistic user-profiling approaches, similar to the one discussed in this paper.

References

[1] U. Varshney, "Recent advances in wireless networking", *IEEE Computer*, vol. 33, no. 6, June 2000.

[2] "Next generation wireless broadband networks", Special issue in *IEEE Network*, vol. 14, no. 5, Sept./Oct. 2000.

[3] E. Gustafsson, A. Jonsson, "Always Best Connected", *IEEE Wireless Communications*, vol. 10, no. 1, pp. 4955, Feb. 2003.

[4] U. Varshney, R. Vetter, "Emerging mobile and broadband wireless networks", *Commun. of the ACM*, vol. 43, no. 6, June 2000.

[5] Institute of Electrical and Electronics Engineers (IEEE), 802 LAN/MAN Standards Committee web site, www.ieee802.org, June 2005.

[6] "Third generation mobile systems in Europe", Special issue in *IEEE Personal Commun.*, vol. 5, no. 2, April 1998.

[7] "Wideband CDMA", Feature topic in the *IEEE Commun. Mag.*, vol. 36, no. 9, Sept. 1998.

[8] Digital Video Broadcasting web site, www.dvb.org, May 2005.

[9] "Management of heterogeneous networks", feature topic in the *IEEE Commun. Mag.*, vol. 36, no. 3, March 1998.

[10] "Fourth generation wireless networks and interconnecting standards", Special issue in the *IEEE Pers. Commun.*, vol.8, no. 5, Oct. 2001.

[11] P. Demestichas, V. Stavroulaki, L. Papadopoulou, A. Vasilakos, M. Theologou, "Service configuration and traffic distribution in composite radio environments", *IEEE Trans. on Systems, Man and Cybernetics Journal*, vol. 33, no. 4, pp. 69-81, Nov. 2003.

[12] P. Demestichas, G. Vivier, K. El-Khazen, M. Theologou, "Evolution in wireless systems management concepts: from composite radio to reconfigurability", *IEEE Commun. Mag.*, vol. 42, no. 5, May 2004.

[13] J. van Bemmel, H. Teunissen, D.-J. Plas, A. Peddemors, "A Reference Architecture for 4G Services", *7th WWRF Meeting*, Eindhoven, The Netherlands, Dec. 2002.

[14] D. Chantrain, K. Handekyn, H. Vanderstraeten, "The Soft Terminal: Extending Service Intelligence from the Network to the Terminal", Alcatel Telecommunications Review, April 2000.

[15] B. Busropan, J. van Loon, F. Vervuurt, R. van Eijk, "Access network selection in heterogeneous networks and the role of the operator", *9th WWRF Meeting*, Zurich, Switzerland, July 2003.

[16] M. Esler, J. Hightower, T. Anderson, G. Borriello, "Next century challenges: Data-centric networking for invisible computing", *5th ACM/IEEE International Conference on Mobile Computing and Networking*, pp. 256-262, Aug. 1999.

[17] M. Weiser, "The computer for the 21st century", Scientific American, vol. 265(3), pp. 94-104, Sept. 1991; reprinted in: *IEEE Pervasive Computing*, pp. 19-25, Jan.-Mar. 2002.

[18] R. Grimm, T. Anderson, B. Bershad, D. Wetherall, "A system architecture for pervasive computing", *9th ACM SIGOPS European workshop: beyond the PC: new challenges for the operating system*, pp. 177-182, Sept. 2000.

[19] M. Weiser, "Some computer science issues in ubiquitous computing", *Commun. of the ACM*, vol. 36, no.7, pp. 75-84, July 1993.

[20] J. Pascoe, N. Ryan and D. Morse, "Issues in developing context-aware computing", *First International Symposium on Handheld and Ubiquitous Computing*, Lecture Notes in Computer Science, vol. 1707, pp. 208-221, Springer-Verlag, 1999.

[21] K. Henricksen, J. Indulska, A. Rakotonirainy, "Modeling context information in pervasive computing systems", *Pervasive 2002,* Lecture Notes in Computer Science, vol. 2414, pp. 167-180, Springer-Verlag, 2002.

