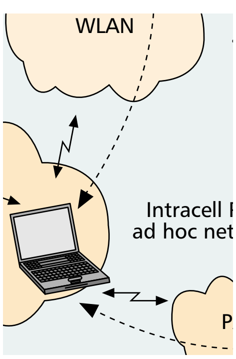
TRANSPARENT IP RADIO ACCESS FOR NEXT-GENERATION MOBILE NETWORKS

DAVE WISELY, BTEXACT
HAMID AGHVAMI, KINGS COLLEGE LONDON
SAMUEL LOUIS GWYN, LUCENT TECHNOLOGIES
THEODORE ZAHARIADIS, ELLEMEDIA TECHNOLOGIES
JUKKA MANNER, UNIVERSITY OF HELSINKI
VANGELIS GAZIS, NIKOS HOUSSOS, AND NANCY ALONISTIOTI, UNIVERSITY OF ATHENS



Advances in the network architecture, enhancements in the signaling protocols, provisioning of end-to-end QoS, world-wide seamless mobility, and flexible service provision are among the major research challenges towards next-generation wireless networks.

ABSTRACT

Advances in network architecture, enhancements in signaling protocols, provisioning of end-to-end QoS, worldwide seamless mobility, and flexible service provision are among the major research challenges toward next-generation wireless networks. The integration and interoperability of all these technologies, along with new truly broadband wireless innovations and intelligent user-oriented services will lead toward the so-called 4G wireless networks. In this article we identify the key issues of an innovative transparent IP radio access system that targets 4G networks.

INTRODUCTION

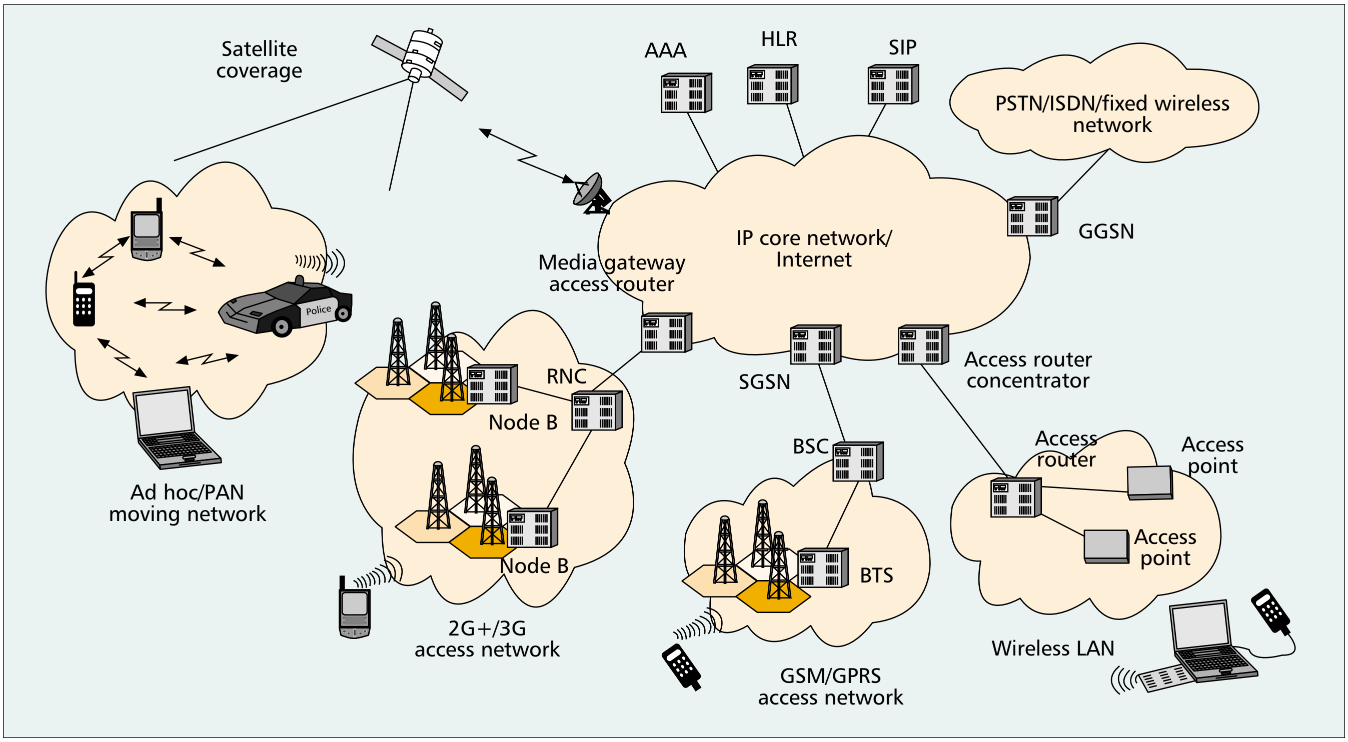
With third-generation (3G) mobile networks already launched in Japan and initial deployment just beginning in Europe, it is an appropriate time to look beyond 3G and ask what the next major development in mobile will be. It is possible [1] to take the view that the progression through the "generations" is simply about new air interfaces and higher bandwidths. There is support in both Japan and the International Telecommunication Union —Radiocommunication Standardization Sector (ITU-R) for this as a definition of fourth-generation (4G) mobile as, potentially, a 100 Mb/s air interface. An alternative view [2] is that there will be an IP core network, through which all traffic will flow, and a large number of access technologies, both fixed and wireless, will provide connectivity. Concrete examples of these include digital subscriber line (DSL) and wireless LAN technologies. Taking the example of the United Kingdom, it is estimated that there will be 5 million DSL connections and 4000 WLAN hotspots by 2005.

IP has fundamentally changed the nature of communications. It has become the de facto standard for data transmission and, with developments in voice over IP (VoIP) such as Universal Mobile Telecommunications System (UMTS) Release 5, is rapidly being used for all applications. In part the impact of IP has been economic, breaking apart the traditionally integrated fixed/mobile operator's value chain and introducing actors like Internet service providers (ISPs) and content providers. One of the reasons for this is that IP networks only transmit packets — services and intelligence are located at the network edge — and the primary differentiator in today's IP networks is bandwidth. IP service creation is not integrated with network service. The economic issues raised by this are illustrated by the Japanese Imode system: users can either visit content within a walled garden of approved sites, or simply connect to the wider network.

