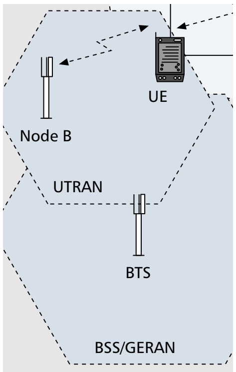
TOWARD A GENERIC "ALWAYS BEST CONNECTED" CAPABILITY IN INTEGRATED WLAN/UMTS CELLULAR MOBILE NETWORKS (AND BEYOND)

VANGELIS GAZIS, NANCY ALONISTIOTI, AND LAZAROS MERAKOS, UNIVERSITY OF ATHENS



The authors analyze the implications of the "ABC" vision in a UMTS/WLAN network context, and reveal important issues that arise. They identify major requirements, point out the limitations of current UTMS/WLAN standards from an "ABC" viewpoint, and discuss key enabling technologies and research efforts.

ABSTRACT

The next generation of mobile communications, broadly referred to as 4G, will be based on a heterogeneous infrastructure comprising different wireless (and wired) access systems in a complementary manner. 4G mobile users will enjoy seamless mobility and ubiquitous access to applications in an *always best connected* (ABC) mode that employs the most efficient combination of available access systems. The ongoing commercialization of 3G cellular mobile networks and their upcoming enhancement with WLAN radio access provides a wireless platform suitable for the introduction of "ABC" capabilities. We analyze the implications of the "ABC" vision in a UMTS/WLAN network context, and reveal important issues that arise. Further on, we identify major requirements, point out the limitations of current UTMS/WLAN standards from an ABC viewpoint, and discuss key enabling technologies and research efforts. We formulate a generic application model for an ABC capability in the interworked UMTS/WLAN architecture and analyze its complexity proving that, in principle, being always best cconnected translates to a family of NP-hard problems. To complement our analysis, we present an object-oriented design of a real-time UML model for an ABC mobile system. Finally, we summarize the advantages of our ABC model and provide directions for future work.

INTRODUCTION

The fourth generation (4G) of mobile communications tends to mean different things to different people: for some it is merely a new higher-capacity (e.g., 100 Mb/s) radio interface, while for others it is interworking of cellular and wireless LAN (WLAN) systems employing a variant of the Mobile IPv6 mobility management protocol (e.g., Hierarchical MIPv6) for intersystem handoff and Internet Engineering Task Force (IETF)-bred authentication, authorization, and accounting (AAA) technologies for seamless roaming. However, the primary distinctive feature of 4G will be a horizontal communication model, where different access systems such as cellular, satellite, WLAN-type systems, short-range connectivity, even wired systems (e.g., asymmetric digital subscriber line, ADSL) will be combined on a common platform to complement each other in an optimum way for different service requirements and radio environments, commonly referred to as *always best connected* (ABC) [1]. Recently, WLAN systems have emerged as a cost-effective infrastructure for public wireless access. From the viewpoint of mobility vs. data rate balance, WLAN systems are complementary to 3G ones. In addition, WLAN systems operate on unlicensed spectrum band, pose few engineering complexities, and are already being manufactured in high unit volumes. Consequently, the capacity increase attained through WLAN comes at a low cost, and over time the provision of affordable high-data-rate services in WLAN hot spots can effectively subsidize the amortization of 3G investments. An integrated Universal Mobile Telecommunications System (UMTS)/wireless LAN (WLAN) architecture will support higher data rates in localized service areas and, combined with global 3G roaming and intersystem mobility, provide an ideal vehicle for the introduction of ABC capability in mobile service provisioning. Standardization work is progressing rapidly to augment 3G cellular networks with WLAN radio access components. The Third Generation Partnership Project (3GPP) initiative is finalizing a 3GPP-WLAN interworking architecture within the scope of 3GPP specifications planned for 3GPP Release 6 [2]. The main goal is to complement the service portfolio of 3GPP cellular networks with WLAN radio access with minimal impact on deployed 3G infrastructures. Figure 1 illustrates the 3GPP UMTS/WLAN architecture.

Dual-mode mobile devices capable of employing their UTMS and WLAN radio interfaces simultaneously will benefit the most from the resulting enrichment of data delivery options. Proactive 3G cellular network operators and mobile terminal manufacturers are working jointly to develop WCDMA/IEEE 802.11b dual-mode handset devices for beyond 3G [3].

