A Ubiquitous Mobile Communication Architecture for Next-Generation Heterogeneous Wireless Systems

IAN F. AKYILDIZ AND SHANTIDEV MOHANTY,
GEORGIA INSTITUTE OF TECHNOLOGY
JIANG XIE, UNIVERSITY OF NORTH CAROLINA — CHARLOTTE

Abstract

Rapid progress in research and development of wireless networking and communication technologies has created different types of wireless systems (e.g., Bluetooth, IEEE 802.11, UMTS, and satellite networks). These systems are envisioned to coordinate with each other to provide ubiquitous high-data-rate services to mobile users. In this article a novel architecture, Architecture for Ubiquitous Mobile Communications (AMC), is introduced that integrates these heterogeneous wireless systems. AMC eliminates the need for direct service level agreements among service providers by using a third party, a network interoperating agent. Instead of deploying a totally new infrastructure, AMC extends the existing infrastructure to integrate heterogeneous wireless systems. It uses IP as the interconnection protocol. By using IP as the gluing protocol, transparency to the heterogeneities of the individual systems is achieved in AMC. Third-party-based authentication and billing algorithms are designed for AMC. New mobility management protocols are also developed to support seamless roaming between different wireless systems.

Introduction

Mobile users are demanding anywhere and anytime access to high-speed data real- and non-real time multimedia services from next-generation wireless systems (NGWS). These services have different requirements in terms of latency, bandwidth, and error rate.

Currently, there exist disparate wireless networks, such as Bluetooth for personal areas, wireless LANs (WLANs) for local areas, Universal Mobile Telecommunications System (UMTS) for wide areas, and satellite networks for global networking. These networks are designed for specific service needs and vary widely in terms of bandwidth, latency, area of coverage, cost, and quality of service (QoS) provisioning. For example, satellite networks can provide global coverage, but are limited by high cost and long propagation delays (from 20–25 ms for low Earth orbit [LEO] satellites to 250–280 ms for geostationary Earth orbit [GEO] satellites). Third-generation (3G) wireless systems like UMTS can deliver a maximum data rate of 2 Mb/s at lower cost and have wide areas of coverage. WLANs support bandwidth up to 54 Mb/s at extremely low cost. It may be noted that the future generation of WLANs are expected to provide data rates in excess of 100 Mb/s. However, WLANs can support only low-mobility users and have small coverage areas. Therefore, none of the existing wireless systems can simultaneously satisfy the low latency, high bandwidth, and ubiquitous coverage needs of mobile users at low cost. This necessitates a new direction in the design of NGWS.

There can be two possible approaches in designing NGWS:

- Develop a new wireless system with radio interfaces and technologies that can satisfy the requirements of the services demanded by future mobile users.
- Intelligently integrate the existing wireless systems so that users may receive their services via the best available wireless system.

The first approach is expensive and needs more time for development and deployment; hence, it is not practical. Therefore, we advocate the use of the second approach, which is a more feasible option [1]. Following the second approach, heterogeneous wireless systems, each optimized for some specific service demands and coverage area, will cooperate with each other to provide ubiquitous "always best connection" [2] to mobile users. In this integrated heterogeneous network architecture, each user is always connected to the best available network or networks.

The integrated NGWS keeps the best features of the individual networks: the global coverage of satellite networks, the wide mobility support of 3G systems, and the high speed and low cost of WLANs. At the same time, it eliminates the weaknesses of the individual systems. For example, the low-data-rate limitation of 3G systems can be overcome when WLAN coverage is available through handover of the user to the WLAN. When the user moves out of a WLAN coverage area, it can be handed over to the overlaid 3G system. Similarly, a satellite network can be used when neither a 3G system nor a WLAN is available. The basic idea is to use the best available network at any time.

The integrated NGWS must have the following characteristics:

- Support for the best network selection based on users' service needs
- Mechanisms to ensure high-quality security and privacy
- Protocols to guarantee seamless intersystem mobility

Moreover, the architecture should be scalable; that is, able to integrate any number of wireless systems of different service providers who may not have direct service level agreements (SLAs) among them.

In this article we present the design of a novel architecture for NGWS, Architecture for Ubiquitous Mobile Communications (AMC). AMC integrates heterogeneous wireless systems using a third party, Network Inter-operating Agent (NIA). Thus, it eliminates the need for direct SLAs among different network operators. AMC achieves transparency to the heterogeneities of individual systems by using Internet Protocol (IP) as the inter-connection protocol. We design protocols for authentication, authorization, and billing. In addition, we develop algorithms for the best network selection and mobility management protocols for intersystem mobility in AMC.

The remainder of the article is organized as follows. We begin by enumerating the design goals of NGWS and contrast the related work in the context of these goals. We then present our proposed architecture, AMC. We describe authentication, authorization, and billing mechanisms followed by intersystem handover management protocols. We discuss the key features of AMC. Finally, we will discuss the future work and conclusions.

Design Goals

In NGWS, users move between different networks as discussed in the previous section. They want to maintain their ongoing communications while moving from one network to another. These heterogeneous networks (WLANs, 3G cellular networks, and satellite networks) may or may not belong to the same service provider. Hence, support for intersystem movement between networks of different service providers is required in NGWS. One way of achieving roaming among networks of different service providers is to have bilateral SLAs among them. This approach is not feasible due to the following reasons:

- First, operators have reservations about opening their network databases (which is required for authentication, billing, and service provisioning when an SLA is established between operators) to all other operators.
- Second, each time a new operator deploys its wireless network, it has to create an SLA with every other operator separately. The number of operators of wireless networks is very large; for example, the number of Global System for Mobile Communications/General Packet Radio Service (GSM/GPRS) operators alone is around 620. Similarly, there are a large number of operators for 3G networks, satellite networks, and WLANs. Given the large number of operators, it is almost impractical for network operators to create direct SLAs with every other operator. It may be noted that to overcome this problem in GPRS global roaming support, the GSM Association has proposed the use of GPRS roaming networks instead of direct SLAs among GPRS operators [3].