The development of terminals, driven by a Moore's law timescale, means that users will soon have much more choice in how they connect to networks and obtain services (in the widest sense). Smaller laptops, PDAs, and tablet PCs are converging with ever smarter mobile phones. The low cost of Bluetooth and 802.11a/b WLANs means that these will soon be incorporated in a majority of these devices. Many commentators believe that ad hoc and personal area networks (PANs) will become commonplace as users create networks within their homes, share connections, and form closed user groups. In addition, universities, enterprises, and hotspot owners are now able to deploy public access WLANs across most of Europe; this activity will greatly increase when 802.11h equipment accepted for use in the 5 GHz bands in Europe is available.

All these developments pose a fundamental

1536-1284/03/$17.00 © 2003 IEEE



■ Figure 1. *Open all-IP network architecture.*

challenge to existing fixed and mobile telecom operators. Users, faced with a range of access technologies, an overcapacity of fiber backbone connectivity, and services available from multiple providers on the Internet, might well opt for the cheapest "bit carrier." Optimal "anywhere" service provision, however, cannot be accomplished in the absence of a number of compelling functions that can, when integrated into a package, retain much more of the value chain for the network operator [3]:

- Flexible and dynamic service creation and discovery
- Network reconfigurability management
- Personalization and context awareness
- Adaptation of services to access technologies and terminals
- Seamless access
- End-to end quality of service (QoS)
- Support for new network topologies such as PANs and ad hoc networks
- Billing and accounting services

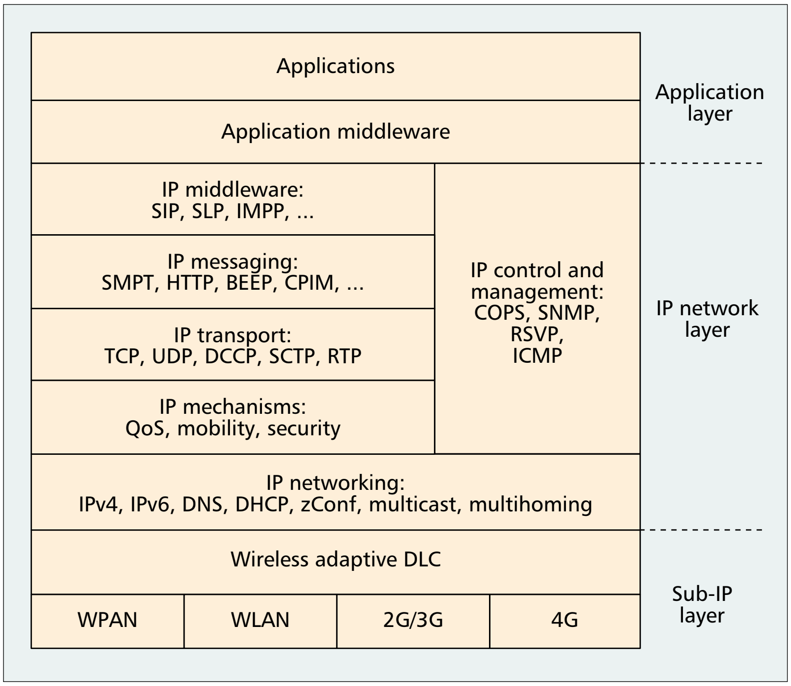
In addition to these enhanced network functions, it is important for network operators to lower costs of IP packet delivery, especially for real-time services such as voice, since this is the ultimate purpose of the network. Transmitting IP packets over wireless links presents many problems, and current solutions are far from optimized. As an example, the common 802.11b WLAN technology is wholly unsuited to providing a cellular voice service; the capacity for voice-only services drops to less than 10 percent of that for data, there is no coordination with layer 3 QoS signaling, encryption is inadequate, and a cellular WLAN network suffers high levels of interference [4].

We believe the integration and interoperability of all these diverse technologies, along with new truly broadband wireless innovations and intelligent user-oriented services will lead toward a new generation of heterogeneous wireless networks, which has already been called 4G. In this article we propose an innovative transparent IP radio access system that targets 4G networks. The article is structured as follows. We describe a heterogeneous open network architecture that is able to fulfill potential future service requirements in a cost-effective way. We identify the trends and research areas in the layers below the IP layer. We address two critical issues in future mobile networks: user mobility and end-to-end QoS. The first is the ability of the user to seamlessly roam in the network, while keeping communication sessions uninterrupted, while the second is the network ability to guarantee the agreed end-to-end quality of the provided services. Users, however, do not care about the technological limitations, the restrictions of the wireless medium, and the heterogeneity of the network. Therefore, an intelligent middleware is described that aims to hide the network and technology details from the end user and provide truly transparent IP radio access.

Open All-IP Network Architecture

By definition it is difficult to make predictions and foresee which network systems and services will succeed in the forthcoming 10 years. However, one thing that can be expected is that since network operators have already made huge investments in network infrastructure, 4G networks are expected to integrate all heterogeneous wired and wireless networks, and provide seamless worldwide mobility [5].

An abstract view of the envisaged network



■ **Figure 2.** *The proposed IP-based protocol stack.*

architecture is shown in Fig. 1. The 4G core IP network will be dominated by connectionless IP packet switching technologies and interface via access routers with multiple wireless access technologies like indoor (e.g., IEEE 802.11a/b, HiperLAN/2, Bluetooth) and cellular 3G/4G networks. Gateway access routers will interface the network with a multitude of wired and air interfaces, inherited from earlier-generation communication systems like the public switched telephone network/integrated services digital network (PSTN/ISDN) and 2G/2G+ technologies: Global System for Mobile Communications (GSM), Digital Cellular Service (DCS), General Packet Radio Service (GPRS), and code-division multiple access (CDMA). Unified mobility management will enable common handing of customer profiles and Internet addresses, user location, and authorization, authentication, and accounting (AAA) functions, facilitating seamless handover between multiple operators' wireless networks.