1536-1284/05/$20.00 © 2005 IEEE

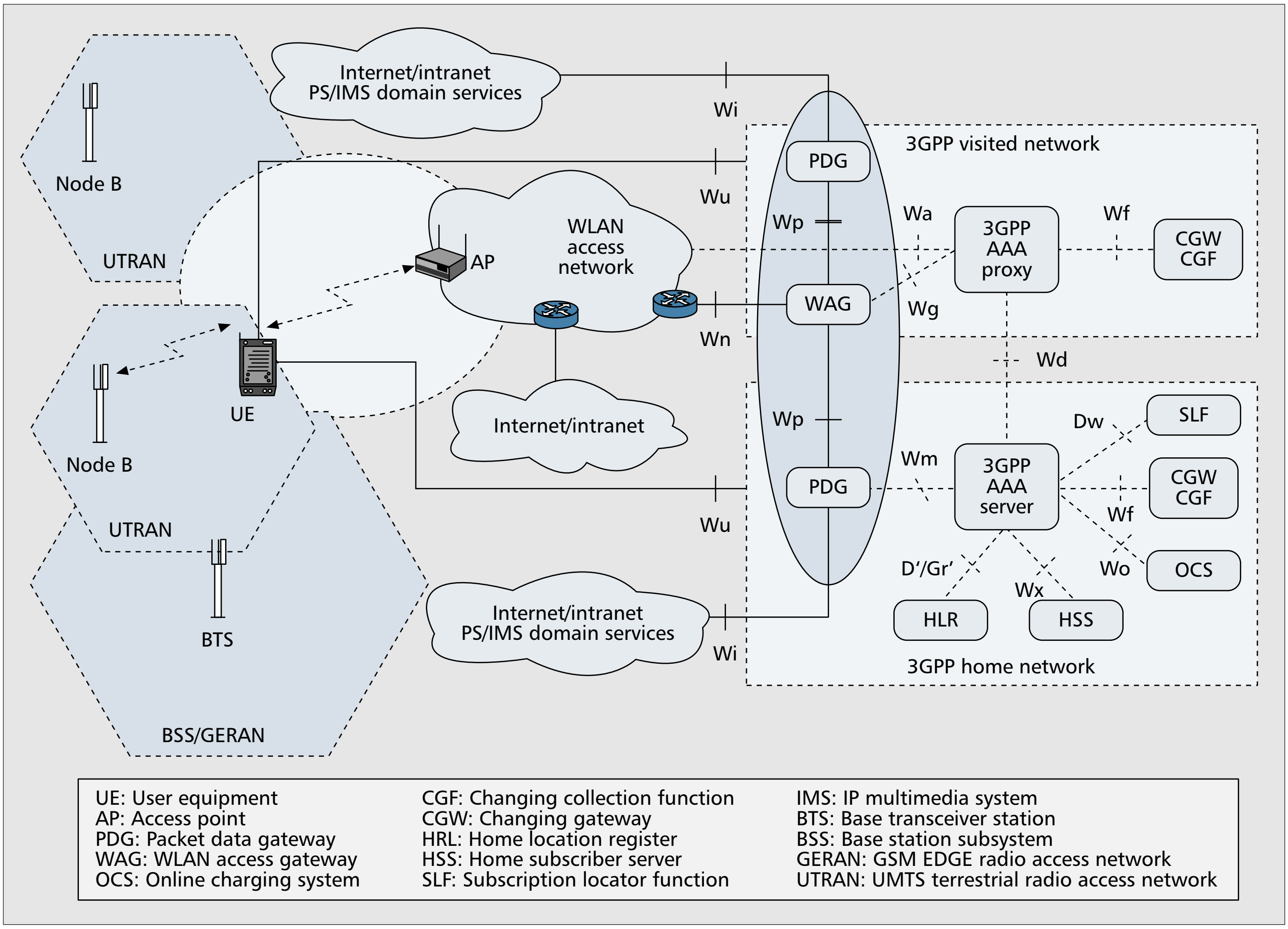


Figure 1. *The 3GPP interworked UMTS/WLAN architecture (also showing GSM/UMTS radio access.segments).*

The rest of the article is organized as follows. We discuss fundamental 4G issues and reveal the implications of the ABC concept in a WLAN/UMTS interworking context. We identify major limitations of current UMTS/WLAN standards for ABC and discuss relevant research efforts. We present UMTS/WLAN network advertisement and selection mechanisms, and point out insufficiencies related to information support for ABC decision making. We introduce our proposed ABC model and analyze its complexity, showing that, in principle, being always best connected belongs to a class of NP-hard combinatorial optimization problems. We present a developed object-oriented UML model of a decision making architecture based on our ABC formulation. Finally, we conclude the article with a qualitative evaluation of the ABC model and directions for future work.

MULTIHOMING MULTI-ACCESS CONCEPTS THAT IMPACT THE 4G SYSTEM ARCHITECTURE

MULTIPLE SERVICE DELIVERY OPTIONS

In an interworked UMTS/WLAN scenario with simultaneous UMTS and WLAN radio coverage, transporting a particular piece of information to the mobile device can be accomplished by either type of radio access network (RAN) types (i.e., UMTS RAN, UTRAN/GERAN and WLAN). One can envision plausible traffic engineering scenarios where end-to-end signaling between application endpoints is routed via UTRAN because of its reliable performance and reliability guarantees (e.g., by using UTRAN RRC acknowledged mode), while (loss-tolerant) media streams are routed via WLAN to take advantage of its greater bandwidth capacity. For an ABC capability, the final decision about which particular RAN to use must depend on a number of factors, such as the respective monetary cost of service, availability of network resources, radio link quality, service requirements, and user preferences.

PERSONALIZATION AND USER PREFERENCES

With multidimensional quality of service (QoS) applications and multiple different pricing models (e.g., volume-based, location-based, flat rate, etc.), various contextual settings and parameters may have to be resolved and employed by a decision-making algorithm that provides an optimal — if not optimum — solution to the instantaneous ABC problem of each particular user. To promote personalization in mobile service provision [4], user preferences must constitute an integral part of context. For example, let us

In a 4G mobile environment, different wireless networks will provide different grades of service, and a mobile device must choose from among a volatile set of radio interfaces the ones that better meet its instant requirements without violating its user's preferences.

consider an interworked UMTS/WLAN mobile network that provides a network-based virtual private network (VPN) capability as a value-added service over its UMTS radio access segment. Business users wishing to access corporate databases with their laptop may be willing to pay a premium price for the greater assurance level of the VPN service, while typical users browsing for Web content may be comfortable with a lower security level — and cost. User preferences are typically expressed through a continuous utility function that models the preference relation over a set of commodities [5].

USER UTILITIES AND QUALITY OF SERVICE

Users employ applications to realize various benefits as long as these (subjective) benefits are perceived to overbalance the respective charges. For the case of communication-based data applications, the perceived benefits are directly analogous to application performance, which in turn depends on the accommodation of QoS requirements for its native signaling and transport of user data (e.g., image, video, text) between peer application endpoints. From a network viewpoint, these factors translate to traffic flows with different QoS profiles that will levy different charges, depending on which access network(s) are employed for transport, thus decreasing user satisfaction. Hence, achieving an optimum performance level for communication-based applications so as to maximize user satisfaction requires honoring the QoS requirements of their traffic flows while minimizing the total charges incurred (i.e., solving a utility maximization problem). Obviously, for any given combination of application traffic flows and QoS level, user utility will decrease as the monetary cost of enjoying that combination increases. However, depending on which charging model and pricing formulae apply, monetary cost may depend on arbitrary traffic flow properties, and its lowest value may be attainable only through the combined use of multiple wireless access networks.

MOBILE DEVICE ARCHITECTURE ISSUES

A multi-access wireless context provides a wide gamut of QoS options that can be efficiently exploited through multihoming capabilities. When multiple routes for exchanging information with peers are available at a mobile device, the latter must decide which particular interfaces to employ at a given time instance. That decision will impact the routing functionality, affecting the selection of the UMTS/WLAN wireless interface for data transmission and the source address of IP datagrams. In addition, if a wireless link interface supports multiple QoS classes, some (radio-specific) QoS signaling and traffic classification (e.g., UMTS Packet Data Protocol, PDP, context signaling) must take place according to the QoS requirements of each particular application. The latter may include multiple QoS parameters (i.e., multidimensional QoS), as is the case in the UMTS QoS architecture [6]. It is a reasonable requirement that QoS-aware end-user applications should be independent of underlying network technologies (e.g., RAN architecture) and signal their QoS requirements solely through generic functional elements in the mobile device architecture that abstract the technical details of wireless technologies and enable the establishment of end-to-end application QoS.

LIMITATIONS OF CURRENT STANDARDS FOR ABC

In a 4G mobile environment, different wireless networks (e.g., UMTS/WLAN) will provide different grades of service, and a mobile device must choose from among a volatile set of radio interfaces the ones that better meet its instant requirements without violating its user's preferences. Unfortunately, the desired flexibility in routing functionality is severely inhibited by the overloading of the IP address with additional semantics (e.g., host identity semantics) that has built up over the years.