[22] E. Homayounvala, S. A. Ghorashi, A. H. Aghvami, "A Bayesian approach to modelling user preferences for reconfiguration", *E2R Workshop on Reconfigurable Mobile Systems and Networks Beyond 3G*, Barcelona, Spain, Sept. 2004.

[23] M. Dillinger, K. Madani and N. Alonistioti, *Software Defined Radio: Architectures, Systems and Functions*, John Wiley & Sons, 2003.

[24] P. Lettieri, M. B. Srivastava, "Advances in wireless terminals", *IEEE Personal Communications*, vol. 6, no. 1, pp. 6-19, Feb. 1999.

[25] J. Irvine, "Adam Smith goes mobile: Managing services beyond 3G with the digital marketplace", European Wireless 2002, Florence, Italy, Feb. 25-28, 2002.

[26] M. O'Droma et al., "Always Best Connected Enabled 4G Wireless World", Proc. of the 12th IST Summit on Mobile and Wireless Communications, Aveira, Portugal, pp.710-716, June 2003, ISBN 972-98368-7.

[27] V. Gazis, N. Alonistioti, and L. Merakos, "Toward a generic Always Best Connected capability in integrated WLAN/UMTS cellular mobile networks (and beyond)", IEEE Wireless Communications Magazine, Vol. 12, no. 3, pp. 20-29, 2005.

[28] M. Bargh, D. Bijwaard, H. Zandbelt, E. Meeuwissen, A. Peddemors, "Mobility management in beyond 3G environments", *9th WWRF Meeting*, Zurich, Switzerland, July 2003.

[29] M. Bargh, H. Zandbelt, A. Peddemors, "Managing mobility with federating service platforms", *IST-EVOLUTE Workshop on "Beyond 3G Evolution of Systems and Services"*, Guildford, UK, Nov. 2003.

[30] M. Liebsch, A. Singh (editors), H. Chaskar, D. Funato, E. Shim, "The Candidate Access Router Discovery Protocol", IETF RFC 4066, July 2005.

[31] Simplicity IST-2004-507558 Project, Deliverable D2302, "Enhanced Performance Evaluation of system mechanisms", Dec. 2004.

[32] G. Judd and P. Steenkiste, "Providing Contextual Information to Ubiquitous Computing Applications", *IEEE International Conference on Pervasive Computing and Communications (PerCom 2003)*, pp. 133-142, March 2003.

[33] C. Perkins, "IP Mobility Support for IPv4", IETF RFC 3220, Jan. 2002.

[34] D. B. Johnson, "Mobility Support in IPv6", IETF RFC 3775, June 2004.

[35] R. Koodli (editor), "Fast Handovers for Mobile IPv6", IETF RFC 4068, July 2005.

[36] K. El Malki (editor), "Low Latency Handoffs in Mobile IPv4", IETF draft-ietf-mobileip-lowlatency-handoffs-v4-10.txt, work in progress, Dec. 2004.

[37] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, E. Schooler, "SIP: Session Initiation Protocol", IETF RFC 3261, June 2002.

[38] C. Ahlund, A. Zaslavsky, "Multihoming with Mobile IP", *6th IEEE International Conference on High Speed Networks and Multimedia Communications*, Lecture Notes in Computer Science, vol. 2720, pp. 235-243, Springer-Verlag, July 2003.

[39] N. Montavont et. al., "Analysis of Multihoming in Mobile IPv6", IETF draft-montavont-mobileip-multihoming-pb-statement-04.txt, work in progress, June 2005.

[40] T. Ernst, "Goals and Benefits of Multihoming", IETF draft-ernst-generic-goals-and-benefits-01.txt, work in progress, Feb. 2005.

[41] P. Demestichas, L. Papadopoulou, V. Stavroulaki, M. Theologou, G. Vivier, G. Martinez, F. Galliano, "Wireless beyond 3G: managing services and network resources", *IEEE Computer Mag.*, vol. 35, no.8, pp. 80-82, Aug. 2002.

