A Service-Agent-Based Roaming Architecture for WLAN/Cellular Integrated Networks

Minghui Shi, Humphrey Rutagemwa, *Student Member, IEEE*, Xuemin Shen, *Senior Member, IEEE*, Jon W. Mark, *Life Fellow, IEEE*, and Aladdin Saleh, *Senior Member, IEEE*

***Abstract*—In this paper, an agent-based integrated service model for wireless local area network (WLAN)/cellular networks and the relevant authentication and event tracking for billing support schemes are proposed. The service model does not require inefficient peer-to-peer roaming agreements to provide seamless user roaming between the WLAN hotspots and the cellular networks, which are operated by independent wireless network service providers. The proposed authentication and event-tracking schemes take the anonymity and intractability of mobile users into consideration and operate independently so that the integrated billing service can be applied to the cellular network, even if it still uses a traditional authentication scheme. Security analysis and overhead evaluation are given to demonstrate that the proposed service model and the supporting schemes are secure and efficient.**

***Index Terms*—Authentication, roaming, service agent (SA), wireless local area network (WLAN)/cellular integrated network.**

I. Introduction

WIRELESS local area network (WLAN) products have become the de facto standard component in mobile devices. Many wireless Internet service providers (WISPs), such as AT&T Wireless, GRIC, iPass, Surf and Sip, and STSN, have set up, are setting up, and will be setting up more and more WLAN hotspots in airports, cafes, bookstores, etc. [1]. With the WLAN products continuing to grow in popularity across both product categories and geographic regions, revenue is expected to hit $3.6 billion by 2008 [2]. In addition, mobile devices having both cellular phone and WLAN capability are available [1]. The demand to integrate multiple mobile computing services into a single entity is preeminent.

Manuscript received July 30, 2006; revised December 15, 2006 and January 13, 2007. This work was supported by a grant from the Bell University Laboratories (BUL). The review of this paper was coordinated by Dr. J. Misic.

M. Shi is with the Centre for Wireless Communications, Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada, and also with McMaster University, Hamilton, ON L8S 4K1, Canada.

H. Rutagemwa, X. Shen, and J. W. Mark are with the Centre for Wireless Communications, Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada.

A. Saleh is with the High-Speed Wireless Access, Wireless Technology Department, Bell Canada, Montreal, QC H3B 4Y8, Canada, and also with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2007.900525

Depending on the interdependence between the WLAN and the cellular network, their coupling can be tight or loose. In tight coupling, the 802.11 network appears to the 3G core network as another 3G access network. In loose coupling, the 802.11 gateway connects to the Internet and does not have any direct link to the 3G network elements such as packet data serving nodes (PDSNs), gateway GPRS support nodes (GGSNs), or 3G core network switches. Generally, a loosely coupled integrated network architecture, as shown in Fig. 1, is preferred [3] due to its flexibility, and it is the major network topology that is considered in this paper.

In Fig. 1, the WLAN and 3G networks work in a complementary way, and the interworking between the WLAN and 3G cellular networks is through an IP network. Mobile IP (MIP) protocol [4], [5] is applied to support the IP mobility for mobile clients. Data can be transported along two different paths, as shown in dashed lines in the figure, depending on the access interface, with which the mobile terminal (MT) is associated. With this loosely coupled architecture, the WLAN hotspot and 3G cellular networks can be operated by different service providers.

The main parts of a 3G network owned by different service providers are the radio access network and the core network. The network has two new functional entities: the Home Agent (HA) and the Foreign Agent (FA), which are specifically added to support the MIP. The PDSN is modified to act as a FA for the cellular network in addition to its originally intended functionality. The WLANs are connected through an IP network (or the Internet) to the 3G core networks. Each WLAN has its own gateway which serves as a FA for mobile users within its coverage. The MTs are provided with MIP clients and can support both 802.11 and 3G access technologies. With the MIP protocol, the IP packets addressed to an MT will be forwarded by its HA at the 3G core network, which is called IP tunneling.

However, for the cellular/WLAN integrated service to be deployed, there are several critical issues in service and security in terms of function, practice, and privacy.

1) No seamless user authentication for the WLAN hotspot: Although user roaming is well defined in the cellular network through authentication, authorization, and accounting (AAA), it is still an open issue in the WLAN networks operated by multiple service providers. Many WISPs provide public WLAN Internet access at the hotspots using a network access server (NAS). The NAS

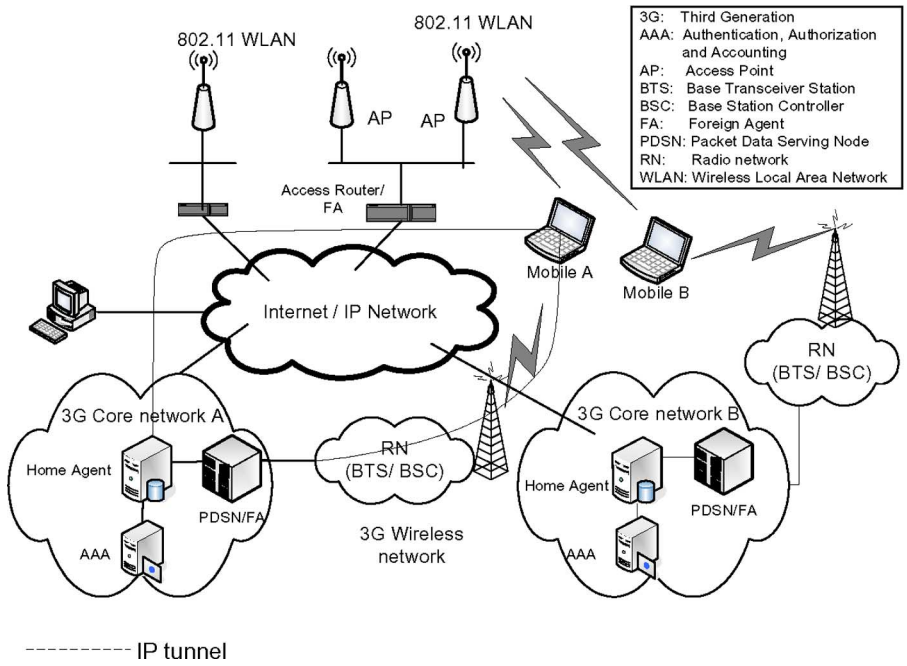

Fig. 1. Loosely coupled 802.11 WLAN and 3G cellular networks.

allows only legitimate customers to use the service and provides intradomain roaming because the hotspots from one WISP share the same customer base. However, it lacks an architecture to provide interdomain roaming and MIP support. Currently, multiple accounts for those service providers are required for a user to use the service in corresponding network territories. Due to the manual interaction between the users and a log on Web page, seamless network service offering is not available. On the other hand, since the network structure of the cellular network is quite different (much more complex and expensive) from the WLAN hotspot, it is difficult to import the authentication scheme [6], [7] used in the cellular network to the WLAN hotspot.

2) Inefficient peer-to-peer roaming service agreement: The number of WISPs is much larger than the number of cellular networks, mainly because of the simplicity and low cost of setting up a WLAN hotspot. For example, each wireless enabled router at home can be considered as a WLAN hotspot. The peer-to-peer service agreement is exercised by the cellular network service providers quite well. However, it is not practical in the WLAN case. Assuming that there are m WISPs, $m(m-1)/2$ agreements in total and $(m-1)$ agreements for each WISP are required to achieve roaming within the network.
3) Difficulty of having a universal roaming service: It is difficult for a service provider to have a roaming agreement with every network service provider. The issue is fine in using the cellular network, since cellular networks overlap each other, and a user is very likely to be able to find a usable one. However, WLAN hotspots do not overlap with each other, and this can result in a situation where the user cannot use the service in some hotspots. Therefore, universal roaming service cannot be achieved.
4) The privacy of user identity: Disclosure of a user identity may also allow unauthorized entities to track his moving history and current location. Any illegal access to information related to the user's location without his knowledge can be a serious violation of privacy. The anonymity and intractability of user identity are not considered in the current wireless communication architecture.

There has been extensive research related to WLAN/cellular network interworking. Most of it focuses on proposing integration architecture [8], modifying network components, such as gateway [9], and analyzing the switching performance of integrated service [10]. However, most solutions are suitable for the service providers who operate their own WLAN and cellular networks. In [11], an authentication scheme via a secured Web page for login over a hypertext-transport-protocol-secured connection is proposed. This type of solution requires interactions from the user and cannot be used in a seamless roaming service. In [12], a roaming and authentication framework for the WLAN/cellular network integration is proposed. However, the anonymity and intractability of user identity, which were not included in [12], has become an important security property for roaming services.