Therefore, there is a need for a new architecture to achieve roaming among heterogeneous networks of different service providers who may not necessarily have direct SLAs among them. We advocate that the architecture of NGWS should have the following characteristics:

- **Economical**: The architecture should try to use as much existing infrastructure as possible and minimize the use of new infrastructures. This will ensure economical and speedy deployment.
- **Scalable**: The architecture should be able to integrate any number of wireless systems of both existing and future service providers.
- **Transparency to heterogeneous access technologies**: The architecture should be transparent to different access technologies of different networks.
- **Secure**: The architecture should be able to provide security and privacy equivalent to existing wireless networks.
- **Seamless mobility support**: The architecture should support seamless mobility management to eliminate connection interruption and QoS degradation during intersystem roaming.

We survey the architectures for the integration of different communication systems proposed in the literature in the next section.

Related Work

The concept of integrating two or more communication systems to get better performance is already in use and has been proven to be highly efficient. The existing integration architectures address the following issues: integration of two specific systems, integration of two general systems, integration of networks of multiple operators but of the same technology, and integration of networks of different operators employing different technologies. These architectures are described below.

In [5, 6], specific pairs of different systems are integrated through an additional gateway, such as interworking of Digital Enhanced Cordless Telephone (DECT) with GSM and of IS-41 with GSM. The additional gateway proposed between a pair of systems takes care of interworking and interoperating issues such as transformation of signaling formats, authentication, and retrieval of user profiles. Similarly, the integration of satellite and terrestrial networks has been studied in [6]. Appropriate interworking units specifically for the considered terrestrial networks are placed at the interface between the satellite and terrestrial systems. In addition, different architectures are proposed to integrate WLAN and 3G systems [7]. All the above architectures are limited to the integration of a specific pair of systems and hence are not scalable to integrate multiple systems.

The boundary location register (BLR) approach [8] is proposed to integrate any two adjacent networks with partially overlapping areas. However, this approach is not scalable in the sense that one BLR gateway is needed for each pair of adjacent networks when integrating multiple networks. Moreover, the above architecture assumes the existence of SLAs between the systems, which is not desirable as discussed earlier.

The GSM Association has proposed an inter-public land mobile network (PLMN) backbone using GPRS Roaming Exchange (GRX) [3] to globally integrate the GPRS networks deployed by various providers who may not necessarily have direct SLAs among them. This architecture uses multiple peer GRX nodes for connecting several GPRS networks. This architecture is limited to only one technology (i.e., GPRS networks).

In the SMART project [9], a new architecture is proposed to integrate the heterogeneous wireless systems. This architecture uses two distinct networks: *basic access network*, and *common core network* for signaling and data traffic, respectively. This architecture is scalable, but requires the development and deployment of the new basic access and common core networks, and hence is not cost-effective.

Heterogeneous network integration using Mobile IP and Session Initiation Protocol (SIP) are proposed in [10, 11]. In these architectures, Mobile IP and SIP use authentication, authorization, and accounting (AAA) agents to carry out authentication and accounting during internetwork roaming. However, these architectures do not have any mechanism to decide the best available network. Moreover, although Mobile IP and SIP are used to carry out intersystem handoff, seamless support of intersystem handoff is not always guaranteed [12].

To the best of our knowledge, none of the above architectures satisfy all the requirements of the NGWS outlined earlier. This is our motivation in designing a new architecture for NGWS with all the design goals. We describe our architecture in the next section.

The Proposed Architecture

First, we describe our motivation for selecting IP to integrate different wireless systems. Then we describe the details of AMC.

IP-Based Interconnection

The integrated NGWS has the following heterogeneities:

- Access technologies: NGWS will include many heterogeneous access networks using different radio technologies (GPRS, cdma2000, UMTS, WLAN, etc.).
- Network protocols: NGWS will have different protocols for transport, routing, mobility management, and so on.

These heterogeneities ask for a common infrastructure to

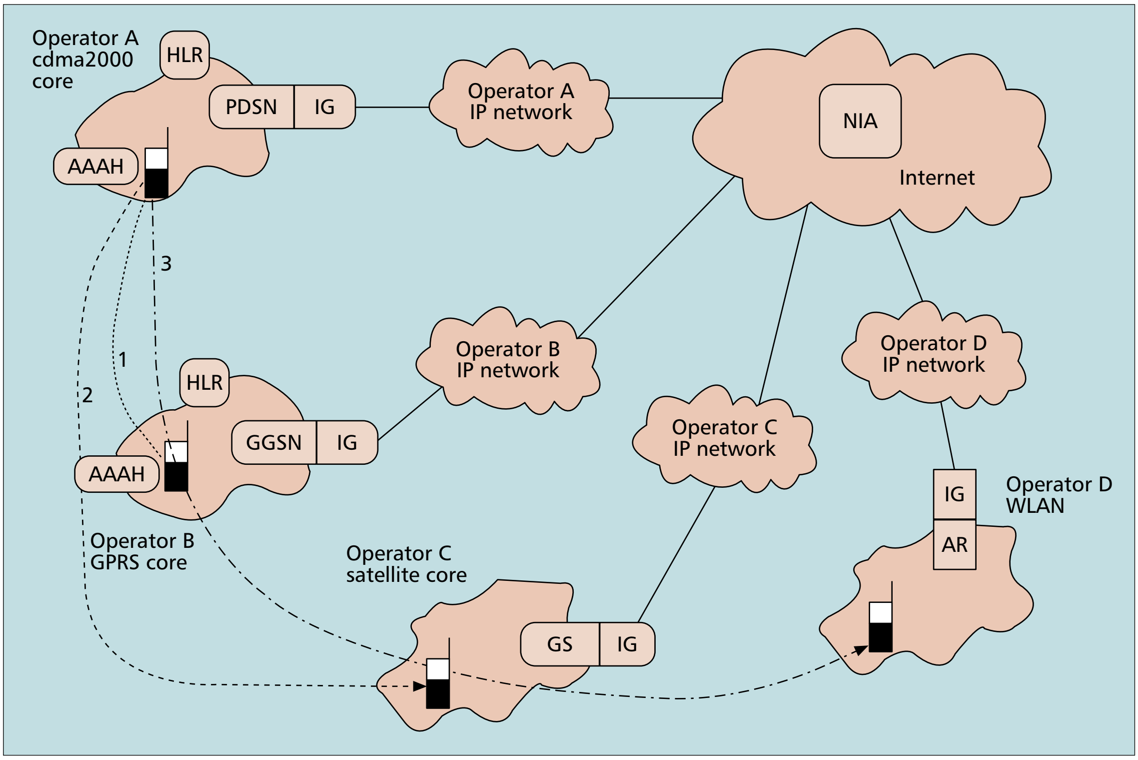

FIGURE 1. *An NIA-based integrated architecture for NG wireless systems.*

interconnect the heterogeneous networks. Since IP provides a globally successful infrastructure for supporting applications in a scalable and cost effective way, it is expected to become the core backbone network of NGWS.