MULTI-ACCESS AND MULTIHOMING RESEARCH EFFORTS

Multihoming capabilities in the Stream Control Transmission Protocol (SCTP) allow an SCTP endpoint to negotiate the use of multiple IP addresses with its peer. However, that primarily serves redundancy purposes in the presence of link/network failures rather than versatile traffic engineering scenarios. A recent development [7] separates internetworking from host identification by introducing an additional layer between the transport and network layers in the IP protocol suite, the Host Identity Protocol (HIP). It defines the host identity (HI) name space for creating cryptographic end-to-end connection identifiers to complement standard routing identifiers (i.e., IP addresses). That decouples mobility management tasks (i.e., managing the bindings between the IP addresses and host link interfaces) from traffic engineering tasks (i.e., assigning traffic flows to link interfaces) and simplifies the simultaneous use of multiple wireless link interfaces. The MIPL proposal describes extensions to the Mobile IPv6 standard that allow mobile devices to direct a particular traffic flow to a specific link interface, taking full advantage of the diversity in wireless access technologies [8]. Through an extension of Mobile IPv6 signaling, the traffic assignment pattern may be modified dynamically according to device mobility and/or application requirements. The MIRAI project investigates mechanisms that will allow multi-mode mobile devices to efficiently use multiple access networks with dissimilar radio technologies [9]. Through an adaptation of the Mobile IP standard, a mobile device with multiple wireless interfaces (e.g., IEEE 802.11b, wideband code-division multiple access [WCDMA], cdmaOne) can multicast data over multiple radio access networks simultaneously and/or switch to a different radio access network whenever necessary. At a logical level, the MIRAI architecture distinguishes the so-called basic access network (BAN) used for transport of MIRAI signaling from the RAN used for regular data transport. At a physical level, however, a particular RAN may simultaneously assume both BAN and RAN roles.

3GPP UMTS/WLAN Network Advertisement and Selection

Network advertisement refers to the (typically unsolicited) communication of network identification information toward a mobile device to inform it of the available access networks. During network selection, the mobile device selects one of the available access networks to engage in subsequent mobility management (e.g., location registration/update) and/or session management procedures (e.g., UMTS PDP context activation). The 3GPP UMTS/WLAN interworking architecture specifies network advertisement and selection procedures for WLAN and 3GPP public land mobile network (PLMN) components.

WLAN Network Advertisement and Selection

For an IEEE 802.11 WLAN, access network advertisement is based on the WLAN network name conveyed in the service set ID (SSID) information element over the WLAN beacon channel. A WLAN mobile device discovers the SSID value using passive and/or active scanning mechanisms according to the IEEE 802.11 standard and may also solicit support for a specific SSID set via a probe request message to the WLAN access point (AP).

WLAN network selection uses three lists of SSID values: a list of available SSIDs discovered during the aforementioned advertisement procedure and two prioritized lists of preferred SSIDs — a user preferred list and a (home network) operator preferred list. The preferred SSID lists are provisioned within the mobile device USIM in the user controlled SSID and operator controlled SSID data files, respectively. During automatic network selection, the WLAN mobile device goes through the preferred SSID list in priority order until it finds an SSID value contained in the available SSID list, processing the user preferred SSID list before the operator preferred SSID list. If a matching SSID value is found, the procedure terminates and a (radio) association to the respective WLAN AP takes place. In manual network selection, the WLAN mobile device processes the aforementioned SSID lists in the same manner, presenting all matching SSID values to the user in priority order for selecting the desired one.

After establishing an association with a WLAN AP, a WLAN mobile device identifies itself by inserting its network access identifier (NAI) in an EAP Response/Identity message. Based on the NAI realm part, the WLAN access network routes EAP authentication signaling to the 3GPP home network AAA server. Should that fail but a network discovery procedure is possible, the WLAN access network conveys the list of PLMN identities it supports to the WLAN mobile device via an EAP Request/Identity message; otherwise, an EAP Failure message is sent.

PLMN Network Advertisement and Selection

In PLMN network selection, the WLAN mobile device first attempts authentication to the home PLMN. If that fails, it resorts to two prioritized lists of PLMN identities: a user preferred list and a (home network) operator preferred list typically provisioned within the mobile device USIM in the user-controlled and operator-controlled PLMN selector for WLAN access data files, respectively, and tries to authenticate with each PLMN identified therein. If all authentication attempts fail or neither of these lists is available in the USIM, it repeats the process by using another pair of prioritized lists of PLMN identities, contained in the user-controlled and operator-controlled PLMN selector with access technolog data files, respectively. If authentication to a PLMN identified therein is not possible, as a last resort the WLAN mobile device will try to authenticate with any PLMN under radio coverage.

Limitations of WLAN/UMTS Network Selection

The described procedures establish link layer associations and protocol state, and determine the set of available wireless interfaces at the UMTS/WLAN mobile device. Assuming that all possible WLAN/PLMN network identities can be ordered according to user preferences, as those are revealed when fully informed about important access-related properties of a WLAN/PLMN network (e.g., available QoS classes, charging model), the priority-based scanning mechanism yields an optimum decision of access network selection. Unfortunately, that is rarely the case. A particular WLAN/PLMN network identifier may be associated with a limited set of QoS classes or a costlier charging model than WLAN/PLMN identifiers of lower priority. Furthermore, even if a WLAN/PLMN access network did provide optimum ABC service (e.g., gratis access) at a past time instance, there is no guarantee that is still does so in the present; UMTS/WLAN signaling mechanisms pertaining to network advertisement and selection cover radio-specific mobility requirements and lower layer protocol provisioning issues, and cannot convey context-dependent or information from higher strata (e.g., pricing information) to the mobile device. Such priority-based mechanisms suffer from a fundamental design limitation: they lack an abstract model and the respective system support for representing user preferences about access network properties of interest to the user.

Limitations of WLAN/UMTS Interface Selection

According to 3GPP specifications, the selection of a suitable link interface for data traffic is an issue of mobile device implementation that should be addressed through standard IETF mechanisms. Unfortunately, the simultaneous use of multihoming capabilities in a multi-access context remains an open issue in IETF. The 3GPP packet-switched (PS) domain specifications merely suggest that some from of traffic flow template (TFT) functionality similar to the one prescribed for the gateway General Packet Radio Service (GPRS) service node (GGSN) [10] should exist in the mobile device, without any further clarifications or recommendations given.

WLAN network selection uses three lists of SSID values: a list of available SSIDs discovered during the advertisement procedure and two prioritized lists of preferred SSIDs — a user preferred list and a (home network) operator preferred list.

