Global Roaming in Next-Generation Networks

Theodore B. Zahariadis, Konstantinos G. Vaxevanakis, Christos P. Tsantilas, and Nikolaos A. Zervos
Ellemedia Technologies, Greece
Nikos A. Nikolaou, Lucent Technologies, The Netherlands

ABSTRACT

Next-generation mobile/wireless networks are already under preliminary deployment. Mobile/wireless all-IP networks are expected to provide a substantially wider and enhanced range of services. However, an evolutionary rather than revolutionary approach to the deployment of a global all-IP wireless/mobile network is expected. To support global roaming, next-generation networks will require the integration and interoperation of mobility management processes under a worldwide wireless communications infrastructure. In this article, global roaming is addressed as one of the main issues of next-generation mobile networks. Apart from the physical layer connectivity and radio spectrum allocation plans, mobility in a hierarchical structured scheme is discussed. An all-IP wireless/mobile network combined with inherited mobility schemes of each network layer and Mobile IP extensions is proposed. In this respect, the mobility management mechanisms in WLAN, cellular, and satellite networks are analyzed, and an all-IP architecture is described and an enhanced roaming scenario presented.

INTRODUCTION

The rapidly growing demand for “anywhere, anytime” high-speed Internet access will be one of the major forthcoming challenges for mobile network operators. As the need for mobility increases, the ability to connect mobile terminals, from laptops and PDAs to future mobile videophones, to the Internet and intranet, and roam freely across geographical boundaries of heterogeneous networks has become a business driver. Next-generation mobile/wireless all-IP networks are expected to provide a substantially wider and enhanced range of services, including global convergence, interoperability, and mobility. To support global roaming, next-generation networks will require the integration and interoperation of mobility management processes under a worldwide wireless communications infrastructure [1]. Terminal and personal mobility will enable users to access services using their personal profile, independent of terminal type and point of attachment to the network. This capability, together with the inherent IP support, is a powerful combination to deliver personalized interactive multimedia services to mobile users.

Deployment of a global all-IP wireless/mobile network, however, is not a straightforward decision. First of all, the potential advantages and added value of such an evolution are not clear to subscribers, while operators have to carry out significant investments to enhance their network infrastructure and obtain expensive frequency licenses. Moreover, although IP is by far the most widely accepted protocol, it still has intrinsic weaknesses, like limited address space, lack of inherent mobility and quality of service (QoS) mechanisms, and poor performance over wireless links.

Alternatively, a phased evolution scenario for the transition from second- to third-generation systems is supported by different commercial and industrial policies and interests. In that development, mature wireless technologies that are already deployed and cover different needs and requirements may be utilized. As far as mobility is concerned, current wireless networks may be organized in three groups: wireless LANs (WLANs) for local area, cellular for wide area, and satellite for worldwide coverage. WLANs are supported by two international standards: IEEE 802.11a and b, and ETSI HIPERLAN I and II, while Bluetooth has been proposed for shorter distances. Regarding cellular networks, High Speed Circuit Switched Data (HSCSD) and General Packet Radio Service (GPRS) are
Terminal mobility applies to the ability of the network to locate a mobile terminal, route incoming or outgoing calls regardless of its network point of attachment, and maintain connections while the terminal roams in the network.

already under deployment in many countries, as evolutions of the Global System for Mobile Communications (GSM) that is the predominant technology in Europe today. Relative standards, like the Cellular Digital Packet Data (CDPD) system, the IS-95 and code-division multiple access (CDMA) systems, are installed in the United States and Asia. Finally, for satellite communications, medium/low Earth orbit (MEO/LEO) networks are considered.

One fundamental aspect of such heterogeneous networks is the ability to dynamically find paths through the network for forwarding information between specific mobile terminals. Connectivity at the physical layer is mandatory, but it does not automatically solve the problem. In a multilayered environment, extra intelligence is required so that the network can find a specific terminal or the terminal can determine the boundaries between wireless networks and switch to the most appropriate one. The routing mechanisms must automatically adapt to changes and failures in the network infrastructure and scale to support millions of mobile subscribers. While routing in the Internet has evolved to meet users’ expectations, the mobile and heterogeneous extensions of session-oriented services pose new challenges.