In this paper, a novel wireless/cellular network integrated service model is proposed. Under the coordination of a service agent (SA), an integrated wireless network service can be offered by independent WLANs and cellular networks, which are not bound by peer-to-peer roaming agreements. Therefore, many small WLAN service providers are able to join the integrated networks, which could greatly increase integrated network service coverage. On the other hand, the end users do not have to be customers of major wireless network operators, and they can purchase the service from the SA instead. The corresponding advantages include reduced support cost, more communication convenience, and higher revenue. The proposed

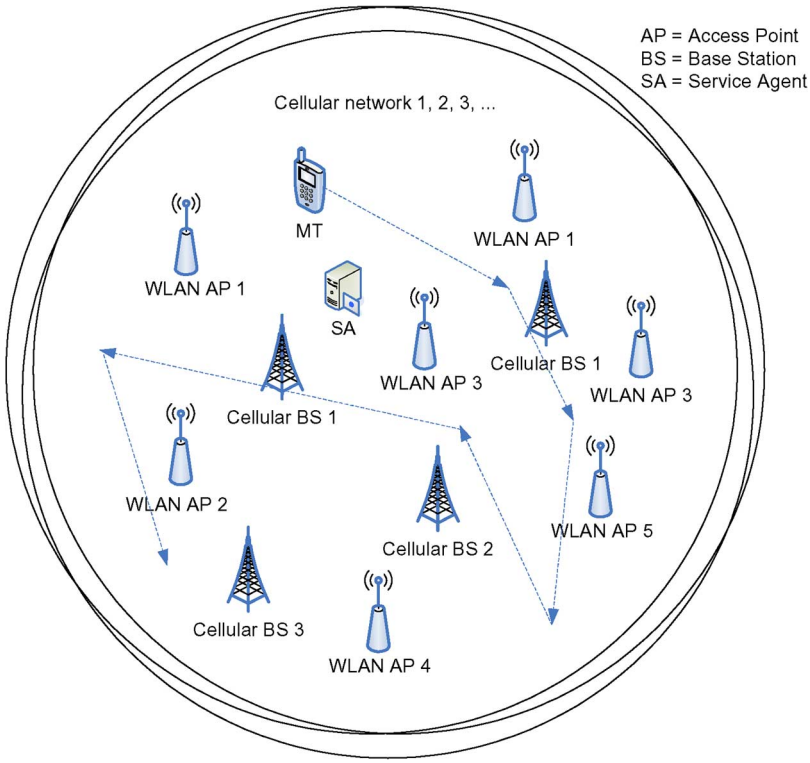

Fig. 2. Typical cellular network and WLAN hotspot deployment. The numbers in the cellular network BS and the WLAN AP labels are coded based on their network operators. SA is the proposed additional network component.

authentication scheme for integrated network service can be elastically applied to both the WLAN and cellular networks or to WLAN alone. It also effectively reduces the extra overhead caused by the SA. It is self-adaptive to the various authentication scenarios due to different ownership of WLAN hotspots. The supporting processes, such as user authentication for roaming and event tracking for billing support schemes, are operated in parallel with the authentication process, such that they can be flexibly deployed to those which do not adopt the proposed authentication scheme, such as the cellular network. In addition, the proposed schemes protect the mobile users' privacy by incorporating the user anonymity and intractability feature, so that the users' real identity is not revealed, and his/her mobility cannot be traced.

The rest of this paper is organized as follows. In Section II, an overview of the proposed service model architecture for cellular/WLAN integration is presented. In Section III, the messaging scheme for the proposed service model is described in detail. Security analysis and overhead evaluation are given in Section IV followed by the conclusion in Section V.

II. WLAN/Cellular Integrated Service Model Architecture

Fig. 2 shows a typical deployment of cellular networks and WLAN hotspots in a coverage area. There are m cellular networks $m = 1, 2, 3, \ldots, n$. The WLAN hotspots are randomly located to offer high-speed network/Internet access service within the area. Due to the disjoint deployment, WLAN hotspots only offer "stationary" wireless network access. Mobile network access service is offered by the WLAN/cellular network integrated infrastructure. The labels of WLAN AP (hotspot) and cellular BS are numbered by their network operators, i.e., service providers. Based on the current WLAN deployment strategy, which avoids unnecessary multiple hardware setup investment, it is assumed that there is only one service provider operating the WLAN service at one hotspot area. Therefore, there is no direct WLAN network-level handover anytime. A cellular network service provider usually operates the WLAN Internet access service at the same time. Let WLAN AP x, $x \in \{1, \ldots, m\}$ be linked with the cellular network x. For example, one service provider operates both cellular network 1 and WLAN hotspots numbered as WLAN AP 1. The rest of the WLAN APs are operated by third-party WISPs. A multimode MT can either connect to a WLAN hotspot or to a cellular network at the user's discretion.

We introduce an additional role, called SA, to the WLAN/cellular integrated network architecture to improve service flexibility and deal with the roaming agreement issue when the number of WLAN operators is large. The cellular network and WLAN are encouraged to have the one-for-all roaming agreement with the SA directly so that the cumbersome peer-to-peer roaming agreements are no longer needed. In the ideal case, one agreement per service provider can achieve universal user roaming. On the other hand, the one-for-all roaming agreement can coexist with the peer-to-peer roaming agreement. A practical strategy could be that a service provider optionally sets up a peer-to-peer agreement with a few major service providers for better performance and a one-for-all roaming agreement with the SA for universal roaming completion.

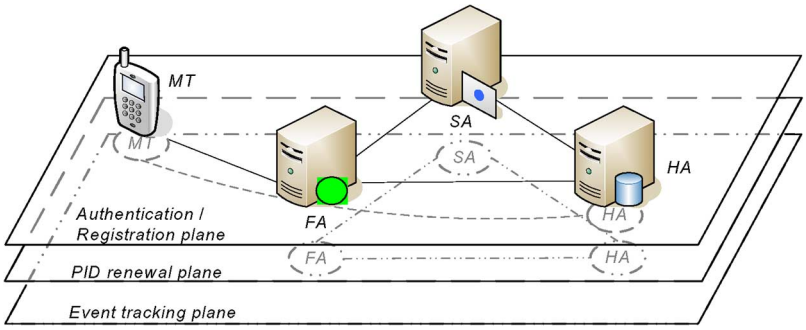

Fig. 3. Overview of service model function and messaging scheme. We do not study the behavior of individual BS or AP devices in this paper, but we focus on their networks (or service providers) as entities because of the nature of mobile-user account management.

In the proposed service model, the MT does not have to be a customer of any physical network operator. The SA can provide cellular/WLAN integrated service itself, which is driven by the following business motivation: The SA purchases bulk wireless service from physical network operators and sells wireless integrated service to the customers so that the network operators spend less on support cost for the end users; the customers receive an improved network access convenience, and the SA gets more profit.

In order to make a clear network operation layout, we abstract the real network parties in a roaming scenario into parties, as shown in Fig. 3. The network which houses the FA and is visited by the MT could be either a WLAN or a cellular network. The HA resides in the MT's home network, who knows everything about the MT, such as identify, shared secret key, etc. Fig. 3 also shows the functions and messaging schemes of the proposed service model, which consists of three layers: 1) authentication and registration plane; 2) pseudoidentifier (PID) renewal for privacy protection of user identity; and 3) event-tracking plane for interoperator billing. The technical details of these functions are given later in this section. The advantage of independent function layer design is to allow the event tracking, which is a key component for integrated service billing, to be easily and uniformly adopted by all network operators with few modifications, or information remapping of existing authentication process, particularly in cellular networks.

A. Authentication and Registration

Before the MT accesses the network service, authentication is performed to verify its legitimacy, and the MT is registered in the network if it is successfully authenticated. Meanwhile, the MT should avoid a rogue service provider via authentication as well. Since the authentication process in the cellular network is mature and has been working well, the proposed authentication scheme is mainly for the WLAN hotspots. It can also be adopted by the cellular network operators if a uniform authentication scheme is preferred. In the authentication plane shown in Fig. 3, the SA acts as an authority, which is trusted by all the parties, and assists the mutual verification of the FA and the HA. Considering the power of the MT and wireless transmission environment, the authentication scheme is designed such that most of the authentication process is executed by servers in the wired network. The messaging part, which directly involves with the MT, consists of two messages and uses symmetric cryptography only.