We regard QoS a multi-dimensional space defined over some performance metrics (e.g., bandwidth, mean delay, mean error rate, etc) that is common across all applications and interoperable to the QoS scheme of each wireless access network.

A GENERIC ALWAYS BEST CONNECTED MODEL

THE PROPOSED ALWAYS BEST CONNECTED MODEL

Our ABC model expands on earlier work in adaptive QoS management [11], which proposes a *utility model* to capture the issues of resource management in an adaptive system with multiple simultaneous sessions where the QoS of individual sessions is dynamically adapted to both available resources and user preferences, and [12], which tackles the problem of maximizing system utility by allocating a finite resource to meet the QoS requirements of applications along multiple QoS dimensions. Formulation of the ABC problem is based on the following concepts:

- Each user application requires the transport of a set of traffic flows to function.
- Each application traffic flow may be accommodated at different QoS levels.
- Each QoS level is uniquely mapped to the amount of computing and communication resources required to support that QoS level.
- Computing and communication resources are finite.
- Application utility is the sum of the utilities of all its traffic flows.
- User utility is the sum of the utilities of all its active applications.

We adopt the common assumption that each application A_i, $i = \{1, 2, \ldots, N\}$ that is active at the user mobile terminal will employ a set $F(A_i) = f_{i1}, f_{i2}, \ldots, f_{iN}\}$ of unidirectional traffic flows in communicating with its peer endpoints. We consider the set $F(A_i)$ to be dynamically volatile; that is, an application may voluntarily increase or decrease the number of its traffic flows as it desires, establishing new sessions or terminating existing ones (within resource limitations, of course). We regard QoS as a multidimensional space defined over some performance metrics (e.g., bandwidth, mean delay, mean error rate) common across all applications and interoperable with the QoS scheme of each wireless access network. These QoS dimensions will be ranked differently by different end-user applications (e.g., a multimedia streaming application will rank end-to-end delay higher than the mean error rate). Let Q denote the QoS space and q a point in the QoS space (i.e., a QoS level), while R denotes the resource space and r a point in the resource space (i.e., an amount of resources). Then $q_{ij} = \{q_1^{ij}, \ldots, q_P^{ij}\}$ is the QoS profile of flow f_{ij}, where P is the number of QoS dimensions.

As in [11, 12], we assume the existence of the following mappings:

- Quality-to-resource mapping $R(q)$: $Q \rightarrow R$, which maps from the QoS level of a traffic flow to the resource levels required to provide that quality. $R(q)$ is neither static nor deterministic nor linear and is interdependent between different traffic flows and different users.
- Quality-to-utility mapping $U(q)$: $Q \rightarrow [0,1] \subseteq \Re$, which maps from the QoS level of a traffic flow to the user utility induced by providing that quality.

In addition to [11, 12], we also assume the existence of the following mapping:

- Resource-to-cost mapping $C(r)$: $R \rightarrow C$, which maps from a specific resource level to the monetary cost of committing the resource at that particular level. We use the resource-to-cost mapping to capture the difference in the pricing model employed by each wireless access network and as a determinant of the monetary cost of each particular traffic flow.

Figure 2 illustrates the ABC problem elements.

We consider that application and user utilities result from the utilities of the individual traffic flows, as follows:

$$\text{Application } A_i \text{ utility}: U(A_i) = \sum_{j=1}^{F_i} u(q_{ij}). \qquad (1)$$

$$\text{User } u \text{ utility}: U = \sum_{i=1}^{N} U(A_i) = \sum_{i=1}^{N} \sum_{j=1}^{F_i} u(q_{ij}). \qquad (2)$$

The net user utility $u(q_{ij})$ induced by the transport of traffic flow f_{ij} at QoS level q_{ij} will be decreased by the monetary cost c_{ij} of the resources employed to guarantee the desired QoS q_{ij} for that particular traffic flow. Thus, the net utility, positive or negative, will be equal to

$$u_{ij}^{\text{net}}(f_{ij}) = u(f_{ij}) - c(f_{ij}) = u(q_{ij}) - c(q_{ij}).$$

Our formulation considers ABC any solution to the user's utility maximization problem, where the user utility is $u_{ij}^{\text{net}}(f_{ij}) = u^{\text{net}}(q_{ij})$.

KNAPSACK PROBLEMS

Knapsack problems are a family of optimization problems that require a subset of some given items to be chosen such that the corresponding profit sum is maximized without exceeding the capacity of the knapsack(s). The typical knapsack problem is formulated as follows. Consider a knapsack of capacity C and a set of N objects $O = \{o_1, \ldots, o_N\}$. Each object o_i has weight w_i and results in profit p_i when wholly included in the knapsack. If a fraction x_i, $x_i \in [0,1]$ of object o_i is placed into the knapsack, a profit of $x_i p_i$ is accrued, and the available knapsack capacity is decreased by $x_i w_i$. The objective is to obtain a filling of the knapsack that maximizes the total profit earned without overflowing the knapsack. Formally, the problem can be stated as

$$\text{Maximize } Z = \sum_{i=1}^{N} p_i x_i \ \text{ s.t. } \sum_{i=1}^{N} w_i x_i \leq C. \qquad (3)$$

Various classes of the knapsack problem exist, like the 0-1 knapsack problem, which occurs when the decision variables $x_i \in \{0,1\}$, or the multiple choice multiple dimension (MMKP) knapsack problem, which arises when there are G groups of items with exactly one item to be selected from each group for inclusion in the knapsack. A variation of the 0-1 and MMKP problems is the use of multiple knapsacks, $K = \{K_1, \ldots, K_M\}$. The generalized form of the knapsack problem is known as the generalized assignment problem (GAP), which is stated as follows: Consider a set of items $O = \{o_1, \ldots, o_N\}$, and a set of knapsacks $K = \{K_1, \ldots, K_M\}$, where knap-