This article addresses one of the main issues of next-generation mobile networks: global roaming. Apart from the physical layer connectivity and radio spectrum allocation plans, mobility in a hierarchical structured scheme should be supported. This will be achieved with evolutionary adoption of an all-IP wireless/mobile network combined with inherited mobility schemes of each network layer and Mobile IP extensions. In this respect we analyze the mobility management mechanisms in WLAN, cellular, and satellite networks, and present an all-IP architecture and an enhanced roaming scenario. Conclusions are summarized in the final section.

MOBILITY MANAGEMENT TODAY

One of the major requirements of next-generation mobile networks will be personal and terminal mobility. Personal (or user) mobility provides the ability for users to access their personal services, independent of terminal type or point of attachment. Personal mobility is a primary concern of service providers. Terminal mobility applies to the ability of the network to locate a mobile terminal, route incoming or outgoing calls regardless of its network point of attachment, and maintain connections while the terminal roams in the network. In this article we focus on terminal mobility and assume that personal mobility is transparent to the network layer.

The wireless network is organized in service regions called cells. According to the type of network, the size of cells varies from a few square meters in wireless LANs to hundreds of square kilometers in satellite networks. In all cases, a central system (access point, base station, or telecommunication satellite) handles the frequencies or channels of each mobile terminal and supports location management and handover. Location management tracks and locates a terminal for delivering of incoming calls, while handover (or handoff) allows for an active connection to remain alive while the terminal roams. Location management handles information concerning the mobile terminal, its original cell, the cell where it is currently located, and paths and routes toward the current location. This information is updated either periodically or on demand when a specific event occurs (e.g., a terminal changes cell, a central station fails or is overloaded), and is stored and retrieved in location or paging databases, independent of the specific network or location management protocol.

Handoff management handles roaming in the same cell (intracell handover) or between cells (intercell handover). In intracell handover, when the signal-to-noise ratio (SNR) falls and the radio channel deteriorates, a dynamic channel allocation (DCA) mechanism is activated that transfers the calls to a new radio channel or frequency with appropriate strength within the cell. In intercell handover connections are passed to a new central station that covers a neighboring area. If the transfer is unimportant, it results in a soft handover, while if the connections have to be re-established, we have a hard handover. Handover may be initiated by either the user terminal or the mobile network. In network-controlled or mobile-assisted handover, the network establishes and handles the connections, searches for new resources, and performs any additional routing or updating functions. In mobile-controlled handover, the mobile initiates handover and seeks available resources, while the network authorizes the operations and handles the control and delay requirements. Finally, intercell handover may be identified between wireless cells of the same type (horizontal handover) or wireless cells with different coverage and hierarchical layers (vertical handover). In horizontal handover coverage and data rates remain the same, simplifying roaming, while in vertical handover bandwidth adaptation is also required.

In the remaining article we initially analyze the case of soft horizontal handover in WLAN, 2G, and satellite networks, and then propose a solution for soft vertical handover toward global roaming in next-generation networks.

MOBILITY IN WIRELESS LANs

Mobility management in WLANs is primarily supported by Internet Engineering Task Force (IETF) Mobile IP [2] and various extensions. The address used by the traditional IPv4 protocol, apart from identifying a specific network node, contains topological information. Under the current form of IPv4, if the mobile terminal moves between different subnetworks without changing its address, the routing process will cease to be operational. On the other hand, if the IP address is modified, all the active connections will be terminated. To overcome this problem and allow a mobile terminal to roam freely around the network while still communicating and maintaining the same IP address, the Mobile IP protocol is utilized.