[42] P. Demestichas, V. Stavroulaki, "Issues in introducing resource brokerage functionality in B3G composite radio environments", *IEEE Wireless Comm. Mag.*, vol. 11, no. 5, pp. 32-40, Oct. 2004.

[43] G. Carneiro, J. Ruela, M. Ricardo, "Cross-layer design in 4G wireless terminals", *IEEE Wireless Comm. Mag.*, vol. 11, no. 2, pp. 7-14, April 2004.

[44] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kauffman, 1988.

[45] V. Terziyan, "A Bayesian Metanetwork", *International Journal on Artificial Intelligence Tools*, vol. 14, no.3, pp. 371-384, World Scientific, June 2005.

[46] D. Heckerman, "A Tutorial on Learning with Bayesian Networks", Microsoft Research, Redmond, WA, Tech. Rep. MSR-TR-95-06, March 1995.

[47] J. Cheng, R. Greiner, "Learning Bayesian belief network classifiers: algorithms and system", *14th Biennial Conference of the Canadian Society for Computational Studies of Intelligence*, Lecture Notes in Computer Science, vol. 2056, pp. 141-151, Springer-Verlag, 2001.

[48] G. F. Cooper, "Computational complexity of probabilistic inference using Bayesian belief networks", *Artificial Intelligence*, vol. 42, no. 2-3, pp. 393-405, March 1999.

[49] P. Spirtes, C. Glymour, R. Scheines, *Causation, Prediction and Search*, MIT Press, 2001.

[50] J. Cheng, D.A. Bell, W. Liu, "An algorithm for Bayesian belief network construction from data", *6th International Workshop on Artificial Intelligence and Statistics (AI & Stats-97)*, pp. 83-90, 1997.

[51] D. M. Chickering, "Learning Bayesian networks is NP-complete". In D. Fischer and H.-J. Lenz, editors, Learning from Data – Artificial Intelligence and Statistics V, Lecture Notes in Statistics, vol. 112, pp. 121–130, Springer-Verlag, 1996.

[52] D. M. Chickering, D. Geiger, D. Heckerman, "Learning Bayesian networks is NP-hard", Microsoft Research, Redmond, WA, Tech. Rep. MSR-TR-94-17, Nov. 1994.

[53] V. Terziyan and O. Vitko, "Bayesian metanetworks for modelling user preferences in mobile environment", *KI 2003: Advances in Artificial Intelligence,* Lecture Notes in Computer Science, vol. 2821, pp. 370-384, Springer-Verlag, 2003.

[54] W. Buntine, "A guide to the literature on learning probabilistic networks from data", *IEEE Trans. on Knowledge and Data Engineering*, vol. 8, no. 2, pp.195-210, April 1996.

[55] D. Geiger, D. Heckerman, "Likelihoods and parameter priors for Bayesian networks", Microsoft Research, Redmond, WA, Tech. Rep. MSR-TR-95-54, Nov. 1995.

[56] D. MacKay, "Bayesian interpolation", *Neural Computation*, vol. 4, no. 3, pp. 415-447, MIT Press, 1992.

[57] D. MacKay, "A practical Bayesian framework for backpropagation networks", *Neural Computation*, vol. 4, no. 3, pp. 448-472, MIT Press, 1992.

[58] S. Haykin, "Neural networks, a comprehensive foundation", *Prentice Hall*, 1998.

[59] D. Specht, "Probabilistic neural networks", *Neural Networks*, vol. 3, pp. 109-118, Elsevier, 1990.

[60] C. M. Bishop, *Neural networks for pattern recognition*, Clarendon Press, 1998.

[61] G. D. Garson, *Neural networks, an introductory guide for social scientists*, SAGE Publications, 1998.

[62] W. Buntine, "Learning classification trees", *Artificial Intelligence Frontiers in Statistics: AI and Statistics III*, Chapman and Hall, 1993.

[63] N. Friedman, M. Goldszmidt, "Building classifiers using Bayesian networks", *Thirteenth National Conference on Artificial Intelligence*, Portland, OR, pp. 1277-1284, AAAI Press, 1996.