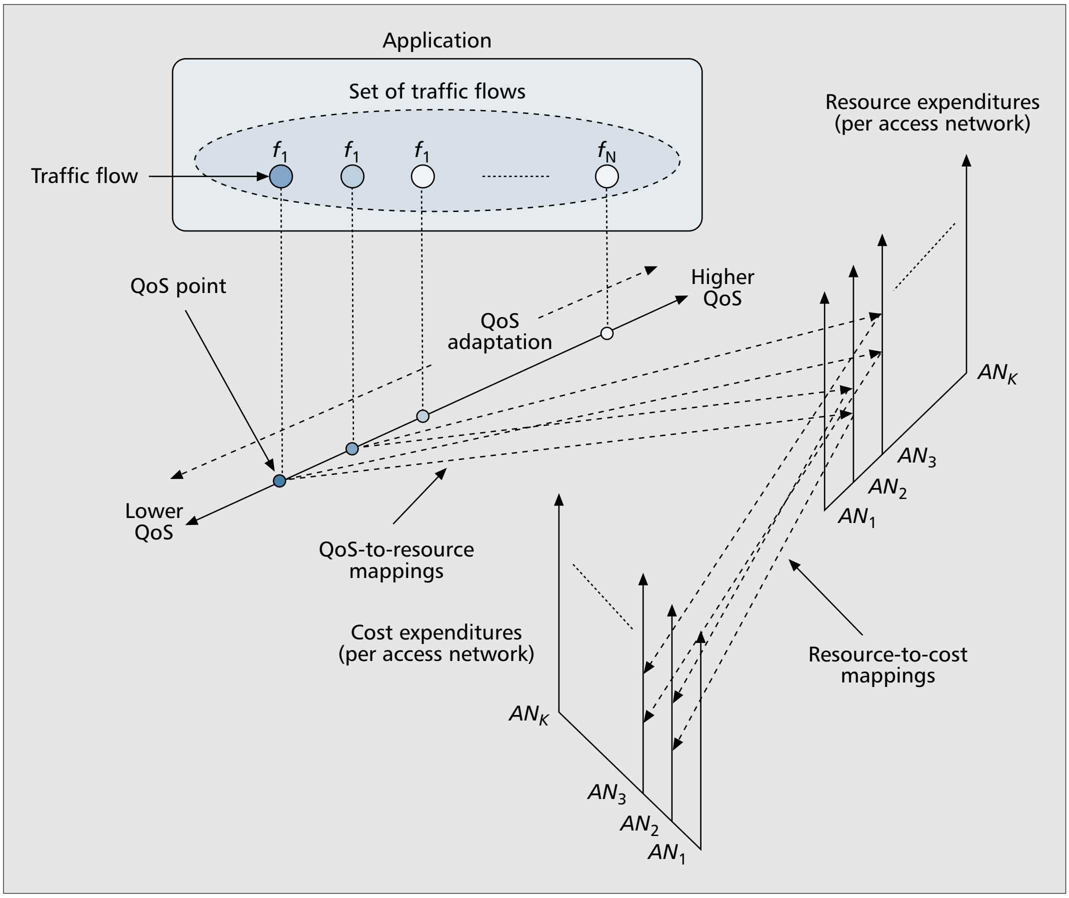


Figure 2. *The always best connected problem parameters and variables.*

Knapsack problems are a family of optimization problems that require a subset of some given items to be chosen such that the corresponding profit sum is maximized without exceeding the capacity of the knapsack(s).

sack K_j has capacity c_j. Let p_{ij} be the profit accrued if object o_i is wholly placed in knapsack K_j and weight w_{ij} the respective decrease in knapsack capacity. The GAP seeks an assignment of objects to knapsacks such that the total profit is maximized without overflowing any of the knapsack(s):

$$\begin{aligned}&\text{Maximize } Z = \sum_{i=1}^{N}\sum_{j=1}^{M} x_{ij}p_{ij}, x_{ij} \in \{0,1\}\\&\text{s.t.}\sum_{i=1}^{N} x_{ij}w_{ij} \le C_j, \forall j \in \{1, \ldots M\}.\end{aligned} \tag{4}$$

The GAP is known to be *NP*-complete, while the problem of deciding if a feasible solution (i.e., one that accommodates all items) exists is *NP*-hard [13].

ALWAYS BEST CONNECTED AS A KNAPSACK PROBLEM

Assuming for a moment that the resource-to-cost mappings are the same for all accessible wireless access networks, our ABC model can be isomorphically mapped to the MMKP with multiple knapsacks as follows:

- The number of traffic flows
  $$F = \bigcup_{i=1}^{N} F_i$$
  is mapped to the number of group of items.
- The QoS profile $q_{ij} = \{q_1^{ij}, \ldots, q_P^{ij}\}$ of each traffic flow f_{ij} is mapped to a group of items, exactly one of which will be placed in a knapsack K_j.
- Each accessible access network is mapped to a knapsack K_j of capacity C_j.
- Total user utility is mapped to the profit accrued by the items included in the knapsacks.
- The resource constraints of the ABC problem are mapped to the resource constraints (i.e., capacity) of the knapsacks.
- Choosing the QoS level is viewed as picking exactly one item from each group to include in a knapsack.

Thus, with constant resource-to-cost mappings, our ABC model maps to an MMKP knapsack problem with multiple knapsacks, which is known to be *NP*-hard [13]. Reverting back to the original formulation where the user utility induced by each traffic flow is dependent not only on the particular QoS level assigned to the traffic flow, but also on the respective cost-to-resource mapping applied by the access network chosen to transport the flow, our ABC model maps isomorphically to the GAP problem, which is also known to be *NP*-hard. Despite its exponential complexity, approximation algorithms that yield an optimal solution within a given bound distance from the optimum solution may be employed to cut down computation time [13]. We should note that because of the duality property, mobile network operators face problems of equal complexity (e.g., deciding which particular

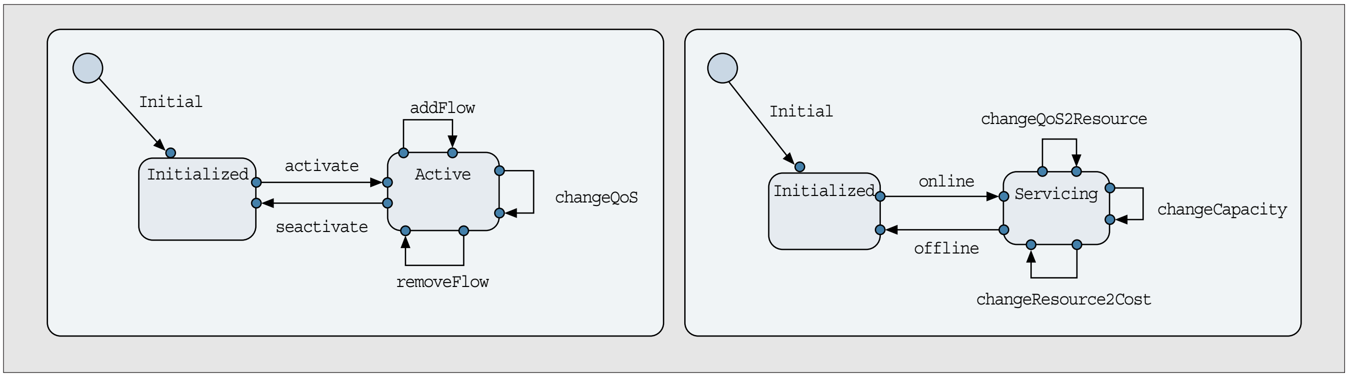


Figure 3. *State machines of the Application and Network capsules.*

flows to admit in each particular RAN with as little resource expenditure as possible.