By using IP as the common interconnection protocol, mobile users may roam among multiple wireless networks in a manner that is transparent to different radio technologies. This is achieved by using Mobile IP [13] to support roaming between different access technologies. This IP-based interconnection solution hides the heterogeneities of the lower-layer technologies from higher layers. Moreover, in NGWS IP-based mobile devices with multiple radio interfaces may switch from one network interface to another by using multiple care-of addresses (CoAs), one for each interface. In this scenario the interface switching is carried out as defined in [14]. Therefore, this approach requires no modifications to the existing heterogeneous radio technologies and provides the greatest transparency to ubiquitous communications in a heterogeneous network environment.

Architecture for Next-Generation Wireless Systems

Architectures requiring direct SLAs among different providers are not feasible because of the reasons mentioned earlier. We propose the use of a third party to integrate heterogeneous wireless systems of different service providers. In this case, an individual network operator needs to establish a direct SLA only with the third party instead of establishing separate SLAs with all other operators.

Our proposed AMC for NGWS is shown in Fig. 1, which consists of cdma2000, GPRS, satellite network, and WLAN of service providers A, B, C, and D, respectively. These systems are connected to the Internet through gateways: cdma2000 is connected to the Internet via a packet data serving node (PDSN), GPRS through a gateway GPRS support node (GGSN), the satellite network through a gateway station (GS), and the WLAN through an access router (AR). It may be noted that AMC can integrate any number of systems of different service providers.

We define two new entities, Network Interoperating Agent (NIA) and Interworking Gateway (IG), for AMC. The NIA functions as the third party and IG as the gateway between a particular system and the NIA. The NIA resides in the Internet, whereas an IG resides in each system and acts as the gateway, as shown in Fig. 1. Instead of being connected to every other system, an IG is connected to only one entity, the NIA. It can be implemented as a separate entity or integrated with the gateways through which individual systems are connected to the Internet (e.g., PDSN, GGSN, GS, and AR, for cdma2000, GPRS, satellite, and WLAN), as shown in Fig. 1. We advocate the latter choice because in this case an IG can be plugged into the existing infrastructure, so it is easy to implement and manage.

In AMC, the network providers do not have to create separate SLAs with every other provider. Instead, they offer roaming services to subscribers of other providers with only one SLA with the NIA. This eliminates the need for separate SLAs between each pair of systems and makes AMC scalable. The NIA is supported by a third-party provider. We assume that the operator of the NIA generates revenue from the network providers who have SLAs with the NIA. It may be noted that network providers charge more from their subscribers when the latter communicate through a foreign network. We advocate providers sharing a part of this revenue generated during intersystem roaming of their subscribers with the NIA. Operators will be interested in using the NIA to support roaming to the networks of other operators as a value-added service feature for their subscribers. For example, a similar business model is used by iPass to provide global remote access services.

The NIA handles the authentication, authorization, billing, and mobility management issues of intersystem roaming. Currently, the AAA broker networks support authentication, authorization, and billing for users belonging to different service providers. However, they cannot handle the mobility management issues, and hence cannot be used as the third party. We describe the components of the NIA and IG in the following subsection.

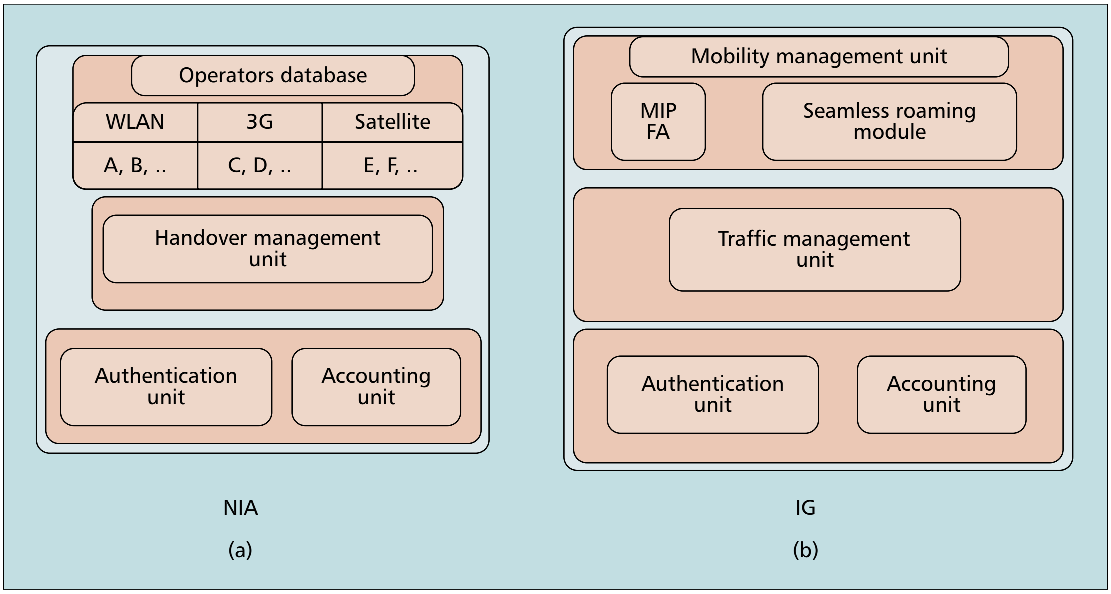

FIGURE 2. *Logical diagram showing the subsystems of NIA and IG.*

Components of the NIA and IG

The subsystems of the NIA are as follows. These are shown in Fig. 2a.