B. PID Renewal

The anonymity of user identity is achieved by a PID. An MT does not use its real identity for network authentication and registration all the time, including when it is roaming in foreign networks. The PID is refreshed periodically and whenever the MT sees the necessity, such as after each authentication session. Independent design of the PID renewal process allows the MT to decide when and how often the PID should be refreshed according to its preference for anonymity strength, wireless channel condition, and battery status. An option for PID refreshment policy can be included in the MT's preference menu.

C. Event Tracking

Successful commercial network service deployment cannot live without a proper billing mechanism. In the proposed WLAN/cellular integrated service model, billing is based on the MT's network accessing activities, which is tracked by a data structure named Event ID. An event ID is defined as an incident of the MT accessing the network resource and is distributed to proper network operators by the event-tracking process, which is independently running from the authentication process. The revenue is partitioned later based on the Event ID records.

III. Messaging Scheme

The messaging scheme for the proposed cellular/WLAN integrated service model includes an authentication scheme with variations, a PID renewal scheme, and an event-tracking scheme. The subsets inherited from the authentication process and the supplementary processes are developed to reduce the authentication latency, enhance the anonymity of user identity, and support billing. We will describe those functions in detail in the following sections. The common symbols, which will be used in the message exchanges thereafter, are shown in Table I.

TABLE I
DESCRIPTION OF SYMBOLS

Symbol	Description
ID_X	X's ID number or a unique global machine number
PID_i	ith pseudo identifier
k_i	ith session key
k_{XY}	Shared key between X and Y
Pub_X	X's public key
$E_k\{.\}$	Symmetric encryption using shared key k
$E_k\langle.\rangle$	Asymmetric encryption using public key k
Sig_X	Digital signature signed by X
$H(.)$	Hash function
$H^x(.)$	x times hash operations recursively
$F(.)$	One way function with defined bit length output
$\|$	Concatenation operation
$fg_{op.}$	A pre-defined fixed number mapped for *op*eration

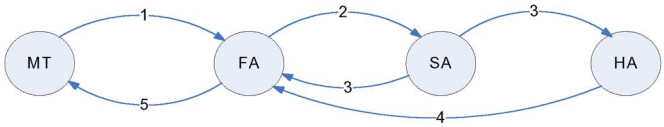

Fig. 4. General authentication message flow of the proposed authentication scheme.

A. Anonymity of User Identity

In the proposed scheme, the PID is derived from the user's real identity. Additional parameters are added to achieve dynamics and secrecy properties. It is renewed periodically and after each authentication. The series of PID is defined as

$$\mathrm{PID}_{j+x} = F\left(\mathrm{PID}_{j+x-1} \| H^x(k_i) \| k_{\mathrm{HM}}\right) \tag{1}$$

where k_i is the ith session key negotiated after each successful authentication, k_{HM} is the shared session key between the MT and the home network $j = 1, 2, \ldots$, and $x = 0, 1, 2, \ldots, F(\cdot)$ is a one-way function which generates the same bit length as PID. $H^x(\cdot)$ is a hash chain function which computes the hash value of $k_i x$ times and outputs the variable itself if $x = 0$. The function PID is used for renewal, which will be discussed in detail in Section III-D1. In the implementation, the MT keeps the results of the last hash operation; therefore, multiple hash operations are not required for each PID renewal. The initial value PID_0 is defined as the identity of the mobile user ID_M. The home network always keeps the mapping relationship of an MT's PID and real identity $(\mathrm{PID}_j \leftrightarrow \mathrm{ID}_M)$. Equation (1) conceals the real identity ID_M in PID_j and provides anonymity of the user identity for an MT without increasing the computational complexity.

B. Proposed Authentication Process

We first introduce the full version of the authentication process for the proposed SA based on the WLAN/cellular integrated service model followed by the description of the inherited subsets. Assume that an MT is accessing a WLAN hotspot (FA) for the first time. Fig. 4 shows the message flow of the proposed authentication scheme. The major authentication stages in the figure are described as follows: 1) The MT submits the authentication request and encapsulated session key negotiation to the FA; 2) the FA forwards the message from the MT to the SA and requests the HA for verification; 3) the SA returns the identity information of the HA and forwards the authentication message from the MT to the HA; 4) the HA verifies the identities of the MT and the FA and returns the session key to the FA; and 5) the FA sends the proof of its knowledge of the session key to the MT. The detailed authentication scheme proceeds as follows.

Step 1) (MT $\rightarrow$ FA) The MT sets the other party of the authentication to its home network and generates a random number as a nonce and a random number as a session key k_i, which can be used in encrypting the information in the communication session after the authentication is successfully processed. The MT can use its International Mobile Subscriber Identity [13] for ID_M and send Message (2) to the FA with current time stamp ts_1

$$M \rightarrow F : \mathrm{PID}_j \| \mathrm{ID}_H \| E_{k_{\mathrm{MH}}}\{\mathrm{ID}_M \| N_M \| k_i \| \mathrm{ts}_1\} \tag{2}$$

where $\mathrm{PID}_j \| \mathrm{ID}_H$ represents the modified AAA user name in the form of username@domain. An alternative form of the authentication request for the case in which the FA already has the information of the MT will be discussed in Section III-C1.

Step 2) (FA $\rightarrow$ SA) On receiving Message (2) from the MT, the FA sets Msg_M to $\mathrm{PID}_j \| \mathrm{ID}_H \| E_{k_{\mathrm{MH}}}\{\mathrm{ID}_M \| N_M \| k_i \| \mathrm{ts}_1\}$ and generates a serial number sn_i, which includes the FA's network operator code, station code, and an event sequence code for the authentication event. By attaching the current timestamp, the FA sends Message (3) to the SA

$$F \rightarrow S : \mathrm{ID}_F \| E_{k_{\mathrm{SF}}}\{\mathrm{Msg}_M \| \mathrm{sn}_i \| \mathrm{ts}_2\} : \mathrm{sig}_F. \tag{3}$$

Step 3) (SA $\rightarrow$ FA, HA) The SA checks the integrity of Message (3) by verifying ID_F and the attached digital signature of the FA. The SA decrypts Message (3) by k_{SF} and looks into Msg_M to get ID_H. Note that the SA cannot read $\mathrm{ID}_M \| k_i \| \mathrm{ts}$ in Msg_M, since it is encrypted by k_{MH}. The SA is aware that the authentication request is intended for the HA. The SA searches the database and fetches pub_F and pub_H. The SA sets Msg_F to $\mathrm{Msg}_M \| \mathrm{sn}_i \| \mathrm{ts}_2$. Besides forwarding the necessary information to the HA, the SA also acts as PKI authority and sends the certified public key information to both the FA and the HA, which will be used by them to verify each other. For publicly available information, such as public key, encryption is not necessary. Digital signature is used to guarantee the integrity. The SA sends Message (4) to the HA

$$S \rightarrow H : \mathrm{ID}_S \| \mathrm{ID}_F \| \mathrm{pub}_F \| E_{k_{\mathrm{SH}}}\{\mathrm{Msg}_F \| \mathrm{ts}_3\} : \mathrm{sig}_S \tag{4}$$

and Message (5) to the FA

$$S \rightarrow F : \mathrm{ID}_S \| \mathrm{ID}_H \| \mathrm{pub}_H \| \mathrm{ts}_3 : \mathrm{sig}_S. \tag{5}$$

Step 4) (HA → FA) The HA checks the integrity of Message (4) by verifying ID_S and the attached digital signature of the SA. The HA retrieves Msg_F, the identity and the public key of the FA, and PID_j by using the shared key k_{SH}. The HA searches ($\mathrm{ID}_M \rightarrow \mathrm{PID}_j$) the mapping database and locates the identity of the MT, who initiates the authentication process and the corresponding k_{MH}. The HA further decrypts Msg_H encapsulated in Msg_F by using k_{MH} and retrieves ID_M of the mobile user, the nonce N_M, and the proposed session key k_i. The HA checks the validity of timing relationships of all the timestamps $\mathrm{ts}_1, \ldots, \mathrm{ts}_3$ to prevent replay attack and verify the identity of the MT by comparing the decrypted ID_M and PID_j with the stored copy. Then, the HA sends Message (6) to the FA

$$H \rightarrow F : \mathrm{ID}_H \| E_{k_F} \langle N_M \| k_i \| \mathrm{PID}_j \| \mathrm{ts}_4 \rangle : \mathrm{sig}_H. \quad (6)$$