Mobile IP introduces two new functional components: the mobility agent (MA) and mobile node (MN). An MA, which could be a home agent (HA) or foreign agent (FA), is
responsible for the mobile terminals of a specific subnetwork, while a MN is located inside a mobile terminal. The subnetwork to which the IP address of a mobile terminal belongs is called the home network, whereas any other subnetwork the mobile terminal might visit is called a foreign network.

When a MN is located on the home network, it operates without any mobility services. Whenever it detects that it has moved to a foreign network, it obtains a care-of address, which can be determined either from agent advertisement messages sent by the FA (a FA care-of address) or using the Dynamic Host Configuration Protocol (DHCP) (a collocated care-of address). The former case is shown in Fig. 1. Having obtained a care-of address, the MN registers that address, through the FA, with its HA (2, 3). From this point on, the HA intercepts all IP packets destined to the MN and tunnels them to the FA (4, 5), where, after decapsulation, they are forwarded to the MN (6).

In the opposite direction, the IP packets sent by the MN are subject to the standard IP routing mechanism and do not necessarily pass through the HA (7).

**MOBILITY IN 2G NETWORKS**

Mobility management in second-generation (2G) cellular networks is supported by two international standards: the Electronic/Telcommunications Industry Associations Interim Standard 41 (EIA/TIA IS-41) mostly used in the United States for the AMPS and IS-54/IS-136 networks, and the GSM Mobile Application Part (MAP) for GSM, DCS-1800, and PCS-1900. In both cases, the call processing and location management functions are based on Signal System 7 (SS7) [1]. As shown in Fig. 2, 2G networks are organized in cells, while the switching center responsible for a specific geographical or logical area is known as the mobile switching center (MSC). Location management is based on location databases, called home location register (HLR) and visitor location register (VLR). VLRs can be considered extra intelligence on each MSC, and contain temporal information for a specific area. HLRs are hierarchical higher databases that contain permanent information for each terminal. The entry of each subscriber is registered in one HLR, including a link to the VLR, which is responsible for the area the terminal is currently visiting.

When a mobile terminal changes base station, it may roam to a cell that corresponds to a new serving VLR. In that case it has to update the information stored in the HLR. Therefore, the terminal initiates an update message (1), which via the base station and MSC is forwarded to the current associated VLR. The VLR checks its local records. If the terminal’s Mobile Identification Number (MIN) is already stored there, no further action takes place, since the terminal has not changed location area. Otherwise, the terminal’s MIN is stored locally and a new update message is forwarded to the HLR (2). The HLR in turn authenticates the terminal and replies with a positive registration acknowledgment to the new VLR (3). Additionally, the HLR may send a registration cancellation message to the old VLR, or a periodical mechanism may automatically update the VLR database and remove out-of-date entries.

Whenever a new connection is initiated (4) the VLR will check its local records again for the called mobile. If both calling and called parties are in the same servicing area, the call is directly routed to the terminal. Otherwise, the VLR of the calling terminal initiates a location request to the HLR (5). The HLR confirms that the terminal is located in this area and sends a route request message (6). This message is forwarded via the VLR to the serving MSC, which allocates a temporary local directory number (TLDN) for the specific terminal. The TLDN is returned to the HLR (7) and forwarded to the calling VLR (8). Using SS7, a path between the MSCs is established (9), and a paging or alerting message is sent to the called mobile terminal.

If the terminal changes VLR while connections are established, all the steps have to be repeated, increasing the signaling overhead, especially when the terminal is far away from the HLR. This has motivated extended research in distributed and hierarchical HLR databases [3].

**MOBILITY IN SATELLITE NETWORKS**

Telecommunication satellites can be categorized in three groups: geostationary Earth orbit (GEO), MEO, and LEO. The advantage of GEO satellites is that they rotate at the same angular velocity as the Earth, always keeping a fixed position in reference to the ground. In this way GEO satellites appear at a fixed latitude and longitude. Moreover, due to the rather long distance from the Earth surface (roughly 36,000 km from the surface), GEO satellites have a very large servicing area of almost 1/3 of the Earth’s surface, from about 75° south to 75° north latitude. The combination of the fixed position along with the very large servicing area provides near-global coverage with a minimum of three satellites in orbit. Communications GEO satellites are very useful, especially for broadcasting services (e.g., TV broadcasting).