•The **authentication unit** is used to authenticate the users moving between two systems belonging to two different service providers, as discussed later.

•The **accounting unit** handles the billing issues between different systems, as discussed later.

•The **operators database** stores information about the network operators who have SLAs with the NIA.

•The **handover management unit** decides if the intersystem handover (ISHO) request should be granted. The handover management unit derives the network access identifier (NAI) from the Mobile IP *Registration Request* message and verifies with the operators database for the existence of the SLA with the home operator of the mobile terminal (MT). When applicable, it also acts as a mediator between different networks (e.g., for transferring user service profiles). In addition, the handover management unit decides the best available network, as discussed later.

The components of the IG are described below. These are shown in Fig. 2b.

•The **mobility management unit** implements Mobile IP [13] (MIP) functionalities using the MIP foreign agent (FA). Note that when a particular wireless system already implements Mobile IP, e.g., cdma2000, there is no need to implement the FA in the IG. In this case, the FA in the IG refers to the FA already implemented in the system. The mobility management unit has a **seamless roaming module** that will implement mobility management protocols for seamless intersystem roaming as discussed later.

•The IG implements traffic monitoring function in its **traffic management** unit. The specific implementation of this unit may be different for different providers based on their policies.

•The **authentication unit and accounting unit** provide authentication and billing support to roaming users.

Security and Billing Support in AMC

In AMC, authentication, authorization, and billing mechanisms are carried out as follows.

Authentication and Authorization

The proposed security architecture for AMC is shown in Fig. 3, where the foreign network (FN) is the network the MT is currently visiting and the HN is the home network of the MT. This architecture glues the security architectures of the FN and HN through the authentication unit (AU) of the NIA (AU_NIA). The use of AU_NIA eliminates the need for any direct security association/agreement between the FN and HN. Both the FN and HN have separate security associations/agreements with AU_NIA. Thus, AU_NIA functions, in essence, as a trusted third party for authentication and authorization dialogs between the FN and HN. The working principle of this third-party-based security architecture is as follows. When a mobile user requests services from an FN and the FN determines it has no SLAs with the HN provider, it forwards the request to AU_NIA to authenticate and authorize the user. Then AU_NIA talks to the HN provider and mediates

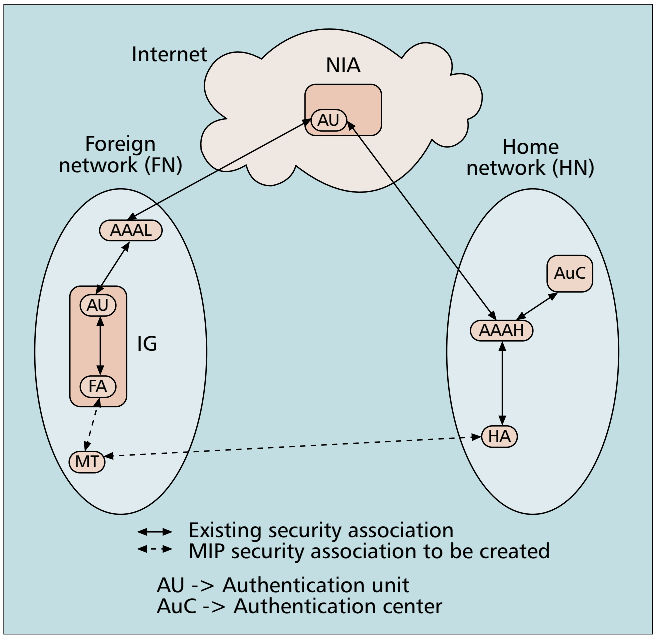

FIGURE 3. *The proposed security architecture for AMC.*

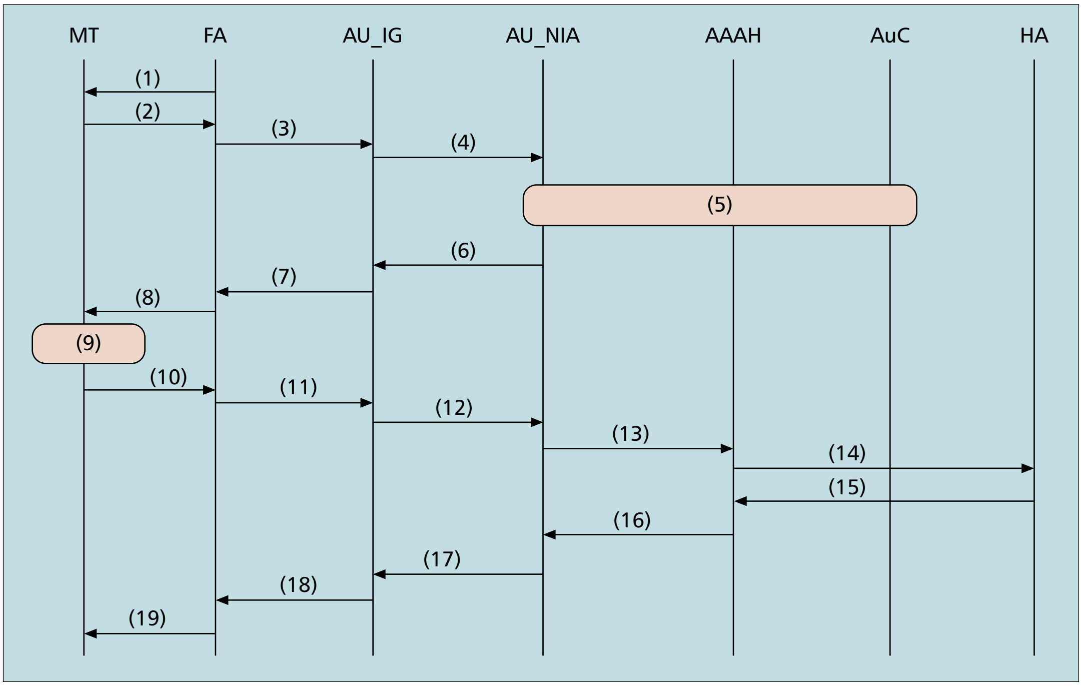

FIGURE 4. *The authentication and authorization signaling messages for AMC.*

between the FN and HN for authentication and authorization message exchanges. Once the user is authenticated, AU_NIA mediates for the creation of security associations/keys that are required between the FN and HN. At the end, the HN and FN will be mutually authenticated and will have session keys for secured data transfer.