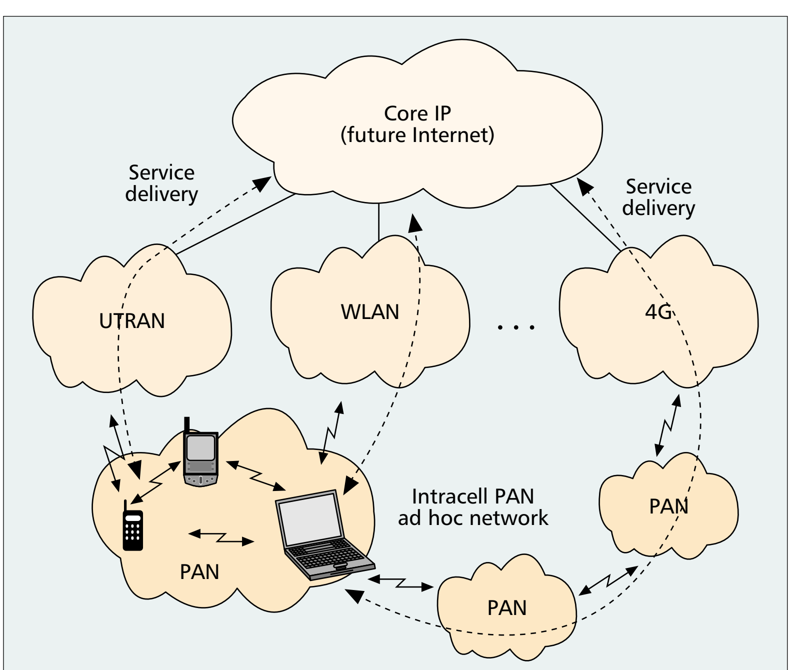
Distributed ad hoc, spontaneous, and in many cases moving networks, deployed in the unlicensed bands, will instantaneously provide wireless personal area networks (PANs). A PAN network may include any collection of devices that belong to or are carried by a networked user (e.g., cell phone, laptop, earphones, GPS navigator, Palm Pilot, beeper) and form his/her personal "PAN bubble." The bubble can expand or shrink dynamically depending on a user's environment and needs. Such access is important when the mobile user enters a new location and aims to quickly sense and control the environment (e.g., gain access/connectivity, control the temperature, adjust the lighting) or be recognized by the environment (e.g., welcome message, uninterrupted communication, automatic selection of background music).

Apart from nodes, mobility is also foreseen for entire subnetworks, which change their point of attachment to the global Internet. An example of a moving network is the PAN of a user who is inside a moving car, or is a passenger on a train or ship that connects his/her PAN to the Internet via the company's private network. Moreover, mobile proxy/ad hoc access points (e.g., the car in Fig. 1) may provide bridging operations to devices that belong to the PAN, but may not have the means to directly communicate with the rest of the IP network. Combining the notion of mobile or moving networks with the security and privacy provided by virtual private networks (VPNs), a new type of service can be introduced, called Mobile VPNs.

In order to support this heterogeneous multifunctional network architecture, the IP-based protocol stack of Fig. 2 is proposed. At the physical layer, an innovative multitechnology, air interface is assumed, able to provide connectivity to existing cellular (2G/3G), WLAN (802.11 a/b/h, HiperLAN/2), wireless PAN (Bluetooth) and future (4G) wireless networks. Apart from adaptability in the physical layer, the proposed wireless adaptive data link control (DLC) will provide mechanisms to interface to the network technology chosen and to provide information to the upper layers. Over the DLC, various IP protocols will provide the means of network communications including transportation, signaling, messaging, QoS, mobility, security, control, and management. Finally, a transparent middleware layer will provide for an abstract view of the underlying mobile network infrastructure and facilitate focused application development.

SUB-IP INNOVATIONS

The aim of future radio access is to ensure the availability of a transparent wireless transport and delivery mechanism that permits the distribution and creation of services in an IP-like



■ Figure 3. *Macro service delivery via direct and indirect wireless access.*

manner, with reliability matching that of wireline networks. Moreover, its architecture must ensure that the cost of wireless data delivery is effective regardless of the access network used. The absolute need for cost effectiveness will ensure that the environment for attractive well priced data services is at least in place and that competitive advantages to an operator can be maintained while allowing the end user a rich set of competitive data services.

In order to achieve cost effectiveness, future wireless infrastructures have to integrate transparently multiple heterogeneous wired and wireless radio access networks that have already been deployed, supported by a full IP-based core network. As shown in Fig. 2, 2G/3G+ cellular, WLAN, PAN, and ad hoc PAN access networks will comprise this heterogeneous environment.

In the envisaged scenario, the traditional (ad hoc) PAN network concept is extended from a composition of devices that form ad hoc connections between themselves to a composition of devices that have cellular and wireless access capabilities. In this respect, the PAN takes on the additional property of virtual terminal equipment(VTE), and the user may be connected, simultaneously or in a sequential manner, to the core network via a variety of wireless access mechanisms, while maintaining intracell connectivity with other users as a participant node of an ad hoc network. This connectivity richness of the extended VTE concept may further enhance the services exploitation potential. For example (Fig. 3), because of the ad hoc capability of the PAN, an individual user may be receiving a service directly via the individual's PAN VTE connection to a wireless access medium, but also via an ad hoc connection through a number of adjacent PAN VTEs over which the user has temporary access rights. In this sense the PAN VTE is also acting as a proxy and ad hoc access point (PA-AP) to the wireless access network. In effect the PA-AP acts as a proxy gateway for the wireless access network so that the service layer is able to share, receive, and create mobile services. The characteristics of these networks are given in Table 1.

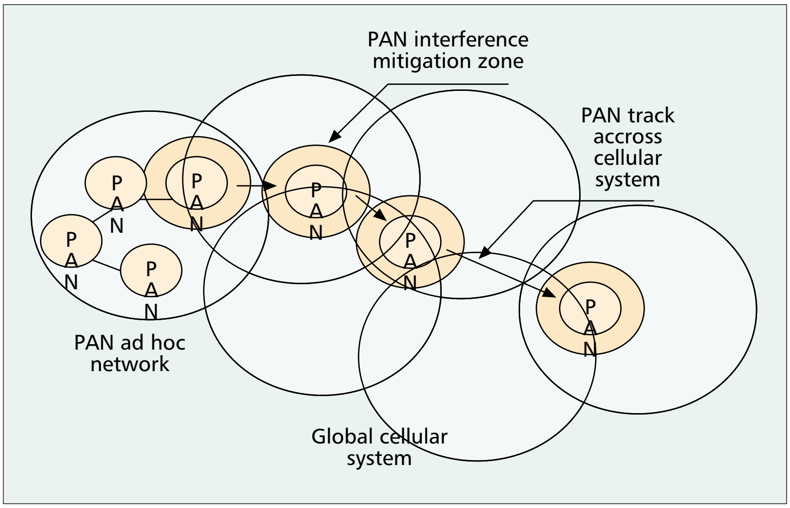
In the wireless service delivery example of Fig. 4, the wireless access network treats the VTE PAN as a single terminal entity. In this environment, the research challenge is to ensure the maintenance of sufficient bandwidth, with low latency, to the PAN VTE via a variety of access mechanisms, while the PAN VTE acts as PA-AP for macro service delivery giving the impression of wire-like quality. The challenge from the mobile PAN environment is to provide enough interference mitigation strategies, so that the internal PAN VTE and PA-AN environments can be maintained flexibly. An example of this challenge is to have macro service delivery achieved by a variety of enhanced cellular systems either simultaneously or sequentially.