[64] D. T. Pham, D. Karaboga, *Intelligent optimisation techniques: genetic algorithms, simulated annealing and neural networks*, Springer-Verlag, 2000.

[65] M. Garey, D. Johnson, *Computers and intractability: a guide to the theory of NP-completeness*, W.H. Freeman and Company, 1979.

[66] C. Papadimitriou, K. Steiglitz, *Combinatorial optimization: algorithms and complexity*, Prentice Hall, 1982.

[67] R. Pindyck, D. Rubinfeld, *Microeconomics*, Prentice Hall, 2004.

[68] V. Marbukh, "On aggregate utility maximization based network management: challenges and possible approaches", *IEEE International Conference on Communications*, vol. 2, pp. 1212-1216, June 2004.

[69] C.-S. Chang, Z. Liu, "A bandwidth-sharing theory for a large number of HTTP-like connections", *IEEE/ACM Trans. on Networking*, vol. 12, no. 5, pp. 952-692, Oct. 2004.

[70] J. Doyle, "Methodological simplicity in expert system construction: the case of judgements and reasoned assumptions", *Technical Report CMU-CS-83-114*, Department of Computer Science, Carnegie-Mellon University, 1983.

[71] L. Forgy, J. McDermott, "OPS: A domain-independent production system", in *Proceedings of the 5th International Conference on Artificial Intelligence*, pp. 933-939, William Kaufmann, 1977.

[72] E. Horvitz, J. Breese, M. Henrion, "Decision theory in expert systems and artificial intelligence", *International Journal of Approximate Reasoning*, vol. 2, pp. 247-302, 1988.

[73] D. Heckerman, E. Horvitz, "On the expressiveness of rule-based systems for reasoning under uncertainty", in *Proceedings of the Sixth National Conference on Artificial Intelligence*, AAAI Press, 1987.
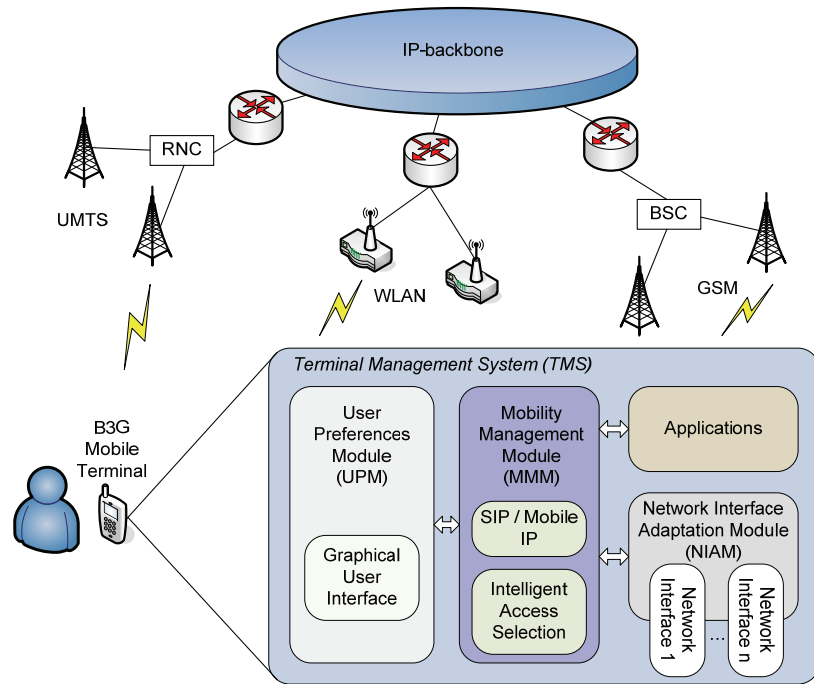
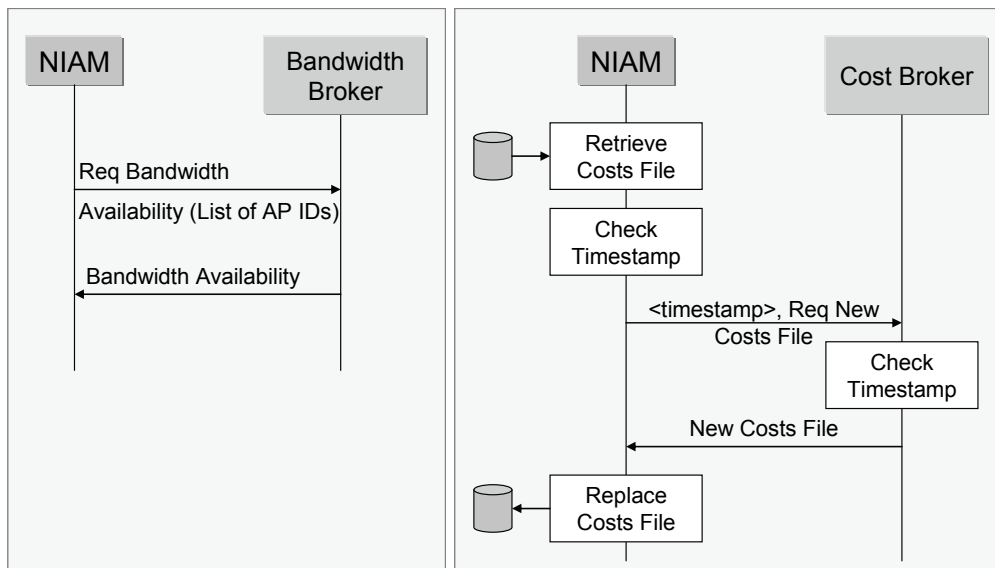Figure 1. The proposed TMS architecture in the context of a typical B3G network infrastructure