Step 5) (FA → MT) The FA should have received Message (5) from the SA sent in Step 3) and Message (6) from the HA in the aforementioned step. When the FA receives Message (5), it checks the timestamp ts_3 and the message integrity by verifying ID_S and the attached digital signature of the SA. The FA then stores the identity and the public key of the HA for later use if the verification is positive. When the FA receives Message (6), it checks the message integrity again by verifying ID_H and the attached digital signature of the HA. It also checks if the timestamp is reasonable. The FA decrypts the message using the public key of the HA retrieved from Message (5) and gets PID_j, the nonce N_M, and the session key k_i proposed by the MT in Step 1), which indicates that the HA has acknowledged the MT and approved its roaming privilege. Since the decryption of Message (6) depends on the reception of Message (5), the FA will temporarily store Message (6) if it is received before Message (5). The FA generates a ticket $Tk_i = Tk\mathrm{ID}_i \| \mathrm{ref}_i \| \exp_i$, where $Tk\mathrm{ID}_i$ denotes the ticket ID, ref_i denotes the ticket reference code, and $\exp_i$ denotes the ticket expiration time. The FA stores the mapping relationship $Tk_i \rightarrow k_i \rightarrow \mathrm{PID}_j$. The purpose of the ticket will be discussed in the next section. The FA sends Message (7) to the HA

$$F \rightarrow H : \mathrm{ID}_F \| E_{k_i} \{ H(N_M) \| N_F \| Tk_i \| \mathrm{ts}_5 \}. \quad (7)$$

Step 6) (MT → FA) After the MT receives Message (7), it checks the identity of the FA and tries to decrypt the message using k_i. If the decrypted $H(N_M)$ is correct and ts_5 is reasonable, it indicates that both the FA and the HA approve the roaming service request. The MT stores the $Tk_i \rightarrow k_i \rightarrow \mathrm{PID}_j$ mapping for later use. The MT sends Message (8) as the acknowledgment of receiving Message (7) and begins to send communication data to the network operated by the FA

$$M \rightarrow F : \mathrm{PID}_j \| E_{k_i} \{ H(N_F) \| \mathrm{ts}_6 \| \mathrm{comm} \}. \quad (8)$$

So far, the MT and the FA have been mutually authenticated. Current WLAN hotspots do not apply data encryption over the wireless link at all, because many WLAN security protocols, such as WEP, WPA, WPA2 (IEEE 802.11i), etc., are more inclined to individual users, and a centralized corporation environment can be used in securing the communication between the MT and the FA.

C. Variation of the Proposed Authentication Scheme

One of the key advantages of the proposed authentication scheme is its self-adaptation. We discuss two common alternatives in the following section: 1) The MT revisits the FA; and 2) the MT purchases the integrated service from the SA/existing peer-to-peer roaming agreement.

1) Localized Authentication: Localized authentication is defined as authentication between the MT and a network operator who has the MT's certain type of credential. One example of localized authentication is that the MT authenticates with the FA when the MT revisits the FA. The definition of "revisit FA" is that the MT visits any hotspot operated by the FA for the second time or more by discovering the hotspot from its beacon or service set identifier (SSID), which is constantly broadcasted. The hotspot is not necessary to be the same hotspot that the MT visited for the first time. Under this definition, occurrence of this "revisit" case would be quite common. Recall the ticket $Tk_i = Tk\mathrm{ID}_i \| \mathrm{ref}_i \| \exp_i$, which is introduced in Step 5) in the aforementioned section. Tk_i is the certificate issued by the FA to the MT to indicate that the MT has been verified by its home network and registered with the FA before. Since the storage of tickets requires extra memory space, the $\exp_i$ parameter defines the absolute time when the ticket is expired so that both the MT and the FA do not keep the ticket forever. If the MT revisits the FA after $\exp_i$ timeout, a full authentication process is required, as described in the aforementioned section. Otherwise, a subset of the proposed authentication scheme, which is composed by Steps 1), 5), and 6), is executed as follows:

$$M \rightarrow F : \mathrm{PID}_{j+c} \| Tk\mathrm{ID}_i \| E_{k_i} \{ \mathrm{ref}_i \| N_M \| k_{i+1} \| \mathrm{ts}_1 \} \quad (9)$$

$$F \rightarrow M : \mathrm{ID}_F \| E_{k_{i+1}} \{ H(N_M) \| N_F \| Tk_{i+1} \| \mathrm{ts}_5 \} \quad (10)$$

$$M \rightarrow F : \mathrm{PID}_{j+c} \| E_{k_{i+1}} \{ H(N_F) \| \mathrm{ts}_6 \| \mathrm{comm} \}. \quad (11)$$

From Message (9), the FA locates the corresponding $Tk_i \leftrightarrow k_i \leftrightarrow \mathrm{PID}_j$ mapping according to $Tk\mathrm{ID}_{i-1}$ and retrieves k_{i-1}. The FA then decrypts Message (9) and verifies whether ref_i matches $Tk\mathrm{ID}_i$ and if ts_1 is within the tolerant range. If it is true, then the FA accepts the new session key $k_i + 1$. The FA grants the access of the MT and sends Message (10) to the MT with a new ticket. The parameter in Message (9) is due to the PID renewal, which will be discussed in Section III-D1. Message (11) serves as an acknowledgment.

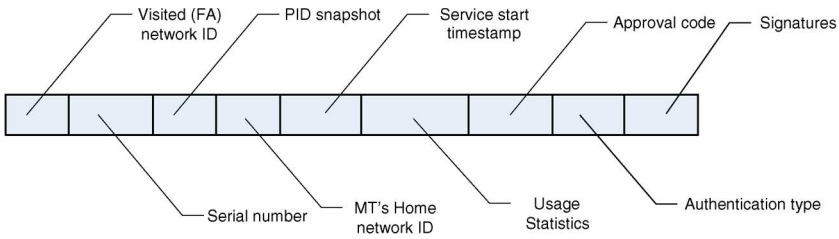

Fig. 5. Event ID data structure.

Note that, in this case, the FA assumes the good standing status of the MT. The HA should notify the FA to revoke the ticket once it finds out that the MT is no longer good. The most recent FA list that served the MT can be obtained by searching the Event IDs in the MT's service record.

Localized authentication also occurs in the following two cases: 1) The MT accesses a network (e.g., WLAN hotspot) operated by its home network when it is in the home network; and 2) the MT accesses WLAN hotspots operated by the visiting cellular network, which supports the ticket and the PID.

2) Integrated Service Offered by SA/Existing Peer-to-Peer Roaming Agreement: Both cases are actually equivalent. Therefore, we begin with the former, which is the other operational mode of the proposed network infrastructure. In this case, the MT does not belong to any real network operator, and it is registered with the SA, which is denoted as for notational convenience. Therefore, SA' is considered as residing in the MT's home network and takes over the work which is designated to the HA in Section III-B. Since the FA and the SA have direct service agreement and share preset secret key, the public key exchange in Step 3) is not necessary, and there is no need to do asymmetric encryption. The subset of the proposed authentication scheme, which is composed by all the steps except Step 3) with minor parameter modifications, is executed as follows:

$$M \rightarrow F : \mathrm{PID}_j \| \mathrm{ID}_S \| E_{k_{\mathrm{MS}}} \{ \mathrm{ID}_M \| N_M \| k_i \| \mathrm{ts}_1 \} \quad (12)$$

$$F \rightarrow S : \mathrm{ID}_F \| E_{k_{\mathrm{SF}}} \{ \mathrm{Msg}_M \| \mathrm{sn}_i \| \mathrm{ts}_2 \} : \mathrm{sig}_F \quad (13)$$

$$S \rightarrow F : \mathrm{ID}_S \| E_{k_{\mathrm{SF}}} \{ N_M \| k_i \| \mathrm{PID}_j \| \mathrm{ts}_4 \} : \mathrm{sig}_S \quad (14)$$

$$F \rightarrow M : \mathrm{ID}_F \| E_{k_i} \{ H(N_M) \| N_F \| Tk_i \| \mathrm{ts}_5 \} \quad (15)$$

$$M \rightarrow F : \mathrm{PID}_j \| E_{k_i} \{ H(N_F) \| \mathrm{ts}_6 \| \mathrm{comm} \} . \quad (16)$$

For the second case, we just need to replace SA by HA in the aforementioned messages. Note that the HA and the FA have shared secret key k_{HF} in this case.