A key feature to achieve this wireless service delivery environment lies in the flexibility it provides to the higher layers so that transparency can be achieved. This flexibility is attained via the application of polymorphic concepts to the radio access layers. In this regard, the aim is to reduce the complexity of the system to the higher layers using techniques that apply the invocation of multiple methods via a single interface; these methods are bound to a given air interface at runtime. It is this single interface, multiple methods, and late binding techniques that give intelligence to the radio access layers and their polymorphism.

Apart from adaptability in the physical layer, it is proposed that wireless DLC will have mechanisms to interface to the network technology

Name	Definition and characteristics
Ad hoc PAN	No infrastructure (private network) • Dynamic topologies • Inter-PAN connection (device to device of different PANs using the same interface) • Inter-PAN gateways (different standard relay)
Cellular	Infrastructure-based • Operator-owned (public network) • Large range • Low/medium data rates
PA-AP	Proxy and ad hoc access point • This is a VTE element that can form *ad hoc* connections between VTEs
PAN	No infrastructure (private network) • Virtual terminal consisting of various devices, all possibly with at least one access interface • Low range • High data rates • Dynamic topologies • Any number of gateways to infrastructure-based networks
VTE	Virtual terminal equipment • From a cellular perspective • Forms a single manageable entity • Composed of PAN wireless access elements
WLAN	Infrastructure-based • Not necessarily operator-owned (public/private network) • Low range • High data rates

■ **Table 1.** *Definitions and characteristics.*



■ **Figure 4.** *A wireless service delivery environment.*

IP Micromobility protocols reduce the round-trip delay and eliminate the end-to-end registration signaling overhead, by handling local movements of the mobile node transparent from the global Mobile IP network.

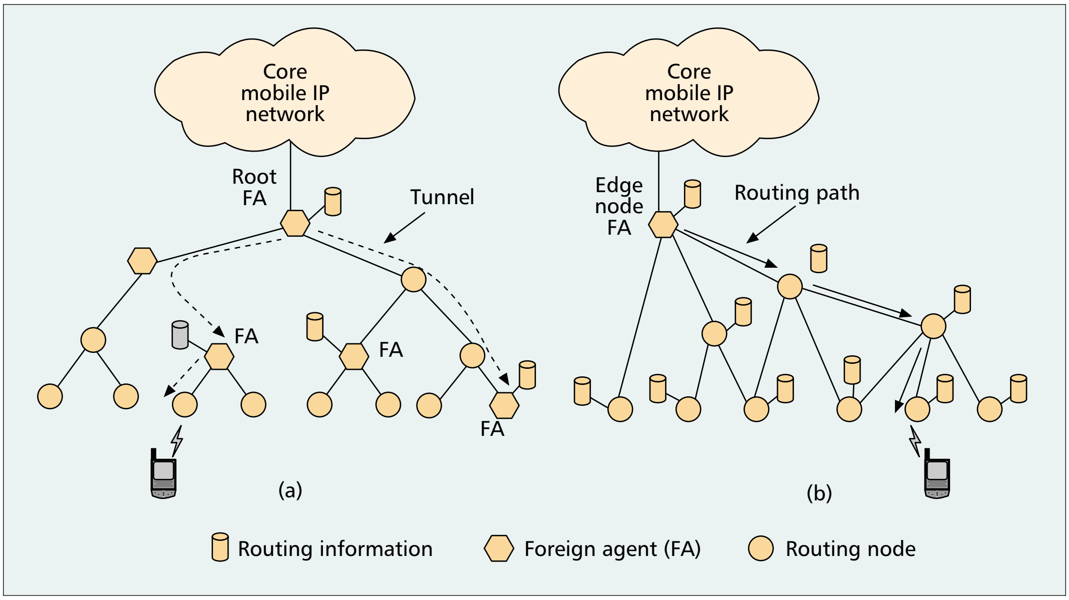


Figure 5. *a) Hierarchical tunneling and b) explicit terminal routing approaches.*

chosen and provide information to the upper layers. Moreover, in cooperation with the middleware layer, it will contain procedures for mapping several control issues of the network to DLC control information. These procedures may refer to special messages such as connection setup or connection release at the network level, which have to be translated to the DLC level. The DLC layer will support QoS over the wireless link for each connection. Thus, the parameters of the network connection have to be transferred to the DLC level to serve the DLC connection accordingly.

Mobility Management

Next-generation mobile networks should provide mechanisms that enable seamless roaming between multiple heterogeneous subnetworks [6]. In more detail, 4G wireless networks should provide:

Personal mobility. This is the ability of the user to access his/her personalized network services while away from his/her home network. The proposed framework aims to provide user mobility based on an intelligent end-to-end location-aware middleware layer.

Terminal mobility. This is the ability of the network to maintain continuous services even when the user's terminal changes locations. Terminal mobility becomes more challenging for streaming applications with strict QoS requirements and devices with multiple heterogeneous radio access technologies. Handing over multihomed terminal mobility will require intertechnology handovers, and resource management initiated by layer 2 triggers and supported by enhanced signaling protocols.

Network mobility is the ability of the network to support roaming of an entire subnetwork, structured or ad hoc. In this case the core and access networks should be able to recognize the moving network as a special roaming node.

Mobile IP, which initially addressed the mobility management problem, introduces significant network overhead in terms of:

- Increased delay due to triangular routing
- Packet loss due to registration delays
- Increased signaling due to frequent terminal registration

Route optimization can improve performance, but cannot provide acceptable service quality, especially for streaming applications like VoIP. The network overhead is drastically increased if mobile hosts frequently change their point of attachment. Based on the assumption that the majority of frequent handovers take place between nodes of the same regional network, many IP micromobility protocols have been proposed aimed to provide fast seamless local mobility. IP micromobility protocols reduce the round-trip delay and eliminate the end-to-end registration signaling overhead by handling local movements of the mobile node transparent from the global Mobile IP network. This is achieved by maintaining regional location databases that map mobile node identifiers to location information, which is used for local delivery of the packets. In order to further reduce registration signaling, IP paging techniques are introduced. Just like cellular networks, micromobility protocols take into account the operational mode of the mobile node and maintain approximate location information for the idle users. In this way, intermediate signaling messages generated by frequent local movements of an idle node are suppressed, with the cost of increased delay when a new call is initiated. In general, micromobility management protocols [7] may be categorized into two groups (Fig. 5).