AN OBJECT-ORIENTED ARCHITECTURE FOR ABC

To model the real-time (and distributed) aspects of ABC we employed the (nearly finalized) real-time extensions to the Object Management Group (OMG) Unified Modeling Language (UML) within an integrated development environment.

UML MODELING ELEMENTS

The main modeling element is the capsule class that possesses all the conventional properties of object-oriented class-based modeling and may additionally exhibit proactive behavior (i.e., it embodies its own state machine and thread of control that allows it to progress computing tasks independent of external stimuli). Capsules support hierarchical containment to arbitrary depths transparently to external entities (i.e., their internal structure and organization are abstracted by the containment facility). Capsule communication takes place solely through message passing over a distinct endpoint termed a *port*. Each port instance is associated with a specific pair of message sets, the latter defining its valid ingoing/outgoing messages. Each capsule instance may comprise multiple port instances that provide a communication path to other capsule instances; that is, port instances are either connected to each other in a static model configuration or connect dynamically during runtime. These features are suitable for modeling volatile environments comprising numerous dissimilar entities where system state may change unexpectedly at any time by various (possibly simultaneous) events. Interworked UMTS/WLAN wireless access networks fall right into this category, loosely classified as event-based distributed systems.

APPLICATION MODELING

The Application capsule (Fig. 3) models an application running in the UMTS/WLAN mobile device, and its attributes include the set of its end-to-end traffic flows. Application behavior is modeled in the Application capsule state machine comprising the following states:

- Initialized, representing the state where the application has completed initialization but does not yet exhibit runtime behavior (e.g., initiation of a new end-to-end traffic flow to its communication peers)
- Active, representing runtime application behavior that includes adding/removing an end-to-end traffic flow and modifying the QoS profile of its traffic flows

ACCESS NETWORK MODELING

The Network capsule (Fig. 3) models an access network and its properties according to the ABC problem formulation (i.e., its capacity and its QoS-to-resource/resource-to-cost mappings). The fluctuation of available wireless capacity (e.g., due to admission/rejection of traffic flows from other mobile devices) is provided by the Network state machine that also models changes of a radio-independent nature (e.g., changes to the resource-to-cost mapping by the network administration) in the following transitions in the servicing state:

- changeCapacity
- changeQos2Resource
- changeResource2Cost

MULTI-ACCESS WIRELESS CONTEXT MODELING

The wider multi-access wireless context is modeled through a generic container capsule class termed Context. Each Context instance may enclose an arbitrary number of Context instances, and Context subclasses include the TerminalContext capsule class and NetworkContext capsules shown in Fig. 4, which represent wireless access network and mobile device containers, respectively. The Context capsule provides a pair of port instances for generic signaling (i.e., control and data) with other Context instances. These are employed to convey terminal-specific information (e.g., the end-to-end application traffic flow set) and network-specific information (e.g., the QoS-to-resource mapping) to an interested entity such as the ABCAlgorithm capsule class described subsequently. Figure 5 presents the internal structure of the TerminalContext and NetworkContext classes, including two utility capsules called TerminalController and NetworkController that are included for modeling support (i.e., they do not model real-life entities).

ALGORITHM MODELING

An ABC algorithm is modeled through the ABCAlgorithm capsule class. Because it is desirable to support applicability of an ABC algo-

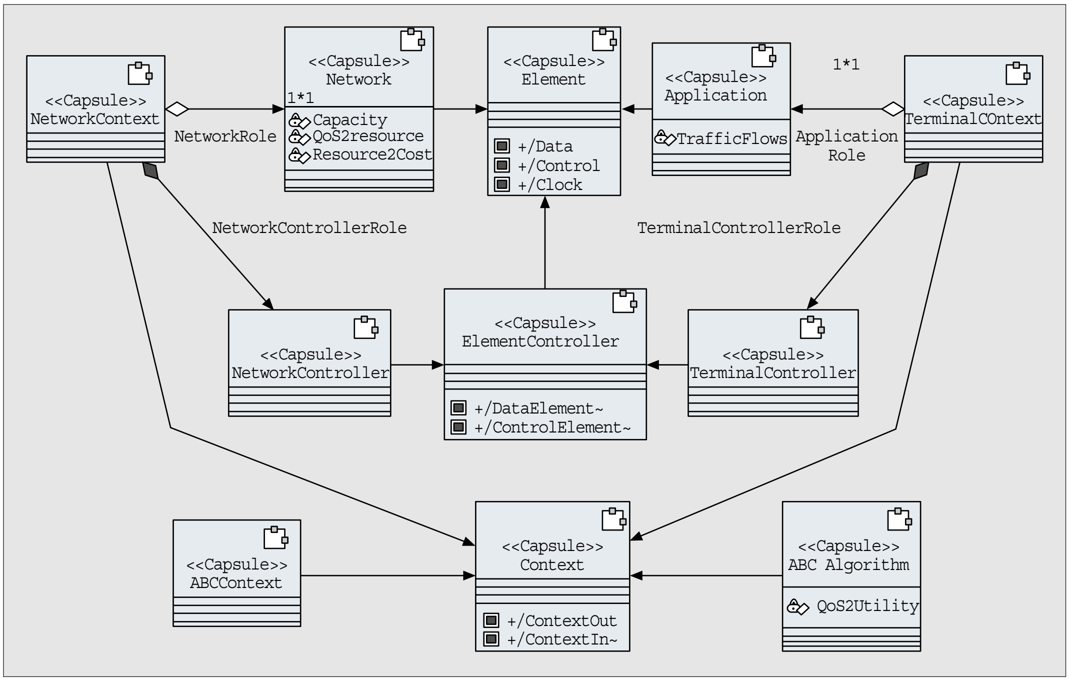


Figure 4. *Top-level class diagram of capsule classes in our ABC object-oriented UML model.*