D. Supplementary Operations

Two supplementary operations are developed in improving the anonymity of the user identity and commercial deployment: PID renewal and event tracking.

1) PID Renewal Process: Based on the algorithm in computing the PID_j series described in Section III-A, both the MT and the HA can compute the next PID independently since both of them always hold updated PID_j and $H^x(k_i)$. Therefore, the main purposes of the PID renewal message exchanges are PID synchronization and recovery in case the PID_j and/or $H^x(k_i)$ values that they are holding are out of synchronization for some reason. The following steps show the PID renewal operation in the normal case:

$$M \rightarrow H : \mathrm{PID}_j \| E_{k_{\mathrm{HM}}} \{ \mathrm{ID}_M \| fg_{\mathrm{pid.renew}} \} \quad (17)$$

$$H \rightarrow M : \mathrm{ID}_H \| E_{k_{\mathrm{HM}}} \{ H(k_{\mathrm{HM}}) \| \mathrm{PID}_{j+1} \} . \quad (18)$$

In the first step, the MT sends Message (17) to the HA. $\mathrm{fg}_{\mathrm{pid.renew}}$ tells the HA that the MT is going to renew its PID. After the message is received by the HA, the HA locates ID_M and the corresponding k_{HM} by searching the PID_j in the $\mathrm{PID}_j \rightarrow \mathrm{ID}_M$ mapping database. The HA then decrypts Message (17) and verifies if the decrypted ID_M matches the record. If it is true, then the HA sends Message (18) to the MT. The message includes PID_{j+1} computed by the HA. After the MT computes PID_{j+1} by itself, it verifies if both results match each other. If it is true, then the MT starts to apply the renewed PID. Otherwise, the MT will continue to use PID_j until proper acknowledgment is received from the HA. Meanwhile, the proposed renewal process is able to detect the exceptions, and the MT and the HA process the exceptional cases as follows.

1) Out-of-synchronization PID_j: The HA is unable to locate information in the mapping database. The HA then sends Message (19) instead of Message (18) to the MT

$$H \rightarrow M : \mathrm{ID}_H \| \mathrm{flag}_{\mathrm{pid.incor}} : \mathrm{sig}_H . \quad (19)$$

Since the MT will not apply a new PID until it receives correct acknowledgment from the HA, the PID at the HA must be PID_{j-1}.

2) Out-of-synchronization $H^x(k_i)$: PID_{j+1} in Message (17) does not match the MT's computed result, although $H(k_{\mathrm{HM}})$ shows that the message is sent by the HA. If either case happens, the MT sends Message (20) to the HA for the PID recovery process so that the HA is able to reconstruct PID_{j+1}

$$M \rightarrow H : \begin{cases} \mathrm{PID}_{j-1} \| E_{k_{\mathrm{HM}}} \{ \mathrm{ID}_M \| \mathrm{fg}_{\mathrm{pid.incor}} \| H^x(k_i) \} , & \text{case 1} \\ \mathrm{PID}_j \| E_{k_{\mathrm{HM}}} \{ \mathrm{ID}_M \| \mathrm{fg}_{\mathrm{hash.incor}} \| H^x(k_i) \} , & \text{case 2} \end{cases} . \quad (20)$$

The HA sends Message (18) to the MT again for acknowledgment, and the PID renewal process is completed.

2) Event Tracking: In the proposed WLAN/cellular integrated service model, billing is based on the MT's network accessing activities, which is tracked by a data structure named Event ID, as shown in Fig. 5. Most of the data fields have

TABLE II
EVENT ID DISTRIBUTION

Scenario	Auth. (Type)	Service provider	MT's home network	Service Coordinator	Event ID distribution
Normal (Section 3.2)	1	FA	HA	SA	FA, HA, SA
Re-visit (Section 3.3)	2	FA	HA	-	FA, HA
Service purchased from SA (Section 3.3)	3	FA	SA	-	FA, HA
Peer-to-peer roaming agreement (Section 3.3)	4	FA	HA	-	FA, HA

either appeared before in the literature or are self-explanatory, so they do not need to be introduced again. In the Event ID data, the usage statistic field includes the MT network activity data such as connection time, upload/download traffic bandwidth, zone and/or service weight, etc. How to adopt these parameters into billing generation solely depends on the roaming agreements and is beyond the scope of this paper. Approval code is a reference number indicating that the event claim is approved by the MT's home network. The attested signatures from the involved network operators show that they agree with those data.

Since different authentication modes have been developed in Sections III-B and C, the distribution of Event ID varies, as shown in Table II. For example, in the "revisit" case, the SA is not involved, and it is unnecessary to send the Event ID to the SA. However, the HA still requires the Event ID for billing purposes.

Based on the nature of billing mechanism, we classify the network operators into three categories: (roaming) service provider, home network, and (roaming) service coordinator. Table II summarizes the roles in the four scenarios. During the authentication process, the network operators keep the received messages at least until the Event ID distribution is complete. Event tracking is implemented as follows.

For the "normal" case, after the SA completes Step 3) in the authentication process, it constructs the Event $\mathrm{ID}_i^{\mathrm{SA}}$ as

$$\mathrm{ID}_S \| \mathrm{sn}_i \| \mathrm{PID}_j \| \mathrm{ID}_H \| \mathrm{ts}_3 \| \mathrm{null} \| \mathrm{null} \| 1 : \mathrm{sig}_S \tag{21}$$

where null indicates not applicable. The SA sends Message (22) to the HA and the FA, respectively, as

$$\begin{aligned} S \rightarrow H &: \mathrm{ID}_s \| E_{k_{\mathrm{SH}}} \left\{ \text{Event ID}_i^{\mathrm{SA}} \right\} \\ S \rightarrow F &: \mathrm{ID}_s \| E_{k_{\mathrm{SF}}} \left\{ \text{Event ID}_i^{\mathrm{SA}} \right\}. \end{aligned} \tag{22}$$

Both the HA and the FA check the data fields with the corresponding information that they have and return Message (23) to the SA

$$\begin{aligned} H \rightarrow S &: \mathrm{ID}_H \| E_{k_{\mathrm{SH}}} \left\{ \text{Event ID}_i^{\mathrm{SA}} \| \mathrm{apv} : \mathrm{sig}_H \right\} \\ F \rightarrow S &: \mathrm{ID}_F \| E_{k_{\mathrm{SF}}} \left\{ \text{Event ID}_i^{\mathrm{SA}} \| \mathrm{apv} : \mathrm{sig}_F \right\} \end{aligned} \tag{23}$$

where apv denotes the approval code.

For all cases, after the FA completes Step 3) in the authentication process, it constructs the Event $\mathrm{ID}_i^{\mathrm{SA}}$ as

$$\begin{cases} \mathrm{ID}_F \| \mathrm{sn}_i \| \mathrm{PID}_j \| \mathrm{ID}_H \| \mathrm{ts}_5 \| \mathrm{stat} \| 1 : \mathrm{sig}_F & \text{Auth:1} \\ \mathrm{ID}_F \| Tk\mathrm{ID}_i \| \mathrm{PID}_{j+c} \| \mathrm{ID}_H \| \mathrm{ts}_5 \| \mathrm{stat} \| 2 : \mathrm{sig}_F & \text{Auth:2} \\ \mathrm{ID}_F \| \mathrm{sn}_i \| \mathrm{PID}_j \| \mathrm{ID}_S \| \mathrm{ts}_5 \| \mathrm{stat} \| 3 : \mathrm{sig}_F & \text{Auth:3} \\ \mathrm{ID}_F \| \mathrm{sn}_i \| \mathrm{PID}_j \| \mathrm{ID}_H \| \mathrm{ts}_5 \| \mathrm{stat} \| 4 : \mathrm{sig}_F & \text{Auth:4} \end{cases} \tag{24}$$

where stat denotes the MT's network usage statistics during the session. The FA sends Message (25) to the MT's home network

$$S \rightarrow H : \mathrm{ID}_F \| E_{k'_{\mathrm{FH}}} \left\{ \text{Event ID}_i^{\mathrm{FA}} \right\} \tag{25}$$

where H could be the HA or the SA, and k'_{FH} could be k_{HF} or k_{SF}, depending on the authentication type. After it receives Message (25), the HA checks the data fields with the local corresponding information and confirms if the following conditions are satisfied during the claimed service time.