The hierarchical tunneling approach: The protocols that fall in this category incorporate a number of foreign agents (FAs) in a tree-like structure. Each FA maintains part of a location database. The entries of the database contain the address of the next FA in the path toward the mobile node. Data packets that address a specific terminal are delivered encapsulated to the root FA of the visiting tree network. Each FA in the lower layers of the tree structure that is the endpoint of a tunnel has to decapsulate the packets, search its location database for a corresponding entry, and then re-encapsulate and forward them through another tunnel down to the terminal's point of attachment. When the terminal moves, it propagates location update messages up to a tunnel endpoint, which is located in both the old and new paths. Examples of micromobility protocols that use the hierarchical tunneling approach are Hierarchical Mobile IP, Hierarchical Mobile IPv6, and the Brain Candidate Mobility Protocol (BCMP). Hierarchical tunneling micromobility protocols may have to introduce new control/signaling messages or require a hierarchical tree structure.

The explicit terminal routing approach: The protocols that fall in this category avoid the overhead introduced by encapsulation and tunneling, with the cost of introducing explicit routing entries for each terminal. Instead of location database, each node keeps a (kind of) routing table that contains an index to the next hop node in the path towards the terminal's point of attachment. When a terminal moves, location update information is propagated explicitly (via signaling messages) or implicitly (via data message snooping) up to an anchor or crossover node, which is located in both the old and new paths or up to the root FA. Examples of micromobility protocols that use the explicit terminal routing approach are Cellular IP, Handoff Aware Wireless Access Internet Infrastructure (HAWAII), and Unified Wireless Access (UniWA). Also in this category fall protocols like "iwander," which are based on Ethernet switching and layer 2 MAC to port number binding. These protocols do not require a hierarchical network structure, but may face scaling limitations due to the increased amount of routing information especially at nodes closer to the edge node.

In the last few years, there has been considerable debate in IETF on suitable fast and seamless handoff extensions. However, the 4G network architecture proposes enhancements in the existing signaling and transport protocols. In the forthcoming years, special focus is expected on many research areas, including:

- Scaling over large networks (e.g., large number of terminals, large number of cells per access router)
- Signaling overhead vs. response delay time (layer 2 vs. layer 3 signaling vs. control protocols)
- Location update vs. IP paging
- Roaming information (terminal velocity/direction/moving model/profile vs. open session requirements/contract vs. network resources/congestion)

End-to-end QoS

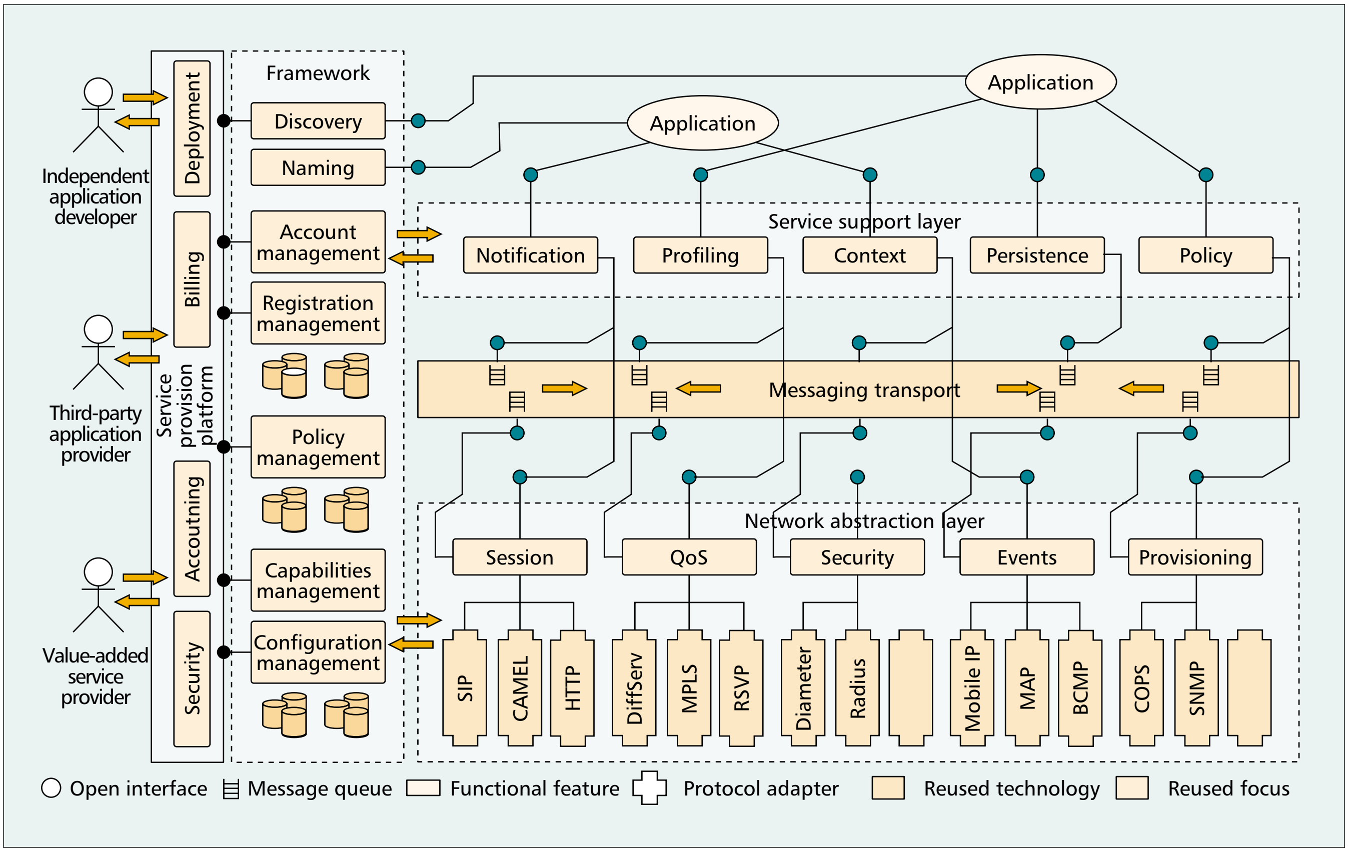
Another key differentiation of next-generation mobile networks will be the provisioning of end-to-end QoS for interactive stream-based multimedia services (e.g., video telephony, videoconference). Currently two ways to provide differentiated treatment of flows may be identified: via differentiated services code points (DSCPs) or application signaling. In the DiffServ-based solution an end host can mark the transmitted packets with proper DSCP values to trigger certain service responses. However, as the DSCP values define only relative priority and their value may not persist end to end, various problems may arise in their interpretation in intermediate nodes. Thus, this method cannot guarantee end-to-end QoS provisioning. On the other hand, various QoS signaling protocols have been proposed. The Resource Reservation Protocol (RSVP) and Integrated Services specifications are among the most widely accepted. However, none of these protocols are foreseen to provide end-to-end QoS in all situations in tomorrow's hybrid networks [8]. The IETF Next Steps in Signaling Working Group is currently defining requirements and architectures for IP signaling, evaluating existing signaling protocols, and looking at new developments. One application of a new signaling protocol is QoS.