MEO satellites rotate at an altitude of around 10,000 km. In contrast to GEO, MEO satellites do not have a fixed position over the Earth. They
rotate at different angular velocities compared to the Earth; thus, they move in reference to the ground. Their orbit period measures about 6 h, while the maximum time during which a MEO satellite is above the local horizon for an observer on the Earth is on the order of a few hours. A global communications system using MEO orbits requires a reasonable number of satellites in two to three orbital planes to achieve global coverage. A prime example of an MEO system is the U.S. Navistar Global Positioning System (GPS).

Finally, LEO satellites rotate in orbits much closer to the Earth at a height of 500–2000 km above the surface of the Earth. Like MEO, LEO satellites do not have a fixed position over the Earth, and move in reference to the ground. The maximum time during which a satellite in LEO orbit is above the local horizon for an observer on the Earth is up to 20 min, while there are long periods during which the satellite is out of view. This may be acceptable for a store-and-forward type of communication system, but not for interactive communications. However, LEO satellites, due to their smaller distance from the surface, may achieve comparable very good end-to-end delay and have lower power consumption requirements for both the mobile terminal and satellite. In order to increase accessibility and make global coverage possible, more than one satellite and multiple orbital planes are utilized. For instance, the IRIDIUM system utilizes 66 satellites (plus 6 in-orbit spares) in 6 orbital planes at an orbital height of 780 km with an orbital period of 100 min 28 s.

Location management in GEO networks may be considered similar to 2G networks. However, in MEO and LEO it is not only the terminal movement, but the satellite movement as well that must be taken into account. In order to better locate a terminal, the coverage area of a single satellite is divided into small cells, called spotbeams [1]. Both the geographic and time dependencies introduce new research concerns in efficient positioning of a terminal to a specific spotbeam.

Handover in GEO networks is not a very common issue. On the contrary, handover in LEO networks is much more important and demanding. As terminals and satellites change position, spotbeam and satellite handovers are defined. The size of the spotbeam is rather small; therefore, satellite handover typically happens every 10 min, spotbeam handover every 38 s [1]. Spotbeam handover occurs on the same satellite; thus, the ability to maintain connections or not depends on the satellite resources and the traffic load on the specific spotbeam. In order to support satellite handover without the need for terrestrial support, intersatellite links (ISLs) are required. ISLs may route calls from one LEO satellite to another, which is located in the same orbit (intraplane) or different orbits (interplane) (Fig. 3). In either intra- or interplane handover, signaling overhead and delay are quite significant factors that must be minimized. An important feature of LEO satellite handovers is that they happen in most cases due to satellite rotation, not to terminal movement; thus, all connections may be transferred as a group to a neighboring satellite, minimizing the time to identify the best ISL for each connection.

The straightforward approach to satellite handover is to establish a new connection each time a handover happens. More sophisticated approaches have been proposed for optimal satellite handover. For example, the route augmentation approach simple extends the original connection with a hop to the next satellite; partial connection rerouting replaces and reroutes only the part of the connection that has been modified [4]; while more routing problems in LEO satellite networks have been addressed in [1, 5].

**Next-Generation Network Architecture**

Next-generation mobile/wireless networks are expected to provide a substantially wider and enhanced range of services. Global convergence, interoperability, and mobility are some of the differentiating factors from current networks. Moreover, the inherent IP support will encour-
age new personalized interactive multimedia services, as well as new broadband applications, such as video telephony, videoconferencing, and mobile Internet.