1) The MT is not being serviced by other networks.
2) The MT is not in power-OFF state.
3) Optionally, the MT is located within the network service provider's physical location by checking the MT's care-of-address or by location-aware application [14].

If all tests are passed, the HA attaches the approval code and its digital signature and sends Message (26) to the FA

$$H \rightarrow F : \mathrm{ID}'_H \| E_{k'_{\mathrm{FH}}} \left\{ \text{Event ID}_i^{\mathrm{FA}} \| \mathrm{apv} : \mathrm{sig}'_H \right\}. \tag{26}$$

Note that authentication type 2 differs from other types. The HA is not notified when the FA offers service to the MT. The key verification point is to validate PID_{j+c} in conjunction with the three test conditions.

IV. PERFORMANCE EVALUATION

In this section, we analyze the security correctness of the proposed protocol and overhead evaluation of the proposed messaging scheme.

A. Security Analysis

A secure protocol designed for roaming services requires the following: 1) prevention of fraud by ensuring that the mobile user and network entity are authentic, i.e., there is a mutual authentication mechanism between a network entity and a mobile user; 2) assuring mutual agreement and freshness of the session key; 3) prevention of replaying attack so that intruders are not able to obtain sensitive data by relaying a previously

intercepted message; and 4) privacy of mobile user's location information during the communication so that it is requisite to provide the mechanism for user anonymity [15].

To prove the correctness of the authentication provided by the proposed security protocol, we use the logic of authentication developed by Burrows, Abadi, and Needham (BAN logic) [16]. Since the conventional notation for security protocols is not convenient for manipulation in the logic of authentication, they introduce the rules in annotating protocols, transforming each message into a logic formula. The BAN logic is the most widely used logic in analyzing authentication protocols [17]. We propose a different notation for the specific session key. We assume that, from the security viewpoint, the function of the unique session key is the same as for the public key. The notation of BAN Logic is shown in Table III.[1]

1) Proof of the Proposed Authentication Scheme[2]*:* Additional notations are listed as follows.

A: MT (A': MT's PID), B: FA, C: SA, D: HA. The security scheme can be described as follows:

1) $A \rightarrow B : A', H, \{A, N_A, k_i, t_1\}_{k_{\mathrm{AD}}}$ from (2);
2) $B \rightarrow C : \{\{A', H, \{A, N_A, k_i, \mathrm{ts}_1\}_{k_{\mathrm{AD}}}, \mathrm{sn}_i, t_2\}_{k_{\mathrm{BC}}}\}_{k_B^{-1}}$ from (3);
3) $C \rightarrow D : \{B, k_B, \{A', H, \{A, N_A, k_i, \mathrm{ts}_1\}_{k_{\mathrm{AD}}}, \mathrm{sn}_i, t_2\}_{k_{\mathrm{CD}}}, t_3\}_{k_C^{-1}}$ from (4);
4) $C \rightarrow B : \{D, k_D, t_3\}_{k_C^{-1}}$ from (5);
5) $D \rightarrow B : \{\{N_A, k_i, A', t_4\}_{k_B}\}_{k_D^{-1}}$ from (6);
6) $B \rightarrow A : \{H(N_A), N_B, t_5\}_{k_i}$ from (7);
7) $A \rightarrow B : \{H(N_B), t_6, \mathrm{comm}\}_{k_i}$.

The idealized protocol can be described as follows:

i3) $C \rightarrow D : \left\{ \overset{k_B}{\mapsto} B, \{\{A, k, N_A\}_{k_{\mathrm{AD}}}\}_{k_{\mathrm{CD}}} \right\}_{k_C^{-1}}$;

i4) $C \rightarrow B : \left\{ \overset{k_D}{\mapsto} D \right\}_{k_C^{-1}}$;

i5) $D \rightarrow B : \{\{k, N_A\}_{k_B}\}_{k_D^{-1}}$;

i6) $B \rightarrow A : \left\{ \left\langle A \overset{N_B}{\rightleftharpoons} B \right\rangle_{N_A} \right\}_k$;

i7) $A \rightarrow B : \left\{ \left\langle A \overset{N_A}{\rightleftharpoons} B \right\rangle_{N_B} \right\}_k$.

Items 1) and 2) are omitted, since Item 2) simply passes Item 1) to C, and the content of Item 1) is included in Item 3).

It is assumed that each principal knows its own secret key and believes that its own nonce and other's timestamp are fresh. The SA also knows other's public key. The above assumptions are summarized as follows:

A1) $A| \equiv A \overset{k_{AD}}{\leftrightarrow} D$;
A2) $A| \equiv \#(N_A)$;
A3) $A| \equiv \#(k)$;

[1]The table is organized from [16].

[2]Because of space limitations, only the full version of the authentication scheme is proved. The subset of authentication schemes for the special cases should comply with this proof.

TABLE III
NOTATIONS OF BAN LOGIC

Symbol	Explanation
$P\| \equiv X$	P believes X, or P would be entitled to believe X. In particular, the principal P may act as through X is true. This construct is central to the logic.
$P \triangleleft X$	P sees X. Someone has sent a message containing X to P, who can read and repeat X (possibly after doing some decryption)
$P\| \sim X$	P once said X. The principal P at some time sent a message including the statement X. It is not Because of space limitation, only the full version of the authentication scheme is proved. The subset of authentication schemes for the special cases should comply with this proof.known whether the message was sent long age or during the current run of the protocol, but it is known that P believed X when he sent the message.
$P\| \Rightarrow X$	P has jurisdiction over X. The principal P is an authority on X and should be trusted on this matter. This construct is used when a principal has delegated authority over some statement. For example, encryption keys need to be generated with some care, and in some protocols certain servers are trusted to do this properly. This may be expressed by the assumption that the principals believe that the server has jurisdiction over statements about the quality of keys.
$\#(X)$	The formula X is fresh, that is, X has not been sent in a message at any time before the current run of the protocol. This is usually true for nonce, that is, expressions generated for the purpose of being fresh. A nonce commonly includes a timestamp or a number that is used only once such as a sequence number.
$P \overset{K}{\leftrightarrow} Q$	P and Q may use the shared key K to communicate. The key K is good, in that it will never be discovered by any principal except P or Q, or a principal trusted by either P or Q.
$\overset{K}{\mapsto} P$	P has K as a public key. The matching secret key (the inverse of K denoted K-1) will never be discovered by any principal except P or a principal trusted by P.
$P \overset{X}{\rightleftharpoons} Q$	The formula X is secret known only to P and Q. And possibly to principals trusted by them. Only P and Q may use X to prove their identities to one another. Often, X is fresh as well as secret. An example of a shared secret is a password.
$\{X\}_k$	This represents the formula X encrypted under the key K. Formally, $\{X\}_k$ is an abbreviation for an expression of the form $\{X\}_k \ from \ P$. We make the realistic assumption that each message is mentioned for this purpose. In the interests of brevity, we typically omit this in our examples.
$< X >_Y$	This represents X combined with the formula Y; it is intended that Y be a secret, and that its presence prove the identity of whoever utters $< X >_Y$. In implementations, X is simply concatenated with the password Y; our notation highlights that Y plays a special role, as proof or origin for X. The notation is intentionally reminiscent of that for encryption, which also guarantees the identity of the source of a message through knowledge of a certain kind of secret.

B1) $B| \equiv \overset{k_B}{\mapsto} B$;
B2) $B| \equiv \overset{k_C}{\mapsto} C$;
B3) $B| \equiv B \overset{k_{BC}}{\leftrightarrow} C$;
B4) $B| \equiv C| \Rightarrow \overset{k_D}{\mapsto} D$;
B5) $B| \equiv \#(N_B)$;
C1) $C| \equiv C \overset{k_{CD}}{\leftrightarrow} D$;
C2) $C| \equiv \overset{k_C}{\mapsto} C$;
C3) $C| \equiv \overset{k_B}{\mapsto} B$;
C4) $C| \equiv \overset{k_D}{\mapsto} D$;
C5) $C| \equiv B \overset{k_{BC}}{\leftrightarrow} C$;

D1) $D| \equiv \overset{k_D}{\mapsto} D$;
D2) $D| \equiv \overset{k_C}{\mapsto} C$;
D3) $D| \equiv C \overset{k_{CD}}{\leftrightarrow} D$;
D4) $D| \equiv C| \Rightarrow \overset{k_B}{\mapsto} B$;
D5) $D| \equiv A \overset{k_{AD}}{\leftrightarrow} D$.