The major problem in providing QoS in a mobile environment is the coupling of QoS and mobility management; that is, how to keep providing the same level of quality to the packet flow during and after a handover. Mobile IP-based networks require the same level of mobility support as GSM networks, which offer voice and data services, but an ongoing call is very rarely dropped due to movement. However, no technology currently exists that would allow fully seamless IP mobility and still be commercially feasible. The question is always about a balance between resource usage and level of service.

There have been several designs for extensions to RSVP to allow more seamless mobility. One solution couples RSVP and mobility management mechanisms, and proposes small extensions to localize the RSVP signaling [9]. The extension allows the mobile host to be in charge of both upstream and downstream reservations and re-set up reservations on both directions immediately following a handover. Also, context transfers studied in the IETF Seamoby Working Group aim at enhancing and minimizing signaling needed during handoffs. Another group of examples are protocols based on advance reservations, where neighboring access points keep resources reserved for mobile nodes moving to their coverage area. When a mobile node requests resources, the neighboring access points are checked too, and a passive reservation is done around the mobile node's current location. The problems with these schemes are that they require topological information of the access network and usually advance knowledge of the handover event. Furthermore, the way the resources reserved in advance are used in the neighboring service areas is still an open issue [10].

What is missing is a common QoS signaling framework, one that application designers could use to make their applications QoS-aware. Currently, there exist too broad a range of QoS

Another key differentiation of next-generation mobile networks will be the provisioning of end-to-end QoS for interactive stream-based multimedia services (e.g., video telephony, videoconference).



■ Figure 6. *The proposed architecture.*

mechanisms, for both signaling and provisioning, which makes it impossible for an application designer to decide on a technology; moreover, the very same signaling protocol must be supported by the network too. A flexible way to make applications QoS-aware is to use middleware that can map the applications' QoS requests to QoS technology used at the IP layer. This is discussed in more detail in the next section.

Still, at the end of the day, the deployment of various QoS mechanisms is a classic "chicken and egg" problem. We need new business models before any operator is willing to invest in expensive new QoS mechanisms. Take, for example, GPRS: the specifications included a number of different QoS parameters (e.g., bandwidths and priorities), but those are not currently used anywhere.

On the other hand, we also need a clear path in deploying QoS in IP networks incrementally, network by network, application by application. Currently, most parts of the technologies needed for providing distinguished services in a best-effort IP network are already available. An architecture that combines per-application signaling and traffic classes within the backbones would be a very interesting solution [11]. There are still some open issues, but the enabling technologies for QoS-enabled services are already here. The question is, which operators make the first move and decide on a migration path from best effort to the integrated services network foreseen by the IETF in the mid 1990s?

Middleware and Services

Typical application development practices in the proposed environment become very complicated by the requirement to anticipate all possible networking contexts during application design and to interpolate network-dependent code within the application's core logic. It is therefore necessary to introduce a middleware architecture that abstracts the technological details of the mobile network infrastructure and facilitates application development, deployment, and management. In general, the term *middleware* denotes a software layer that resides between the different parts of a distributed application in order to shield them from the heterogeneity of the underlying platforms. In the case of next-generation mobile service provision, the role of middleware is not restricted to handling fundamental (albeit important) tasks like seamless "binding" and communication between distributed application components; it should also incorporate higher-level functionality to facilitate adaptable service provision. Furthermore, deployment and use of diverse distributed applications may require provisioning — or even reconfiguration — of network equipment in multiple administrative domains, an issue best tackled by service provision platforms that can act as a control point for service provision flexibility and adaptability.

In a heterogeneous infrastructure applications face an enormous range of operational parameters, thereby making adaptation a critical issue. Adaptability requires constant awareness of the evolving service provision context (e.g., user preferences, user location and status, mobile terminal capabilities). A system is context-aware if it explores context to provide relevant information and/or services to the user, where relevance depends on the user's current tasks and preferences. Thus, context awareness is regarded as the ability to autonomously adapt, either proactively or reactively, to observed or announced changes in the environment (e.g., changes in the user's location or the mobile terminal's power drain rate). Consequently, middleware must support a flexible context definition scheme and also provide mechanisms for context capture, collection, management, and distribution to concerned applications using open interfaces that impose no restrictions on application architecture or implementation. Middleware services must hide the implementation details of underlying networks while exposing particular contextual aspects (e.g., available QoS classes, resource status) important for applications to behave efficiently and adapt when necessary.

Adaptation strategies may also be applied to functional network elements in order to dynamically adjust their grade of service according to specific application criteria. For instance, in the case of a multimedia streaming application whose bandwidth requirements cannot be met by the current wireless bearer service, the possible adaptation strategies are:

- To lower the video and audio codec resolutions
- To negotiate higher bandwidth for the bearer service
- A combination of both

The application of such two-way adaptation strategies in heterogeneous wireless environments requires a network abstraction layer that exposes the common functional features of underlying mobile networks to the middleware intelligence in a technologically agnostic way. By using open interfaces, mobile networks of different technologies can be "plugged in" to the middleware and make their signaling capabilities available to applications.

An important aspect of contextual information concerns user preferences. The middleware must enable the retrieval and handling of user preferences in a generic way so that applications or even the middleware itself may align their adaptation strategy to the optimal trade-off between network QoS, application presentation fidelity, communication security, and overall service cost [12]. Since users may formulate arbitrarily complex preferences that must be translated to reconfiguration actions on mobile network elements without jeopardizing their integrity, flexible and generic configuration management emerges as a paramount issue. Figure 6 below illustrates the main layers and components of the proposed policy-enabled middleware architecture, highlighting the major research issues.

The **framework** enforces security, integrity, and access policies by acting as an initial reference point for all access to and subsequent use of middleware services through its discovery interface. Parlay and OSA also adopt a framework-based architecture. Furthermore, the framework coordinates complex procedures (e.g., policy and configuration management) that require interactions across different layers of the middleware architecture.

The **service support layer** provides a set of open interfaces categorized according to major application concerns, such as contextual awareness, notification requirements, as well as policy, profiling, and persistence services. Application components may discover and use these interfaces in a technologically neutral way via the framework.

The **network abstraction layer** provides a set of generic reusable functional blocks that plug in to the middleware architecture, providing an up-to-date view of the collective capabilities of the underlying wireless networks, whether cellular, satellite, ad hoc, PAN, M-VPN, or any similar combination. In addition, they export open interfaces that allow higher layers to use selected infrastructure functionality, including retrieval of network state information and triggering of reconfiguration actions.