Figure 2. Information flows between the NIAM and the Bandwidth and Cost Brokers
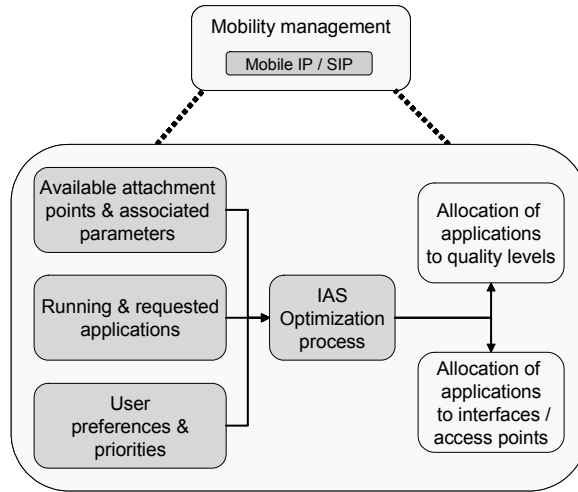
Figure 3. Mobility management at the mobile terminal
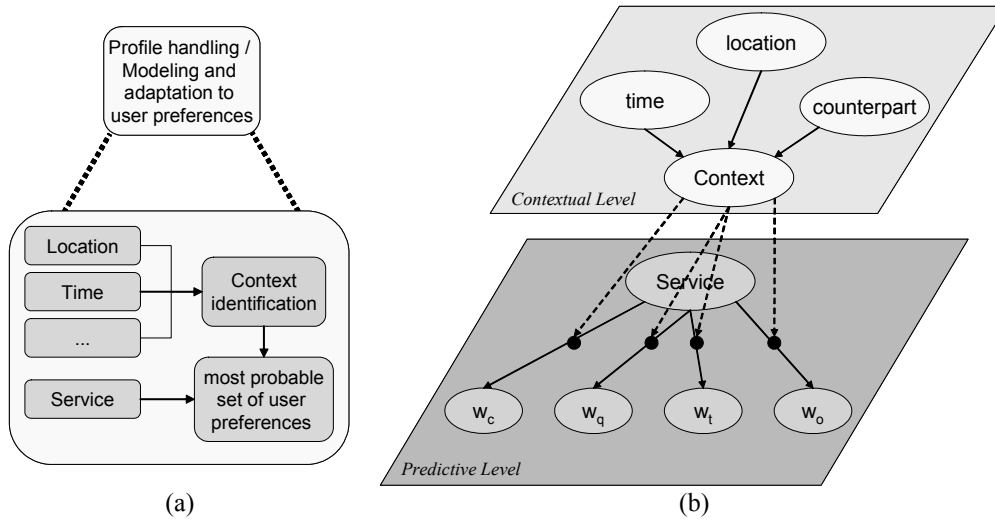