Proof: From i3)

$$D \triangleleft \{\{A, k, N_A\}_{k_{\mathrm{AD}}}\}_{k_{\mathrm{CD}}} \tag{27}$$

$$D \triangleleft \left\{\overset{k_B}{\mapsto} B\right\}_{k_C^{-1}}. \tag{28}$$

By applying D3) and D5) to (27), we have

$$D| \equiv A| \sim k, N_A. \tag{29}$$

By checking the timestamp, D believes that N_A and k are fresh, and we further have

$$D| \equiv A| \equiv k \tag{30}$$

$$D| \equiv A| \equiv N_A. \tag{31}$$

By applying C2), C3), D2), and D4) to (28), we have

$$D| \equiv C| \equiv \overset{k_B}{\mapsto} B \tag{32}$$

$$D| \equiv \overset{k_B}{\mapsto} B. \tag{33}$$

Similarly, from i4), by applying C2), C4), B2), and B4), we have

$$B| \equiv C| \equiv \overset{k_D}{\mapsto} D \tag{34}$$

$$B| \equiv \overset{k_D}{\mapsto} D. \tag{35}$$

From i5)

$$B \triangleleft \{\{N_A, k\}_{k_B}\}_{k_D^{-1}}. \tag{36}$$

By applying (33), (35), and B1), we have

$$B| \equiv D| \sim k, N_A. \tag{37}$$

By applying (30) and (31), we have

$$B| \equiv D| \equiv A| \equiv k \tag{38}$$

$$B| \equiv D| \equiv A| \equiv N_A. \tag{39}$$

From i6)

$$A \triangleleft \{N_A, N_B\}_k. \tag{40}$$

By applying A2) and A3), we have

$$A| \equiv B| \sim N_B, N_A. \tag{41}$$

By checking the timestamp, A believes N_B is fresh. We have

$$A| \equiv B| \equiv A \overset{N_B}{\rightleftharpoons} B. \tag{42}$$

By applying A1)

$$A| \equiv B| \equiv N_A \tag{43}$$

$$A| \equiv B| \equiv k \tag{44}$$

$$A| \equiv B| \equiv D| \equiv N_A \tag{45}$$

$$A| \equiv B| \equiv D| \equiv k \tag{46}$$

$$A| \equiv D| \equiv N_A \tag{47}$$

$$A| \equiv D| \equiv k. \tag{48}$$

Since B does not know N_A and k, unless they are told by D, and at the same time, A trusts its home network D

$$A| \equiv B| \equiv A \overset{N_A}{\rightleftharpoons} B \tag{49}$$

$$A| \equiv B| \equiv A \overset{k}{\leftrightarrow} B. \tag{50}$$

From i7)

$$B \triangleleft \{N_B\}_k. \tag{51}$$

By applying B5) and (49), we have

$$B| \equiv A| \sim N_b \tag{52}$$

$$B| \equiv A| \equiv A \overset{k}{\leftrightarrow} B \tag{53}$$

$$B| \equiv A| \equiv A \overset{N_A}{\rightleftharpoons} B. \tag{54}$$

■

Therefore, the main conclusions are (49), (50), (53), and (54), which show that A and B are mutually authenticated. At the same time, k and N_A are shared with A's trusted party D as well, which allows the service provider to decrypt the communication when they are requested by law enforcement units.

2) Robustness of Proposed Security Schemes: The proposed security schemes are robust to resist certain attacks and sniffing. The intruder cannot impersonate all the parties. In both authentication and event-tracking schemes, the messages between wired parties, such as service providers and SAs, are identified by their digital signatures. The proposed authentication scheme can resist replay and middleman attacks. The intruder cannot act as the MT since he cannot generate a meaningful Message (2), which can be detected by the HA.

The user anonymity and intractability are implemented by dynamic PID. The dynamics are achieved by recursive hash operations of the session key, and the confidentiality is achieved by the involvement of secret shared key and one-way function so that the FA cannot generate a valid PID series while knowing the session key. Depending on the user's preference and the unit's power condition, PID can be regularly refreshed. Since the PID refresh is an independent process, the PID can be used in cellular networks to replace the International Mobile Equipment Identity (IMEI) when the traditional authentication is used so that user anonymity and intractability are supported in the entire integrated network. The PID recovery processes have

TABLE IV
DATA LENGTH OF PARAMETERS

Parameter	Data Length (HEX digits)
ID_M/PID	15
ID network (IP address)	8
Nonce	16
Session key	32
Timestamp	8
Hash function output	16
Ticket ID (ticket)	8
Ref (ticket)	4
Exp (ticket)	8

the capability of detecting and restoring unsynchronized PID in the HA due a poor wireless channel or a software fault.

The replay attack is prevented by implementing an encrypted timestamp mechanism. The intruder cannot update the encrypted timestamp in the replayed message, which can be identified by the legitimate users if its timestamp is out of the predefined range.

B. Overhead Analysis

The overhead is critical since the security protocols are implemented in the mobile devices in the wireless environment. Therefore, heavy computation by the mobile is not feasible [18]–[20]. Since the bandwidth is lower and the channel error is higher in the wireless networks than those in the wired networks, it is important for the security protocols to minimize the message size and the number of message exchanges.

The data length of parameters in authentication messages related to the MT is shown in Table IV. The total authentication message exchange via the wireless link is less than 0.11 kB. The MT does not deal with asymmetric cryptography computation, except for verifying the HA's digital signature for once during the PID synchronization recovery. The selection of cryptographic algorithm and the data length for the parameters can be adjusted in balancing the communication/computation overhead and security strength.

In the following section, dynamic overhead characteristics of the proposed service model will be discussed. We first describe our analytical model. Then, we develop general expressions that describe the performance metrics of our interests. Finally, we present and discuss numerical results from various network scenarios and settings.

1) Analytical Model: The integrated WLANs and the cellular networks under consideration are shown in Fig. 2. The cellular networks usually have wide coverage area, whereas the WLAN hotspots have small coverage area and are only available within distinct hotspots. When an MT roams to a new area, it usually registers in one of the cellular networks and does not switch to other cellular networks. Therefore, it is reasonable to assume that a single large cellular network is overlaid on disjoint WLAN hotspots. Without loss of generality, we consider one cellular network covering n WLAN hotspots.

In order to analyze the overhead performance, we consider a simplified but reasonable model where a mobile user roams, as shown in Fig. 6. We assume that the residence times in the wireless networks are exponentially distributed random variables

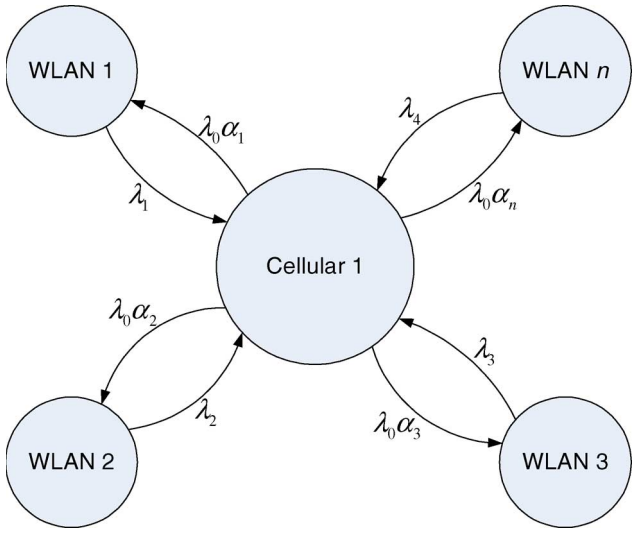

Fig. 6. Mobile user roaming/mobility.

with parameters λ_i (for $i = 0, 1, \ldots, n$). To differentiate the accessing rate from the WLAN hotspot i, we further consider the parameter α_i (for $i = 1, 2, \ldots, n$), which is defined as the probability of a mobile user roaming from the cellular network to the WLAN hotspot i. Therefore

$$\sum_{i=1}^{n} \alpha_i = 1. \tag{55}$$

Let $X(t) \in \{\text{Cellular1}, \text{WLAN1}, \ldots, \text{WLAN}n\}$ denote a process that tracks the network, to which the MT is connected at time t. Therefore, the process $X(t)$ can be modeled as a continuous time Markov process. By sampling the random process $X(t)$ after time instances $t_k = k\tau$, where $\tau \ll (1/\lambda_i)$, for $i = 0, 1, \ldots, n$, the new sampled process $X(t_k)$ is a Markov chain [21] in the state space $\{0, 1, \ldots, n\}$, which is defined by the transition probability matrix

$$\mathbf{P} = \begin{bmatrix} 1-p_0 & \alpha_1 p_0 & \alpha_2 p_0 & \cdot & \cdot & \cdot & \alpha_n p_0 \\ p_1 & 1-p_1 & 0 & & & & \\ p_2 & 0 & 1-p_2 & & & & \\ \cdot & & & \cdot & & & \\ \cdot & & & & \cdot & & \\ \cdot & & & & & \cdot & \\ p_n & 0 & 0 & 0 & 0 & 0 & 1-p_n \end{bmatrix} \tag{56}$$

where $p_i = 1 - e^{-\lambda_i \tau}$, for $i = 0, 1, \ldots, n$. For a given network setup, the mean residence time(s) $1/\lambda_i$, $\forall i$, the roaming probability α_i in WLAN hotspots, and the sampling time interval τ can be empirically obtained.