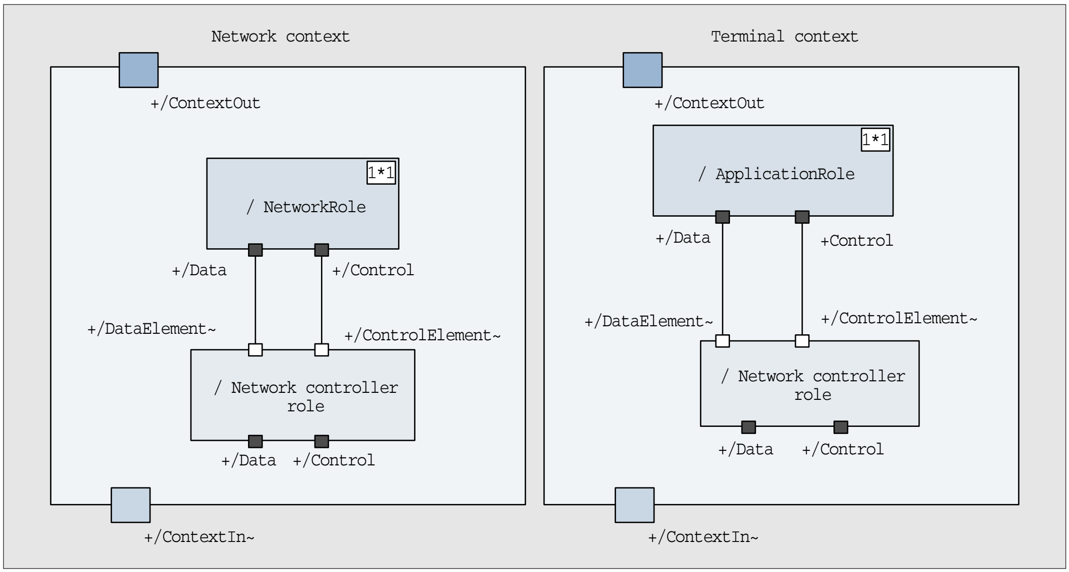


Figure 5. *The internal structure of NetworkContext and TerminalContext capsules.*

The wider multi-access wireless context is modeled through a generic container capsule class termed Context. Each Context instance may enclose an arbitrary number of Context instances, and Context subclasses include the TerminalContext capsule class and the NetworkContext capsules

rithm in as wide a variety of multi-access situations as possible, the ABCAlgorithm capsule inherits the generic Context class properties, thus becoming composable into any Context instance, like its sibling TerminalContext and NetworkContext capsule classes. The main ABCAlgorithm capsule attribute is the QoS-to-utility mapping pertaining to user preferences. A particular multi-access wireless scenario including multiple wireless access networks and multiple mobile devices can be modeled by an appropriate instantiation of TerminalContext and NetworkContext classes within an enclosing Context instance, as illustrated in Fig. 6. This may include an instance of the ABCAlgorithm capsule that undertakes decision making for the ABC problem discussed previously.

DISCUSSION OF MODELING FEATURES

Our formulation of the ABC problem captures important multi-access properties of interworked UTMS/WLAN access networks in a unified generic model:

- Application QoS adaptation in multiple dimensions
- User-centric preferences, including aversion to cost
- Resource management considerations

Although the exponential complexity may become a limiting factor in large problem sizes, in the majority of cases problem size is expected to be quite small (i.e., on the order of tenths), and the computation delay negligible. Even if a large problem size occurs, approximation algo-

Although the exponential complexity may become a limiting factor in large problem sizes, in the majority of cases, problem size is expected to be quite small (i.e., in the order of tenths), and the computation delay negligible.

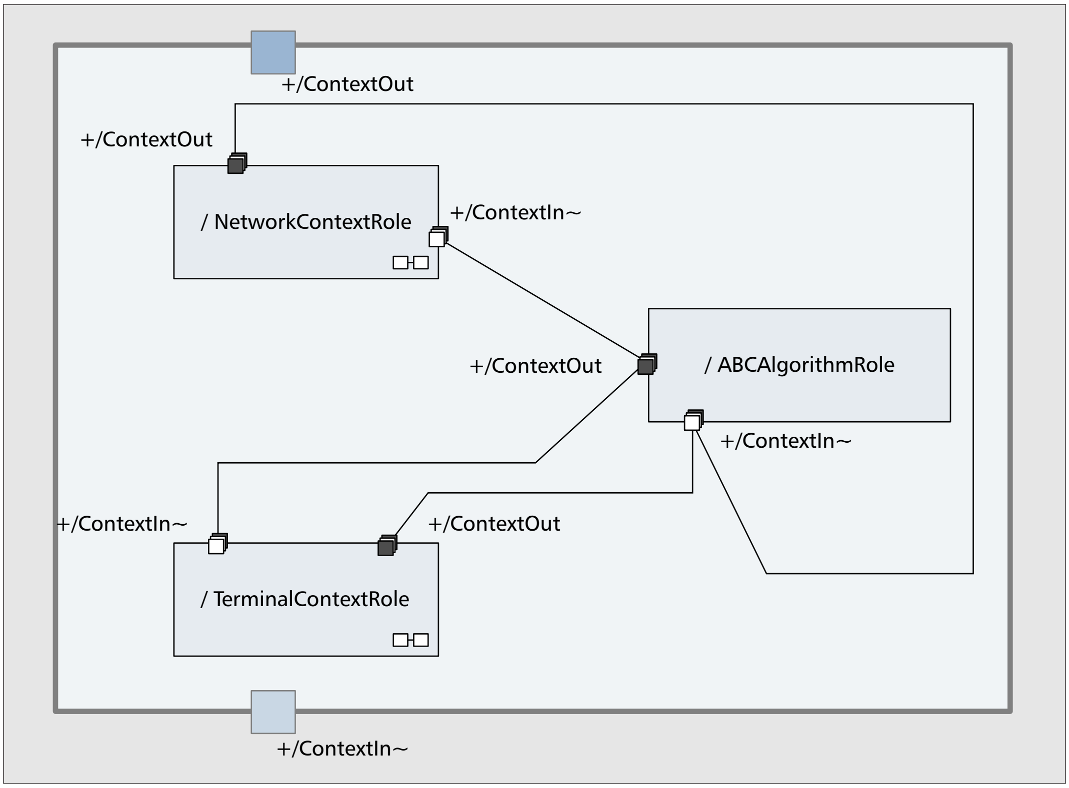


Figure 6. *Top-level structural view of our ABC object-oriented model.*

rithms may be employed to reduce the time it takes to compute an optimal solution [13].

From a modeling point of view, our ABC system model provides explicit modeling support for important parameters that inarguably affect the computation of an ABC solution in a UMTS/WLAN interworking context — and beyond. For instance, loss of radio coverage due to device mobility can be modeled through an appropriate structural manipulation of the model (e.g., by disconnecting a NetworkContext instance from the ABCAlgorithm). Other events of interest to ABC (e.g., traffic flow rate adaptation, consumption of radio bandwidth) can be readily accommodated as well. Thus, it provides an ideal instrument for studying the impact of various factors on the optimal ABC solution and its effect on the traffic distribution across multiple wireless access networks.

CONCLUSIONS AND FUTURE WORK

The integration of cellular and WLAN wireless networks will augment wide-area medium-rate cellular-based services with localized high-rate WLAN services. Because of its diversity, the resulting service set may optimally accommodate the requirements of a large customer base with disparate service requirements. The introduction of always best connected capabilities will further enhance the user experience by facilitating optimal data rate adaptation and traffic distribution over multiple UMTS/WLAN networks according to user preferences. We analyzed the ABC problem's complexity and formulated a generic utility-based mode that readily applies in an interworked UMTS/WLAN environment. We have shown that, in principle, ABC belongs to a class of NP-hard combinatorial optimization problems. We see future extensions of this work in simulation of pragmatic scenarios regarding access network selection problems in full-fledged 4G mobile environments, focusing on the effect of pricing models on the derivation of the optimum solution.