We integrate the authentication, authorization, and Mobile IP registration processes as defined in [10]. The architecture in Fig. 3 shows the existing security associations along with the required MIP security associations so that the FN will be able to deliver services to a roaming MT. We use Extensible Authentication Protocol (EAP) over Diameter for end-to-end mutual authentication between an MT and its home AAA server (AAAH). When the MT roams into the FN domain, the authentication, authorization, and MIP registration are carried out as described below. The signaling messages for this procedure are shown in Fig. 4. Here, we use EAP-SIM [15] to illustrate the authentication process. Note that any other authentication scheme (e.g., EAP-AKA, EAP-SKE, EAP-TLS) can also be used.

1) When the MT hears MIP *Agent Advertisement* (*step 1*), it sends an MIP Registration Request including *Mobile-AAA Authentication and Authorization* extensions (as defined in [16]) to the FA located in the IG (*step 2*). The MT also includes a SIM Key Request extension [17] and an NAI [18] (e.g., MT@relam) in its MIP Registration Request.

2) When the FA receives the MIP Registration Request and finds the Mobile-AAA authentication and authorization extensions, it learns that the MT is a roaming user and forwards the MIP Registration Request to the authentication unit in the IG (AU_IG) (*step 3*). Based on the NAI in the MIP Registration Request, the AU_IG recognizes that the FN does not have a direct SLA with the HN of the MT and forwards the MIP Registration Request to the AU_NIA, either directly or through other AAA proxies (*step 4*).

3) The AU_NIA examines the NAI of the received MIP Registration Request message and forwards it to the Home AAA server (AAAH). Once the AAAH receives the MIP Registration Request, it first verifies the Mobile-AAA authentication and authorization extensions. If these extensions are valid, it contacts the home authentication center (AuC) of the MT. The AuC generates *n* triplets (RAND, SRES, K_c), where RAND denotes a random number, SRES denotes the response, and K_c is the key used for encryption. Then the AuC calculates message authentication code (MAC) for the RANDs (MAC_RAND) as defined [17]. The AuC sends the RANDs and MAC_RAND to the AAAH, which forwards those to the AU_NIA (all these constitute *step 5*). Then, the AU_NIA forwards these to the AU_IG (*step 6*). Finally, the AU_IG forwards these to the FA (*step 7*). The FA sends an MIP Registration Reply message to the MT containing the RANDs and MAC_RAND (*step 8*). The MT derives the corresponding SRES and K_c values using its SIM card and the received RANDs. It also calculates MAC_RAND and validates the authenticity of RANDs by comparing the calculated MAC_RAND with the received MAC_RAND, thus confirming that the RANDs are generated by the HN. If the MAC_RAND is valid, the MT calculates an MAC for its SRES values as defined in [17] (*step 9*). The MAC_SRES is used by the AuC to know if the SRES values are fresh and authentic. The MT also generates security association keys: K_{MT_FA} for the FA and K_{MT_HA} for the HA as defined in [17]. These keys are used to authenticate subsequent MIP registrations until the key lifetime expires.

4) Now, the MT sends another MIP Registration Request message to the FA containing *SRES* extension [17] and Mobile-AAA authentication and authorization extensions (*step 10*). When the FA detects the presence of Mobile-AAA authentication and authorization extensions, it forwards the MIP Registration Request message to the AU_IG (*step 11*), which forwards it to the AU_NIA (*step 12*). The AU_NIA forwards the MIP Registration Request message to the AAAH (*step 13*). After successful authentication and authorization (this may require the interaction of the AAAH and AuC), the AAAH forwards the MIP Registration Request to the HA (*step 14*) containing K_{MT_HA} security key. The HA carries out the registration for the MT as defined in [13], extracts the K_{MT_HA} key, and sends MIP Registration Reply to the AAAH (*step 15*). The AAAH forwards the MIP Registration Reply (containing K_{MT_FA} and the K_c keys) to the AU_NIA (*step 16*). Then the AU_NIA forwards the MIP Registration Reply to the AU_IG (*step 17*). The AU_IG forwards it to the FA

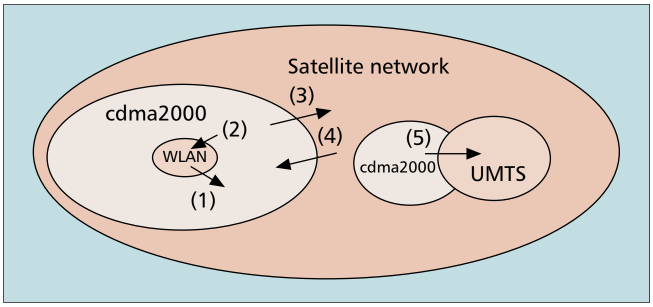

FIGURE 5. *Next-generation integrated systems scenario.*

(*step 18*). The FA extracts K_{MT_FA} and K_c keys and sends an MIP Registration Reply to the MT (*step 19*). The K_c keys are used for secure data transfer between the MT and the FA, providing confidentiality and integrity to the data traffic. If necessary, a FA-HA security association key can be generated by the AuC as defined in [17] and distributed to the FA and HA as a part of authentication process.

Billing

Once the MT is authenticated and authorized by the FN, the *accounting unit* of the IG (ACU_IG) maintains a per user accounting record based on the charging policy of the FN provider (connection duration, amount of data transferred, etc.). It transfers the accounting information either on a per session basis or in real time to the AAAL server of the FN domain. The AAAL server collects and consolidates the accounting information for the MT and forwards it as FN access call detail records (FN CDRs) to the accounting unit of the NIA (ACU_NIA). The NIA is capable of interpreting FN CDRs. However, it may happen that the HN of the MT supports a different CDR format. Then the NIA first converts the FN CDR format to the CDR format supported by the HN and forwards the final CDRs to AAAH for billing purposes. ACU_NIA is responsible for the interoperation of different billing schemes supported by different network providers.