Messaging transport provides message-based transport services for all middleware components, thereby supporting asynchronous communications and facilitating loose coupling of interacting components. Ideally, the actual network protocols employed for information transfer would be dynamically selected according to contextual information.

The **service provision platform** exploits middleware capabilities through the framework to support secure and flexible service deployment for third-party-developed applications and value-added services. Furthermore, it engages during service discovery to customize the list of applicable services according to mobile terminal capabilities, wireless system properties (e.g., traffic load), and user privacy preferences.

In summary, the middleware architecture defines a set of functional capabilities and open interfaces that can be used as building blocks for innovative application development in next-generation mobile networks. By using the middleware API and interfaces, operator- and third-party-developed applications will be able to discover user preferences, be informed of specific events in the network infrastructure (e.g., handoff), and make informed adaptation decisions (e.g., content adaptation) when necessary. These middleware services are further exploited and integrated into the service provision platform facilities for independent (third-party) application providers.

The middleware architecture defines a set of functional capabilities and open interfaces that can be used as building blocks for innovative application development in next-generation mobile networks.

CONCLUSIONS

In the future mobile networks will display a much richer topology than today's 2G and 3G integrated systems: users will create personal area networks, they may have multiple connection options, and mobile VPNs will be dynamically modified and created. We believe that IP connectivity will form the basis of the networking functionality of such networks, providing packet delivery, QoS, and mobility support; and we have developed an outline IP architecture for such a system. Given the open nature of IP, network operators, we believe, have two main goals: reducing the costs of delivering IP packets to

In the future mobile networks will display a much richer topology than today's 2G and 3G integrated systems: users will create personal area networks, they may have multiple connection options, and mobile VPNs will be dynamically modified and created.

mobile users and offering access to common services over any access technology through functions such as service discovery, payment, and adaptation.

In this article we present concepts, such as an adaptive data link control layer and reconfigurable radio, that should enhance the efficiency of IP multimedia service delivery over wireless links and help support users connecting through new or extended network domains not controlled by their network provider.

Also presented are the trends in mobility and QoS frameworks for the mobile access network, highlighting the need for a common QoS signaling framework to support end-to-end QoS over these heterogeneous networks and for further developments of an intergraded, and hence efficient, QoS mobility solution. One of the major outstanding issues is that of securing QoS signaling, since QoS is a valuable network resource; this is particularly acute for handovers where a high level of signaling between mobile nodes and access routers is required. This issue is unresolved and will potentially lead to considerable delays in completing handover.

Current middleware technologies are not suited to mobile environments; seamless access requires the adaptation of services as dictated by users' profiles. Multimedia service may be required to be delivered over "complex" connections, where, for example, voice and video travel over different access paths. It is the job of the middleware to abstract the underlying network services and performance. We have developed a mobile middleware architecture to support the concept of seamless access, breaking the functionality down into service support, network abstraction, and messaging layers.

Acknowledgment

This work has been performed in the framework of the TIPRA consortium. The authors would like to acknowledge the contributions of their colleagues from BTexact, Telefonica ID, Portugal Telecom, TeliaSonera, NTT, TML(NEC), Fujitsu, Mitsubushi, Alcatel, Teletel, Qualcomm, Lucent Technologies, Thales, TTI, U4EA, Ellemedia, Alphamicro, ITRI, CEA-LETI, Agora, Kings College London, University of Surrey, University of Bristol, UPC, University of Murcia, Toshiba, University of Athens, UPM, University of Helsinki, University of Singapore, and CEA.

References

[1] N. Nakajima and Y. Yamao, "Development of 4th Generation Mobile Communication," *Wireless Commun. and Mobile Comp.*, vol. 1, no. 1, Jan. 2001.

[2] J. Pereira, "Fourth Generation: Now, It Is Personal," *Proc. 11th Int'l. Symp. Pers., Indoor and Mobile Radio Commun.*, 18–21 Sept. 2000, vol. 2, pp. 1009–16.

[3] N. Houssos *et al.*, "Value Added Service Management in 3G Networks," *NOMS '02*, Florence, Italy, Apr. 2002.

[4] L. Burness *et al.*, "Wireless LANs — Present and Future," to be published in *BTTJ* Special Issue on Mobile Systems, Sept. 2003.

[5] T. Zahariadis, "Trends in the Path to 4G," *IEE Commun. Eng. Mag.*, Feb. 2003, pp. 12–15.

[6] T. Zahariadis *et al.*, "Global Roaming in Next-Generation Networks," *IEEE Commun. Mag.*, vol. 40, no. 2, Feb. 2002, pp. 145–51.

[7] A. Cambell *et al.*, "Comparison of IP Micromobility Protocols," *IEEE Wireless Commun.*, Feb. 2002, vol. 1, pp. 72–82.

[8] J. Manner and X. Fu, "Analysis of Existing QoS Signalling Protocols," IETF NSIS WG, Feb. 2003, draft-ietf-nsis-signalling-analysis-01.txt.

[9] J. Manner and K. Raatikainen, "Extended Quality-of-Service for Mobile Networks," *IEEE/IFIP IWQoS 200)*, Karlsruhe, Germany, June 6–8, 2001; Springer LNCS Series, vol. 2092, pp. 275–80.

[10] B. Moon and H. Aghvami, "RSVP Extensions for Real-Time Services in Wireless Mobile Networks," *IEEE Commun. Mag.*, Dec. 2001, pp. 52–59.

[11] Y. Bernet *et al.*, "A Framework for Integrated Services Operation over Diffserv Networks," RFC 2998, Nov. 2000.

[12] N. Alonistioti, N. Houssos, and S. Panagiotakis, "A Framework for Reconfigurable Provisioning of Services in Mobile Networks," *ISCTA 2001*, Ambleside, U.K., 2001, pp. 21–26.

Additional Reading

[1] C. Perkins, "IP Mobility Support," Internet RFC 2002, Oct. 1996.

[2] A. Festag, H. Karl, and G. Schafer, "Current Development and Trends on Handover Design for All IP Wireless Networks," Tech. Univ. of Berlin, TKN-00-007, v. 1.3, Aug. 2000.

[3] R. Ramjee *et al.*, "IP-Based Access Network Infrastructure for Next-Generation Wireless Data Networks," *IEEE Pers. Commun.*, Aug. 2000, pp. 34–41.

[4] T. Zahariadis and N. Nikolaou, "Unified Wireless Access in Hot-Spot Environment," *IEEE Commun. Lett.*, vol. 6, no. 6, June 2002, pp. 259–61.

[5] C. Keszei *et al.*, "Evaluation of the BRAIN Candidate Mobility Management Protocol (BCMP)," IST Global Summit 2001, Barcelona, Spain, 2001.