(a)   (b)

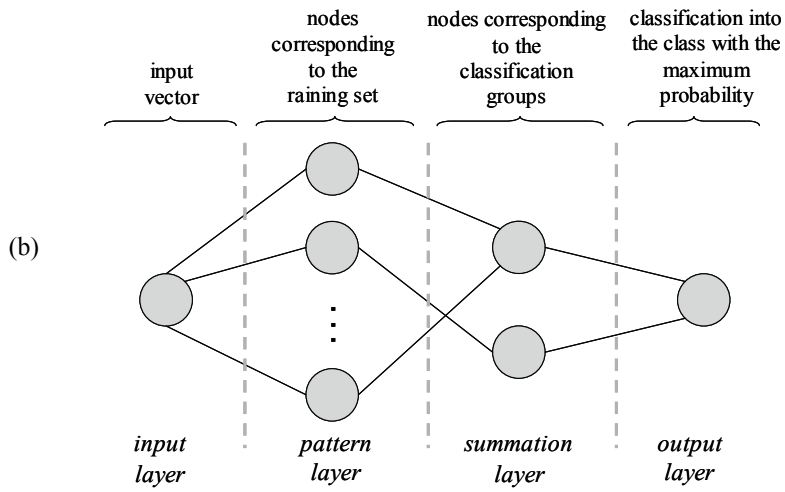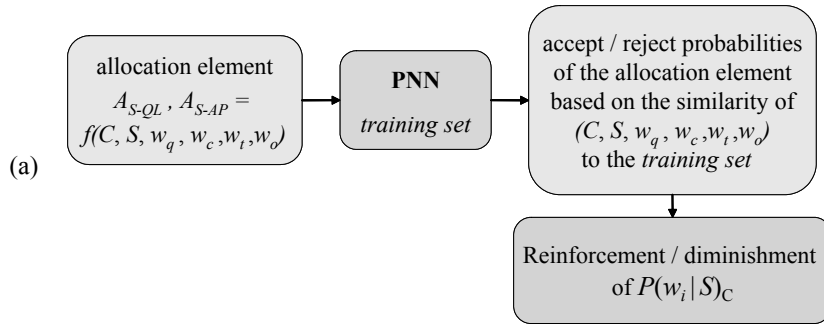Figure 4. (a) Profile handling at the mobile terminal; (b) A Bayesian meta-network for modelling user preferences