REFERENCES

[1] ITU, http://www.itu.ch/
[2] 3GPP; "Technical Specification Group Services and System Aspects; 3GPP System to Wireless Local Area Network (WLAN) Interworking; System Description," 3G TS 23.234, http://www.3gpp.org/
[3] "NTT DoCoMo Develops First Dual-Mode WCDMA/IEEE 802.11b Cellular Handset Capable of Operating in FOMATM and WLAN Sites," http://www.nttdocomo.com/presscenter/pressreleases/press/pressrelease.html?param[no]=470, July 13, 2004.
[4] V. Gazis *et al.*, "Evolving Perspectives on 4th Generation Mobile Communication Systems," *IEEE 13th Int'l. Symp. Pers., Indoor and Mobile Radio Commun.*, Lisbon, Portugal, Sept. 2002.
[5] A. Mas-Collel, M. D. Whinston, and J. R. Green, *Microeconomic Theory*, Oxford Univ. Press.
[6] 3GPP, "Technical Specification Group Services and System Aspects; Quality of Service (QoS) Concept and Architecture," 3G TS 23.107, http://www.3gpp.org/
[7] R. Moskowitz, "Host Identify Protocol," Internet draft, http://www.ietf.org/, work in progress.
[8] K. El-Malki, H. Soliman, and C. Castelluccia, "Per-Flow Movement in MIPV6," Internet draft, http://www.ietf.org/, work in progress.
[9] M. Inoue *et al.*, " Novel Out-of-Band Signaling for Interworking Between Heterogeneoous Networks," *IEEE Wireless Commun.*, vol. 11, no. 2, Apr. 2004, pp. 56–63.
[10] 3GPP, "Technical Specification Group Services and System Aspects; General Packet Radio Service (GPRS); Service Description," 3G TS 23.060, http://www.3gpp.org/
[11] S. Khan, "Quality Adaptation in a Multisession Multimedia System: Model, Algorithms and Architecture," Ph.D. thesis, Dept. of Elec. and Comp. Eng., Univ. of Victoria, 1998.

[12] C. Lee *et al.*, “On QoS Optimization with Discrete QoS Options,” *IEEE Real Time Tech. and Apps. Symp.*, 1999.
[13] D. Pisinger, “Algorithms for Knapsack problems,” Ph.D. thesis, Dept. of Comp. Sci., Univ. of Copenhagen, 1995.

BIOGRAPHIES

VANGELIS GAZIS (gazis@di.uoa.gr) received his B.Sc. and M.Sc. (communication networking) degrees from the Department of Informatics and Telecommunications at the University of Athens, Greece, in 1995 and 1998, respectively, and his M.B.A. from Athens University of Economics and Business in 2001. From 1996 until now he has been with the research staff of the Communication Networks Laboratory (CNL) at the Department of Informatics and Telecommunications, University of Athens. He has participated in national and European research projects (MOBIVAS, ANWIRE) of the IST framework program. He specializes in reconfigurable mobile systems and networks for beyond 3G, metadata and ontology languages, reflective and component middleware, adaptable services, and open API frameworks for telecommunications. He is currently pursuing a Ph.D. in the Department of Informatics and Telecommunications, University of Athens.

NANCY ALONISTIOTI (nancy@di.uoa.gr) has B.Sc. and Ph.D. degrees in informatics and telecommunications (University of Athens). She is now a lecturer in the Department of Informatics and Telecommunications of that university. She worked for seven years at the Institute of Informatics and Telecommunications of NCSR Demokritos in the areas of protocol and service design and test, mobile systems (UMTS), open architectures, and software-defined radio systems and networks. She specializes in reconfigurable mobile systems and networks for beyond 3G and adaptable services, pervasive computing, and context awareness. Moreover, she has wide experience in formal specification and testing of communication protocols and services, and design of object-oriented mobile applications. She has participated in several national and European projects, (CTS, SS#7, ACTS RAINBOW, EURESCOM etc) and was technical manager of the IST-MOBIVAS and IST-ANWIRE projects, which focused on reconfigurable mobile systems and networks and their service provisioning. She is now a member of the PMT and WP, leader of the FP6 IST-E2R project on reconfigurability, and she also participates as University of Athens technical manager in the IST-LIAISON project focusing on location-based services in working environments. She is co-editor and author of *Software Defined Radio, Architectures, Systems and Functions* (Wiley, 2003). She is a TPC member for many conferences in the areas of mobile communications and mobile applications for systems and networks beyond 3G. She has over 55 publications in the areas of mobile communications and reconfigurable systems and networks.

LAZAROS MERAKOS (merakos@di.uoa.gr) received a Diploma in electrical and mechanical engineering from National Technical University of Athens, Greece, in 1978, and M.S. and Ph.D. degrees in electrical engineering from State University of New York, Buffalo, in 1981 and 1984, respectively. From 1983 to 1986 he was on the faculty of the Electrical Engineering and Computer Science Department, University of Connecticut, Storrs. From 1986 to 1994 he was on the faculty of the Electrical and Computer Engineering Department, Northeastern University, Boston, Massachusetts. During the period 1993–1994, he served as director of the Communications and Digital Processing Research Center, Northeastern University. During the summers of 1990 and 1991, he was a visiting scientist at the IBM T. J. Watson Research Center, Yorktown Heights, New York. In 1994 he joined the faculty of the University of Athens, where he is presently a professor in the Department of Informatics and Telecommunications, and director of the Communication Networks Laboratory (UoA-CNL) and the Networks Operations and Management Center. Since 1995 he has led the research activities of UoA-CNL in the area of mobile communications, in the framework of the Advanced Communication Technologies and Services (ACTS) and Information Society Technologies (IST) programs funded by the European Union (projects RAINBOW, Magic WAND, WINE, MOBIVAS, POLOS, ANWIRE). His research interests are in the design and performance analysis of broadband networks, and wireless/mobile communication systems and services. He has authored more than 150 papers in the above areas. He is Chairman of the Board of the Greek Universities Network, the Greek Schools Network, and a member of the Board of the Greek Research Network. In 1994 he received the Guanella Award for the Best Paper presented at the International Zurich Seminar on Mobile Communications.

From a modeling point of view, our ABC system model provides explicit modeling support for important parameters that inarguably affect the computation of an ABC solution in a UMTS/WLAN interworking context and beyond.