Intersystem Handover Protocols

Two types of intersystem roaming may arise in AMC. They are:

- Roaming between fully overlapping systems, which can be further classified as:
  –Roaming from a lower-tier system to a higher-tier system (e.g., (1) and (3) in Fig. 5)
  –Roaming from a higher-tier system to a lower-tier system (e.g., (2) and (4) in Fig. 5)
- Roaming between partially overlapping systems (e.g., (5) in Fig. 5)

Note that a lower-tier system supports greater bandwidth than a higher-tier system.

When any of the above intersystem roaming types occurs, intersystem handover (ISHO) is carried out. ISHO is also referred as *vertical handoff*. It is essential that the applications running on an MT remain unaware of the roaming to ensure uninterrupted services with minimum QoS degradation. This can be achieved by reducing the handoff failure probability and latency to values tolerable by the applications.

Several issues need to be addressed during vertical handoff. When an MT is accessible through multiple fully overlapping systems, first, based on the service needs of a user, the best communication network should be determined. Then the handoff initiation time is determined to guarantee successful intersystem roaming. The authentication, authorization, and accounting procedures are then carried out before the MIP registration process.

Best Network Selection

The NIA helps each MT to be "always best connected" [2] by selecting the best available network for communications. Several factors influence the design of policies on the best network selection for vertical handoff. Monetary cost, network conditions, power consumption, user activity history, and the required QoS from applications are considered as the decision metrics. Moreover, the best network selection also affects the distribution of the overall system load.

We develop a hybrid network selection scheme that combines terminal-based and network-based selection mechanisms. The terminal-based mechanism allows MTs to periodically collect dynamic network conditions and determine the best reachable network for handoff by themselves. The network-based mechanism makes a globally optimized selection and achieves load balancing for the whole system. The objective of the proposed scheme is to provide satisfactory overall performance of the whole system as well as take into account user preferences. It is a two-level decision-making scheme. At the first level, each MT monitors and collects the dynamically varying network conditions for decision making at the terminal side. At the second level, the *handover management unit* inside the NIA finds the optimal user distribution for each individual network based on global observations. The decision made by this central controller is fed back to the first-level decision as adjustments. The details of the proposed best network selection scheme is in [19].

Handoff Initiation Time Estimation

After the best available network is selected, the next challenge is to determine the right time to start handoff procedures. Currently, there are several proposals that use physical and MAC layer sensing to determine the appropriate time for vertical handoff initiation. In these algorithms, the implicit assumption is that the signaling delay associated with vertical handoff is constant. Based on this assumption, these algorithms initiate the vertical handoff when the received signal strength (RSS) of the serving network goes below a certain fixed threshold value, S_{th}. However, in a real scenario, the vertical handoff signaling delay varies from a few seconds to several tens of seconds depending on several factors (e.g., traffic level in the backbone network, wireless link quality, and distance between the user and its home network). Therefore, protocols designed based on a fixed vertical handoff signaling delay have poor performance.

We propose the use of a dynamic RSS threshold to eliminate the effect of signaling delay variation. We predict the handoff signaling delay in advance. Moreover, we estimate the speed of MTs. Then we determine a dynamic threshold value for the RSS, S_{dth}, based on the handoff signaling delay and speed information such that if the vertical handoff procedures are initiated when the RSS of the serving network goes below S_{dth}, they are completed before the user moves out of the coverage area of the serving network. The *seamless roaming module* in the IG implements the algorithm for the estimation of S_{dth}. Details of this proposed scheme is in [20].

ISHO Protocols for Fully Overlapping Systems

ISHO Protocols for Lower- to Higher-Tier Roaming — When an MT is moving out of the coverage area of a lower-tier system, the goal is to switch it to the overlaying higher-tier system before the lower-tier link breaks. We describe the associated mobility management protocols using Fig. 6.

The MT first enables its interfaces for the higher-tier systems and determines the best network to be handed off to (*step 1*). When the handoff initiation time is determined, it

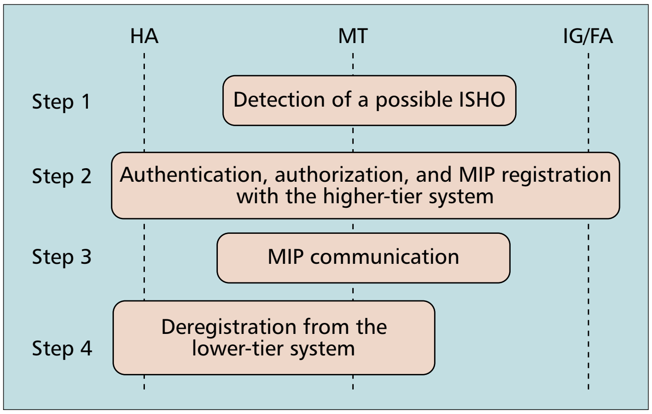

FIGURE 6. *Steps for intersystem roaming.*

registers with the higher-tier system using MIP [13] registration procedures. Authentication and authorization procedures are combined with MIP registrations as discussed earlier (*step 2*). The MT also maintains its registration with the lower-tier system using simultaneous mobility binding [14] with both systems.

After successful registration with the higher-tier system, the MT uses both the lower- and higher-tier systems for downlink traffic, but uses only the lower-tier system for uplink traffic as long as it is within the coverage area of the lower-tier system to take advantage of the higher data rate of the lower-tier system (*step 3*). With an established connection with the higher-tier system, the ongoing communications of the MT can be immediately switched to the higher-tier system when it moves out of the lower-tier one. This ensures a seamless ISHO. Once it moves out of the coverage area of the lower-tier system, it uses only the higher-tier system for its communications (*step 4*).