Figure 5. (a) Use of a neural network for the training of the system; (b) Probabilistic neural network architecture
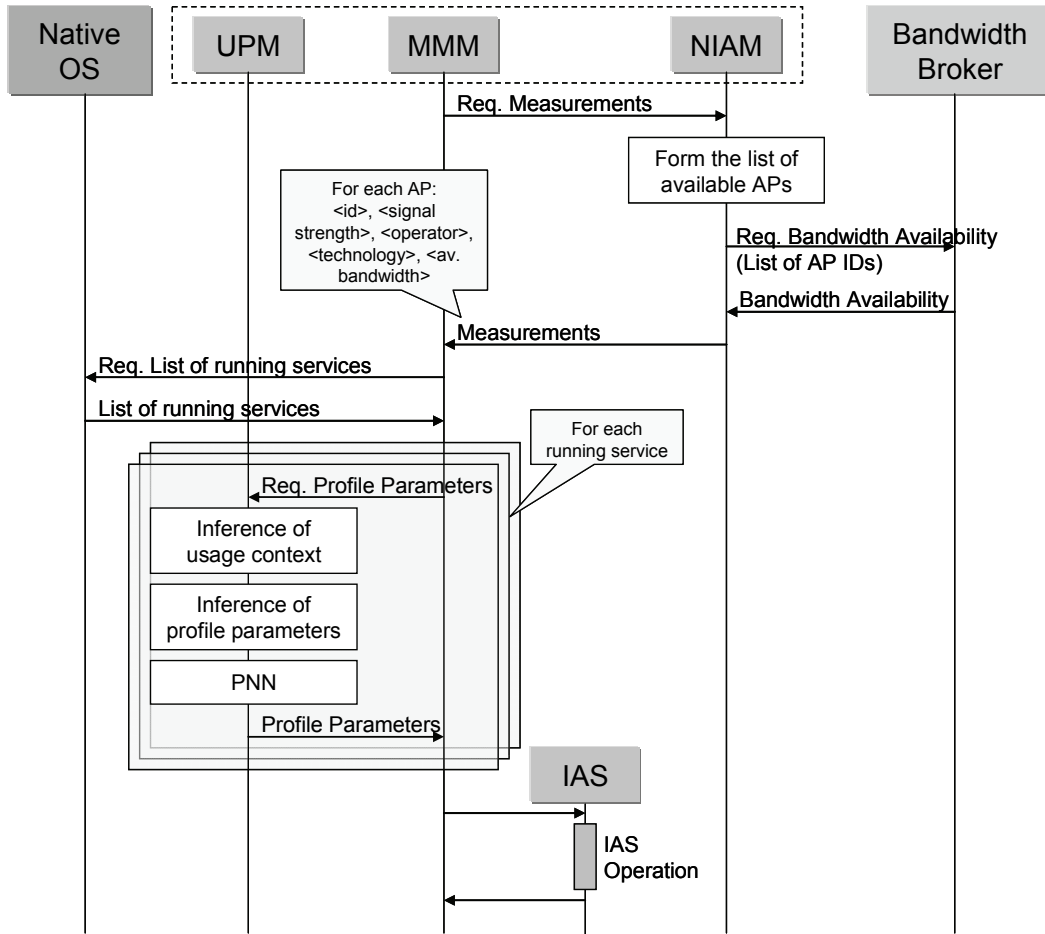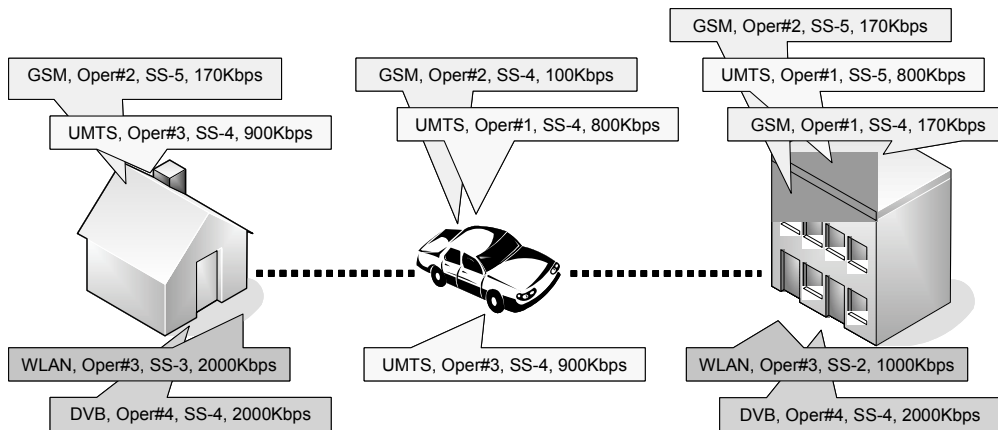
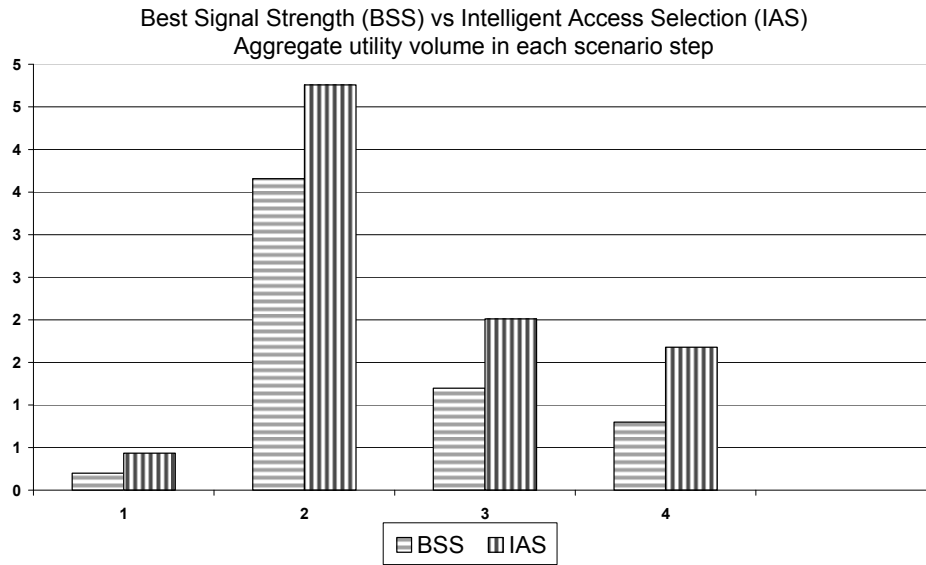Figure 6.Typical sequence diagram