Deployment of a global all-IP wireless/mobile network [6], however, is expected through evolutionary rather than revolutionary steps. Vendors promote the new profitable IP services that a network will allow, while postal, telephone, and telegraph companies (PTTs) look to maximize profit and return of investment based on existing equipment. As a result the wireless network infrastructure may be organized in a cell hierarchy, based on technology either already deployed or still under development (Fig. 4). Starting from the home cell, coverage in private buildings (e.g., house, office) or in public “hot-spot” locations (e.g., airport, train station, conference center) may be provided by access points (AP). IEEE 802.11, HIPERLAN, Bluetooth, and Home-RF are alternative technologies that may be deployed. The APs may also provide connectivity in picocells, while a combination with pico-GSM or DECT can also be considered. Moreover, fixed wireless access via central stations (CSs) and remote stations (RSs) may provide wireless access up to macrocells in suburban areas. Horizontal mobility to mobile terminals that move with different speeds in micro- or macrocells may be provided utilizing 2G and 2G+ networks (GSM, HSCSD, GPRS, EDGE, CDPD, IS-95, CDMA). Connectivity and mobility in satellite cells are provided via GEO, MEO, or LEO satellites and Fixed Earth Stations (FESs) or mobile satellite terminals (STs).

In order to support both horizontal and vertical roaming in such a complex environment, the first step is to gain connectivity at the physical layer. In this respect, either multimode or adaptive terminals are considered. For example, terminals equipped with commercial wireless LAN (e.g., IEEE 802.11b), cellular (e.g., GSM/GPRS), or satellite network interface cards may be introduced, while soft radio techniques have also been proposed. Global roaming, however, requires integration and interoperation of the mobility management processes of each independent network. IP is the most widely accepted protocol; thus, mobility based on IP will be leveraged.

A detailed architecture of an all-IP wireless/mobile network architecture is shown in Fig. 5. WLAN, 2G, and 3G cellular and satellite networks are selected as alternative radio access networks. Due to different physical and protocol characteristics, each radio access network consists of different base stations and radio control nodes, connected to the common core network via a service support node (SSN). This may be an MSC+ for cellular networks, an IP L1/L2 switch for the WLAN, or an FES. The SSN also provides the VLR or FA functionality, respectively, in cooperation with an extended HLR or home subscriber server (HSS). HSS maintains user profiles and may integrate or cooperate with a remote authentication dial-in user server (RADIUS) and/or an authentication, authorization, and accounting (AAA) server for user authentication and authorization. Interoperability between circuit-switch-based networks and packet-based networks is also mandatory; thus, the common core network of the proposed all IP architecture supports both circuit-based connections and packet-based transmission. Access to both the public switched telephone network/integrated services digital network (PSTN/ISDN) and the Internet is provided via interworking function (IWF) units, voice gateways, firewalls, or generally gateway support nodes (GSNs). Additional servers (e.g., DHCP, DNS) provide complementary services in the IP domain, while

![Figure 4. Cell hierarchy of a next-generation network.](image)
mobility servers provide mapping between Universal Personal Telecommunications (UPT) numbers and dynamic IP addresses.

The architecture is based on enhancements of existing equipment, so we may assume that horizontal roaming will be handled by the specific network roaming mechanisms. For example, a direct extension toward roaming of IP traffic in GPRS networks has been proposed by the GSM Association, in the form of a GPRS Roaming Exchange (GRX) architecture that carries traffic between mobile operators’ networks [7]. In the vertical roaming scenario, however, the terminal should have a more active role and initiate the specific roaming mechanisms. Starting from the WLAN cell, whenever a terminal is activated it has to obtain a valid IP address. This may be a preconfigured IP address or most probably a dynamically allocated one via a local or distributed DHCP/DNS server. Moreover, the mobility server responsible for the specific hot spot or AP may authenticate the terminal via a centralized or distributed RADIUS/AAA server. User/terminal authentication and authorization based on MAC/ password pair or Subscriber Identification Module (SIM) card could be considered. The RADIUS+ server communicates with the VLR and/or HLR+ servers and associates the hot-spot user with his/her cellular database entry; thus, the user will receive a single bill for all services. After registration and authentication, the user can roam in the hot-spot while communicating via the WLAN connection. Moreover, a virtual private network (VPN) can be set up via higher-layer protocols (e.g., IPSec, L2TP), and the mobile terminal can access the corporate intranet. When the user moves outside the hot-spot coverage area, the terminal has to initiate a vertical roaming mechanism and soft handover to a GSM; a UMTS or satellite network can be activated. The mobile network intermediate nodes (node B, RNC in UMTS, BSC, BTS in GSM, FES in satellite) are considered transparent to mobile Internet traffic, since they control radio resources and handover decisions only at the physical layer. The terminal communicates with the mobility servers located in the IP part of the network, and a new authentication/authorization process is initialized. After the connection at the physical layer has been established, various Mobile IP extensions may be applied [8, 9] to keep connections uninterrupted.