ISHO Protocols for Higher- to Lower-Tier Roaming — A higher-tier system completely overlaps a lower-tier system. Therefore, when an MT roams from a higher-tier system to a lower-tier system, the MT can always keep its connection with the higher-tier system to ensure no connection loss.

The MT initiates an ISHO by sending an MIP Registration Request message to the FA located in the IG of the corresponding lower tier system. The FA determines that the MT is a roaming user and starts the process of ISHO by forwarding the MIP Registration Request message to the NIA through AU_IG. The NIA determines if the MT has the permission to access the lower-tier system using its *operators database* as discussed previously. If the outcome is yes, the NIA proceeds with the MIP registration process along with authentication and authorization as discussed earlier (*step 2*). After successful registration with the lower-tier system, the MT starts communicating through the lower-tier system and deregisters from the higher-tier one (*steps 3 and 4*).

ISHO Protocols for Partially Overlapping Systems

In case of adjacent systems, when the MT detects that it is moving out of the coverage of the serving system, it enables the interfaces and searches for an available system (*step 1*). When it finds the new system, it registers with that system using MIP [13] registration procedures. Authentication and authorization procedures are combined with MIP registration as discussed earlier (*step 2*). The MT also maintains its registration with the old system using simultaneous mobility bindings to both the systems for a predefined period of time to avoid ping-pong effect during the ISHO. After successful registration with the new system, it uses both the old and new systems for downlink traffic. It uses only the new system for uplink traffic (*step 3*). After the specified time period, if it does not move back to the old system, it deregisters from the old system and uses only the new system for its communications (*step 4*).

Characteristic Features of AMC

In this section we qualitatively evaluate our proposed architecture, AMC, in the context of the design goals stated earlier.

Economical — AMC uses the access and core network infrastructure of the existing wireless systems. It does not require any change to the infrastructure of the existing networks. AMC achieves integration of heterogeneous networks by adding only one new entity, the integration gateway (IG), to the individual networks. Hence, it is economical.

Scalability — AMC can integrate any number of wireless systems of different providers who may not have SLAs among them by using the NIA as the third party. Therefore, it is scalable. To further enhance the scalability, we propose the hierarchical NIA structure to integrate the wireless networks globally. In this hierarchical structure, wireless networks of various providers are integrated at the regional (e.g., city) level through the first-tier NIAs. These regional NIAs of a particular country or several countries are then integrated through the second-tier NIAs, followed by integration of the second-tier NIAs through the third-tier NIAs to realize global integration. The exact number of tiers and the number of NIAs at each tier depend on several factors, such as the number of network providers in that tier and the number of roaming users. Determination of the number of NIAs required for a particular deployment scenario can be carried out. This is beyond the scope of this article.

These NIAs can be owned by a single operator or multiple operators with SLAs among them. Note that the number of NIA operators is small. Hence, the required SLAs among NIA operators is only a few. Therefore, the scalability of AMC is not compromised when multiple operators own NIAs. An NIA operator is responsible for aspects of heterogeneous wireless system integration of a particular region and supports their interworking with other wireless systems globally through the establishment of SLAs with other NIA operators.

In this hierarchical NIA structure, a network operator only needs to have SLAs with a set of nearby first-tier (a.k.a. regional) NIA operators to be able to provide its subscribers with global access.

The NIA is involved only during the ISHO process and transfers the control signals between two systems. Once the ISHO is over, the data traffic of the roaming users does not go through the NIA as discussed previously. Therefore, the load on the NIA is limited.

Transparency to Heterogeneous Access Technologies — By using IP as the common interconnection protocol in AMC, mobile users may roam among multiple wireless networks in a manner that is completely transparent to different radio technologies.

Security — AMC adopts state-of-the-art security mechanisms such as SIM to provide security and privacy equivalent to existing wireless networks.

Seamless Mobility Support — AMC supports seamless intersystem mobility using Mobile IP (MIP) as the mobility management protocol. We further improved the performance of ISHO by using a dynamic RSS threshold for ISHO initiation. This reduces latency and packet loss during ISHO.

Future Work and Conclusion

We plan to carry out the following tasks to demonstrate the performance of our proposed architecture, AMC:

- **Resource management in integrated wireless systems**: In integrated wireless systems, managing resources among multiple wireless networks in an economical and computationally practical way is crucial to preserving system robustness. If each network manages its resources and reacts to congestion independently, it can lead to instability. How to fairly and efficiently share network resources among multiple networks is important to integrating multiple networks together. This could be achieved by efficient algorithms implemented inside the handover management unit of the NIA.
- **QoS provisioning**: QoS provisioning in heterogeneous wireless networks introduces new mobility management problems such as timely service delivery and QoS negotiations during ISHO. Both the handover management unit in the NIA and the mobility management unit in the IG could be involved for QoS provisioning.
- **Determination of the number of NIAs**: As discussed earlier, the exact number of NIAs at each tier depends on several factors. Determination of the optimum number of NIAs required can be carried out for each tier.

In this article we present AMC, a third-party-based architecture, to integrate heterogeneous wireless systems. The design goals of AMC are cost, scalability, transparency, security, and seamless mobility support. AMC reduces the cost of architecture deployment by using the access and core network infrastructures of existing wireless systems. We show how AMC can integrate heterogeneous wireless systems of different operators who may not necessarily have direct service level agreements among them. Furthermore, security equivalent to the existing wireless systems is achieved in AMC. Finally, intersystem handover is implemented to achieve seamless roaming.

Acknowledgments

This work is supported by the National Science Foundation under contract ANI-0117840.

References

[1] S. Y. Hui and K. H. Yeung, "Challenges in the Migration to 4G Mobile Systems," *IEEE Commun. Mag.*, vol. 41, no. 12, Dec. 2003, pp. 54–59.

[2] E. Gustafsson and A. Jonsson, "Always Best Connected," *IEEE Wireless Commun.*, vol. 10, no. 1, Feb. 2003, pp. 49–55.

[3] "Inter-PLMN Backbone Guidelines," GSM Assn. classifications, v. 3.4.0, Mar. 2003.