Figure 7. Network coverage in scenario

Best Signal Strength (BSS) vs Intelligent Access Selection (IAS)
Aggregate utility volume in each scenario step

Figure 8. Utility computed for the voice call and video streaming services

Table 1. Example of a conditional probability table

	$S = S_1$	$S = \ldots$	$S = S_n$
$w_i = a$	$P(w_i = a \mid S = S_1)$	$\ldots$	$P(w_i = a \mid S = S_n)$
$w_i = b$	$P(w_i = b \mid S = S_1)$	$\ldots$	$P(w_i = b \mid S = S_1)$
$w_i = c$	$P(w_i = c \mid S = S_1)$	$\ldots$	$P(w_i = c \mid S = S_1)$
$w_i = d$	$P(w_i = d \mid S = S_1)$	$\ldots$	$P(w_i = d \mid S = S_1)$

Table 2. User preferences (implicit)

	Leisure context	*Business context*
General Preferences	'cost' is more important than 'quality'	'quality' is more important than 'cost'
Voice 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'

Table 3. User's answers to randomly selected questions

Context	Service	Accept				Reject			
		Q	C	O	T	Q	C	O	T
Leisure	*Voice call*	4	1	3	2	1	4	2	3
	Video streaming	4	1	2	3	2	4	1	3
	Web browsing	4	2	3	1	1	3	4	2
Business	*Voice call*	1	4	2	3	4	1	2	3
	Video streaming	2	4	1	3	3	1	4	2
	Web browsing	3	4	1	2	4	1	2	3

Table 4. Allocations selected per scenario phase

Scenario step	Service	Context	Prioritization				Allocation		
			Q	C	O	T	*Tech.*	*Oper.*	*QoS*
(a)	Web browsing	Leisure	4	2	3	1	WLAN	#3	1
(b)	Voice call	Business	1	4	3	2	UMTS	#1	5
	Web browsing	Leisure	4	2	3	1	UMTS	#1	4
(c)	Video streaming	Business	2	4	1	3	DVB	#4	5
	Voice call	Leisure	4	1	3	2	WLAN	#3	1
	Web browsing	Leisure	4	2	3	1	WLAN	#3	1
(d)	Web browsing	Business	3	4	1	2	WLAN	#3	5

Table 5. IAS and BSS service configurations in scenario

Scenario step	Service	IAS		BSS	
		AP name	QoS level	AP name	QoS level
(a)	Web Browsing	WLAN #3	1	GSM #2	1
(b)	Voice Call	UMTS #1	5	UMTS #3	5
	Web Browsing	UMTS #1	4	UMTS #3	4
(c)	Video Streaming	DVB #4	5	UMTS #1	3
	Voice Call	WLAN #3	1	UMTS #1	1
	Web Browsing	WLAN #3	1	UMTS #1	1
(d)	Web Browsing	WLAN #3	5	GSM #2	3