CONCLUSIONS

Next-generation mobile/wireless networks are already under preliminary deployment. However, since the licenses and equipment cost are still very high, PTTs try to maximize return of investment on existing networks and increase profit margins. Global roaming in current and next-generation networks is an important issue that will boost mobile Internet in the years to come. In this article we consider a hierarchical cell architecture consisting of infrastructure either installed or under development and discuss soft horizonal mobility management mechanisms in case of WLAN, 2G, and satellite networks. Final-
ly, the case of vertical handover is discussed. An enhanced roaming scenario is presented initiated by the mobile terminal and supported by an all (Mobile) IP network.

REFERENCES

[8] K. Vaxevanakis et al., “Software Toolkit for Providing Unin-


BIographies

Theodore B. Zahariadis [M] (zaharia@surfer.amc.bell-labs.com) received his Ph.D. degree in electrical and computer engineering from the 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, New Jersey, and subsequently as technical manager of Ellemedia Technologies (affiliate of Bell Laboratories of Advanced Technologies EMEA), Athens, Greece. From 1990 to 1996 he was a technical consultant to the Institute of Communications, where he was responsible for the interactive TV activities, INTRACOM S.A., where he participated in the Internet over Satellite system, proposed for the 1997 European IT Prize, and INTRASOF S.A., where he was a major designer of a TMN system, built on behalf of IBM. At present, he is leading R&D of end-to-end interactive multimedia services. He is responsible for the development of QoS multimedia services (data, voice, video) for wireless networks implementation, home network control and management, and 3G core network services. He has participated in many European R&D programs. He is a member of the Technical Chamber of Greece.

Nikolaos A. Zervos [M] received his Ph.D. degree in electrical engineering from the University of Toronto, Canada, his M.A.Sc. degree in systems and computing science from Carleton University, Ottawa, Canada, and his Dipl. in electrical and mechanical engineering from National Technical University of Athens, Greece. He is currently managing director of Ellemedia Technologies (affiliate of Bell Laboratories of Advanced Technologies EMEA), Athens, Greece. He is currently directing exploratory development of core network technologies in the areas of home networking, wireline/wireless systems, and broadband networking components. He joined the Data Communications Research Department of Lucent Technologies in April 1985. From 1985 to 1993 his research was in the general area of communication and information theory with emphasis in mathematical statistics, signal space coding, and statistical signal processing. His research/development work has been directed toward state-of-the-art digital transmission system architectures capable of establishing record high transmission rates over (wired and wireless/mobile) narrowband and broadband (voiceband, radio, twisted pair, cable TV) multiple access media. He was an instrumental researcher on exploratory development of high data rate loop systems over UTP. He is one of the world’s experts on bandwidth-efficient digital transmission. His research has led to data rates up to 155 Mb/s for UTP-based ATM LAN and broadband access applications such as the broadband telephone network. He is author or co-author of nine patents in the areas of data transmission and digital signal processing. He was an editor of IEEE Transactions on Communications, co-chair and technical program chair of the 1991 IEEE Communication Theory Workshop, chair of the 1995 IEEE Communication Theory Mini-Conference, and technical program vice-chair of IEEE ICC ’96, and is co-chair of ICC 2002, to be held in New York.