[6] A. Campbell *et al.*, "Design, Implementation, and Evaluation of Cellular IP," *IEEE Pers. Commun.*, Aug. 2002, pp. 42–29.

[7] D. Wisely *et al.*, *IP for 3G*, Wiley, 2002.

[8] IST EU Project BRAIN del. 2.2, "BRAIN Architecture Specifications and Models, BRAIN Functionality and Protocol Specification," http://www.ist-brain.org

[9] IST EU Project MIND (Mobile IP Network Development) del. 1.2, "Top-level Architecture for Providing Seamless QoS, Security, Accounting and Mobility to Applications and Services," http://www.ist-mind.org

Biographies

DAVE WISELY (dave.wisely@bt.com) has worked for BT for 15 years in the fields of networks and mobility research. He pioneered optical wireless free-space links in the early 1990s, constructing a 4 km 1500 nm system across central London using optical amplifiers. He has worked in the field of mobility for the past eight years, looking first at wireless ATM and HIPERLAN type 1 systems, and more recently into WLAN and cellular mobile systems. He was one of the pioneers of an all-IP solution for future developments of 3G. He currently heads up BT's UMTS and 4G research unit and was also technical manager for the influential EU IST BRAIN/MIND project. He has just published a book entitled *IP for 3G* and is currently leading an initiative on 2G WLANs providing QoS and handover over campus-sized networks.

HAMID AGHVAMI [SM] obtained his M.Sc. (1978) and Ph.D. (1981) degrees from King's College, University of London. He continued as a postdoctoral research associate working on digital communications and microwave techniques projects sponsored by EPSRC, joining the academic staff in 1984. He is presently director of the Center for Telecommunications Research at King's. He does consulting on digital radio communications systems for British and international companies. He has published over 230 technical papers and lectured worldwide. He was a ComSoc distinguished lecturer, and a member, chair, and vice-chair of technical programs and organizing committees of many international conferences. He chaired the Communications Chapter of the U.K. and Republic of Ireland (1990–1998). He is a Fellow of the IEE.

SAMUEL LOUIS GWYN served in the Royal Navy from 1981 to 1991 as a nuclear reactor specialist. On leaving the Royal Navy he went on to study at Queen Mary and Westfield College, University of London, receiving an M.Eng. in communication engineering in 1995. Further studies at Queen Mary and Westfield College led to a Ph.D. in the application of nonlinear fynamics to teletraffic modeling in 1999. Since November 1998 he has been with the Global Wireless Systems Research Department of Bell Laboratories, Lucent Technologies, United Kingdom, where he is involved in the development of advanced protocols and network architectures for wireless communications. His current research interests include nonlinear dynamics, complexity theory, agent-based systems, software architectures and infrastructures, software protocols, 4G systems, mobility, and resource management.

THEODORE B. ZAHARIADIS (zahariad@ellemedia.com) received his Ph.D. degree in electrical and computer engineering from National Technical University of Athens, Greece, and his Dipl.-Ing. degree in computer engineering from the University of Patras, Greece. Since 1997 he has been with Lucent Technologies, first as technical consultant to ACT, Bell Laboratories, subsequently as technical manager of Ellemedia Technologies (affiliate of Bell Laboratories), Athens, Greece, and since November 2001 as a technical director. At present, he is leading R&D of multimedia platforms and 3G core network services. He is responsible for the development of QoS multimedia service platforms, home network systems, and wireless protocols implementation. His research interests are in the fields of broadband wireline/wireless/mobile communications, interactive service deployment over IP networks, embedded systems, and residential gateways. He has authored more than 30 papers, and served as principal guest editor of many special issues of magazines and journals.

JUKKA MANNER received his M.Sc. in computer science with a minor emphasis in industrial management from the University of Helsinki in 1999 and a Ph.Lic. in 2002. He has just completed his Ph.D. thesis on IP QoS in mobile environments for public defense. His research focuses on mobile and wireless communications, and QoS in future-generation mobile networks. He has participated in several national research projects (funded by the National Technology Agency of Finland and industry) and in EC projects BRAIN (IST-1999-10050) and MIND (IST-2000-28584). He has actively participated in the IETF, for example, in working groups Seamoby, NSIS, and TSVWG. He has published over 20 articles in conferences and journals.

VANGELIS GAZIS (gazis@di.uoa.gr) received his B.Sc. and M.Sc. (communication networking) from the Department of Informatics at the University of Athens, Greece, in 1995 and 1998, respectively, and his M.B.A. from the Athens University of Economics and Business in 2001. From 1995 until now, he has been with the research staff of the Communication Networks Laboratory (CNL) at the Department of Informatics in the field of mobile ad hoc networks and cellular systems (MOBIVAS, ANWIRE). In parallel, he worked with a number of established companies in the IT sector as consultant. He is currently pursuing a Ph.D. in the Department of Informatics. His research interests include flexible and adaptable service provision, management of reconfigurable systems and services, quality of service, billing, and business model issues in 3G/4G mobile networks.

NIKOS HOUSSOS (nhoussos@di.uoa.gr) received a B.Sc. degree from the Department of Informatics at the University of Athens in 1998 and an M.Sc. (with distinction) in telematics (communications and software) from the Department of Electronic and Electrical Engineering, University of Surrey, United Kingdom, in 1999. Since 1999 he is pursuing a Ph.D. at the Department of Informatics and Telecommunications, University of Athens. He is also a staff member of the Communication Networks Laboratory of the University of Athens. He is involved in projects MOBIVAS (including workpackage leadership), ANWIRE, and POLOS of the European Union IST framework. His main research interests include middleware platforms and business models for beyond 3G mobile service provision, service adaptability, and network reconfigurability management.

NANCY ALONISTIOTI (nancy@di.uoa.gr) has B.Sc. and Ph.D. degrees in informatics and telecommunications (University of Athens). 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, CORBA, intelligent networks, and software-defined radio systems and networks. She specializes in the formal specification and testing of communication protocol and services, design of object-oriented SDL platforms, mobile applications, reconfigurability, adaptable services, design of test cases for conformance testing using TTCN, design of IN services, and CORBA-based architectures and services. She has participated in several national and European projects, (CTS, SS#7, ACTS RAINBOW, EURESCOM) and is technical manager of the IST-Mobivas project. She has worked as an expert at the National Telecommunications Commission and is currently project manager, senior researcher in the CNL (University of Athens) and member of the project management team of the Greek Academic Network, GUNET. Her current research includes mobile systems, software reconfigurable radio systems and networks, adaptability, service management and download, and design of open architectures and platforms.

It is the job of the middleware to abstract the underlying network services and performance. We have developed a mobile middleware architecture to support the concept of seamless access, breaking the functionality down into service support, network abstraction, and messaging layers.