[4] V. Garg and J. E. Wilkes, "Interworking and Interoperability Issues for North American PCS," *IEEE Commun. Mag.*, vol. 34, no. 3, Mar. 1996, pp. 94–99.

[5] S. Ghaheri-Niri and R. Tafazolli, "Cordless-cellular Network Integration for the 3rd Generation Personal Communication Systems," *Proc. IEEE VTC '98*, vol. 1, 1998, pp. 402–08.

[6] F. D. Priscoli, "Interworking of a Satellite System for Mobile Multimedia Applications with the Terrestrial Networks," *IEEE JSAC*, vol. 17, no. 2, Feb. 1999, pp. 385–94.

[7] M. Buddhikot *et al.*, "Design and Implementation of a WLAN/cdma2000 Interworking Architecture," *IEEE Commun. Mag.*, vol. 41, no. 11, Nov. 2003, pp. 90–100.

[8] I. F. Akyildiz and W. Wang, "A Dynamic Location Management Scheme for Next Generation Multitier PCS Systems," *IEEE Trans. Wireless Commun.*, vol. 1, no. 1, Jan. 2002, pp. 178–89.

[9] P. J. Havinga *et al.*, "The SMART Project: Exploiting the Heterogeneous Mobile World," *Proc. 2nd Int'l. Conf. Internet Comp.*, Las Vegas, NV, June 2001, pp. 346–52.

[10] S. Glass *et al.*, "Mobile IP Authentication, Authorization, and Accounting Requirements," IETF RFC 2977, Oct. 2000.

[11] J. Rosenberg *et al.*, "SIP: Session Initiation Protocol," IETF RFC 3261, June 2002.

[12] I. F. Akyildiz, J. Xie, and S. Mohanty, "A Survey of Mobility Management in Next-generation all-IP-based Wireless Systems," *IEEE Wireless Commun.*, vol. 11, no. 4, Aug. 2004, pp. 16–28.

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

[14] R. Wakikawa *et al.*, "Multiple Care-of Addresses Registration," Internet draft, draft-wakikawa-mobileip-multiplecoa-03.txt, June 2004, work in progress.

[15] H. Haverinen and J. Salowey, "EAP SIM Authentication," Internet draft, draft-haverinen-pppest-eap-sim-16.txt, Dec. 2004, work in progress.

[16] C. Perkins, P. R. Calhoun, and J. Bharatia, "Mobile IPv4 Challenge/Response Extensions (Revised)," Internet draft, draft-ietf-mip4-rfc3012bis-03.txt, Dec. 2004, work in progress.

[17] H. Haverinen, N. Asokan, and T. Maattanen, "Authentication and Key Generation for Mobile IP Using GSM Authentication and Roaming," *Proc. IEEE ICC 2001*, vol. 8, 2001, pp. 2453–57.

[18] P. Calhoun and C. Perkins, "Mobile IP Network Access Identifier Extension for IPv4," IETF RFC 2290, Mar. 2000.

[19] J. Xie and I. F. Akyildiz, "A Hybrid Control Resource Allocation Scheme for Policy-enabled Handoff in Wireless Heterogeneous Overlay Networks," submitted for publication, 2004.

[20] S. Mohanty and I. F. Akyildiz, "A Cross-Layer (Layer 2 + 3) Mobility Management for Next Generation Wireless Systems," submitted for publication, Mar. 2004.

Biographies

IAN F. AKYILDIZ [F'95] (ian@ece.gatech.edu) is the Ken Byers Distinguished Chair Professor with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, and director of the Broadband and Wireless Networking Laboratory. He is Editor-in-Chief of *Computer Networks* (Elsevier) and *Ad Hoc Networks Journal* (Elsevier). His current research interests are in sensor and wireless networks. He is an ACM Fellow (1996). He served as a National Lecturer for ACM from 1989 until 1998 and received the ACM Outstanding Distinguished Lecturer Award for 1994. He received the 1997 IEEE Leonard G. Abraham Prize (IEEE Communications Society) for "Multimedia Group Synchronization Protocols for Integrated Services Architectures" in *IEEE Journal of Selected Areas in Communications* in January 1996, the 2002 IEEE Harry M. Goode Memorial Award (IEEE Computer Society) with the citation "for significant and pioneering contributions to advanced architectures and protocols for wireless and satellite networking," the 2003 IEEE Best Tutorial Award (IEEE Communicatons Society) for his article entitled "A Survey on Sensor Networks" (*IEEE Communications Magazine*, Aug. 2002), and the 2003 ACM SIGMOBILE award for his significant contributions to mobile computing and wireless networking.

SHANTIDEV MOHANTY (shanti@ece.gatech.edu) received his B.Tech. (Hons.) degree from the Indian Institute of Technology, Kharagpur, and his M.S. degree from Georgia Institute of Technology, Atlanta, in 2000 and 2003, respectively, both in electrical engineering. He is currently a graduate research assistant with the Broadband and Wireless Networking Laboratory and a Ph.D. candidate at the School of Electrical and Computer Engineering, Georgia Institute of Technology. His current research interests include wireless networks, mobile communications, mobility management, ad hoc and sensor networks, and cross-layer protocol design. From 2000 to 2001 he worked as a mixed signal design engineer for Texas Instruments, Bangalore, India. He worked as a summer intern for Bell Labs, Lucent Technologies, Holmdel, New Jersey, during the summers of 2002 and 2003, and for Applied Research, Telcordia Technologies, Piscataway, New Jersey, during the summer of 2004.

JIANG XIE (jxie1@uncc.edu) received a B.E. degree from Tsinghua University, Beijing, China, in 1997, an M.Phil. degree from Hong Kong University of Science and Technology in 1999, and M.S. and Ph.D. degrees from Georgia Institute of Technology in 2002 and 2004, respectively, all in electrical engineering. She is currently an assistant professor with the Department of Electrical and Computer Engineering at the University of North Carolina-Charlotte. She was a research assistant in the Broadband and Wireless Networking Laboratory at Georgia Institute of Technology from August 1999 to April 2004. Her current research interests include resource and mobility management of wireless networks, QoS provisioning, and next-generation Internet.