2) Performance Analysis: The metric of our interest in studying the performance is the overhead cost, which is defined as the average cost per network roaming. The overhead cost can be computed as

$$C = \frac{1}{2} C_0 + \frac{1}{2} \sum_{i=1}^{n} \alpha_i C_i \tag{57}$$

TABLE V
PARAMETERS FOR NUMERICAL RESULTS

Parameter	Values
Mean residence times $(1/\lambda_0, 1/\lambda_1, 1/\lambda_2, 1/\lambda_3)$	(100,100,100,100) slots
Symmetrical WLAN hotspots roaming probability $(\alpha_1, \alpha_2, \alpha_3)$	(1/3,1/3,1/3)
Asymmetrical WLAN hotspots roaming probability $(\alpha_1, \alpha_2, \alpha_3)$	(1/4,1/2,1/4)
Number of service providers in WLAN1 and WLAN3	1
Number of service providers in WLAN2	10
Sampling time interval τ	1 slot
Cellular network average roaming overhead cost C_0 in traditional scheme	2 hops
Cellular network average roaming overhead cost C_0 in proposed scheme	2 hops
WLAN 1, 2, and 3 hotspots average roaming overhead cost with/without cached credentials $\{C_i^{ticket}/C_i^{no\_ticket}\}$ in traditional scheme	(2/2,2/4,2/4) hops
WLAN 1, 2, and 3 hotspot average roaming overhead cost with/without cached credentials $\{C_i^{ticket}/C_i^{no\_ticket}\}$ in proposed scheme	(2/2,2/5,2/4) hops

where $C_i = \Pr_i\{\text{ticket hit}\}\, C_i^{\text{ticket}} + (1 - \Pr_i\{\text{ticket hit}\}\, C_i^{\text{no\_ticket}}$. C_0, C_i^{ticket}, and $C_i^{\text{no\_ticket}}$ are the average overhead costs in the roaming cellular network, the WLAN hotspot i with cached credentials, i.e., ticket TK_i, and the WLAN hotspot i without cached credentials, respectively. $\Pr_i\{\text{ticket hit}\}$ is the ticket hit probability in the WLAN hotspot i, which is defined as the probability of revisiting the WLAN hotspot i before the timeout t_i^{ticket} set for the cached credential expires. $\Pr_i\{\text{ticket hit}\}$ is equal to the probability of revisiting the WLAN hotspot i within the time interval t_i^{ticket}. Therefore, $\Pr_i\{\text{ticket hit}\}$ can be found as the ith element in the first row of the k-step transition matrix computed as $\mathbf{Z}^k$, where $k = \lfloor t_i^{\text{ticket}}/\tau \rfloor$. The matrix $\mathbf{Z}$ is obtained by replacing the ith row of the matrix $\mathbf{P}$, which is defined in (56), by the ith row of its identity matrix.

3) Numerical Results: Numerical results are obtained by considering the scenario where a mobile user roams to a region with one cellular network and three types of WLAN hotspots. The cellular network is a foreign network and has a roaming agreement with the mobile user's home network. The first type of WLAN hotspot (WLAN1) is affiliated with the cellular network, and both networks are tightly coupled. The second type of WLAN hotspot (WLAN2) is owned by the third-party service provider. The third type of WLAN hotspot (WLAN3) is owned by the mobile user's home cellular networks. The second and the third WLAN hotspots have the capability to cache the MT credentials, and they are loosely coupled to the cellular network. The overhead cost is computed by using (57), and the parameter values are given in Table V. The WLAN1 ticket timeout (t_1^{ticket}) is set to zero. The WLAN2 ticket timeout (t_2^{ticket}) and WLAN3 ticket timeout (t_3^{ticket}) take on different values.

In Fig. 7, the overhead costs for the traditional and proposed schemes are compared at various values of WLAN2 ticket timeouts (t_2^{ticket}) for asymmetrical and symmetrical WLAN hotspot roaming probabilities. In each case, the WLAN3 ticket timeout (t_3^{ticket}) is set to zero (i.e., no user's credentials are cached in WLAN3). From Fig. 7, it can be seen

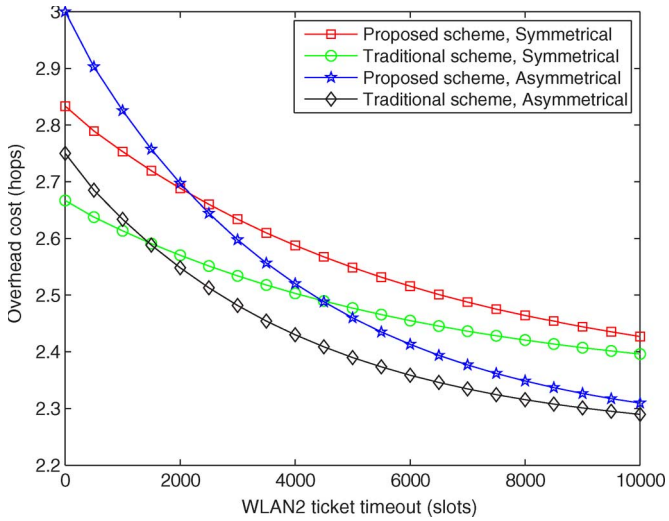

Fig. 7. Overhead cost versus WLAN2 ticket timeout for the proposed and traditional schemes.

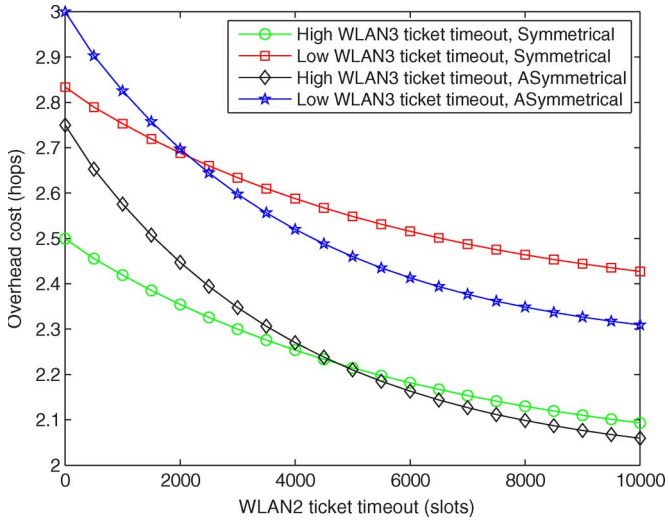

Fig. 8. Overhead cost versus WLAN2 ticket timeout for the high and low WLAN3 ticket timeouts.

that, as the WLAN2 ticket timeout increases, the overhead costs of both the proposed and traditional schemes decrease and converge to the limiting point. Furthermore, at the low values of the WLAN2 ticket timeout, the traditional scheme has lower overhead cost than the proposed scheme; however, as the WLAN2 ticket timeout increases, the difference in overhead cost decreases to zero. The observations suggest that, by choosing an appropriate value of the WLAN2 ticket timeout, the proposed scheme can provide a better and more convenient roaming than the traditional scheme with the same overhead cost.

The overhead cost of the proposed scheme is further studied at various values of WLAN2 ticket timeouts for low WLAN3 ticket timeout $(t_3^{\text{ticket}} = 0)$ and high WLAN3 ticket timeout $(t_3^{\text{ticket}} = 2000$ slots), respectively, for the asymmetrical and symmetrical WLAN hotspot roaming probabilities. Fig. 8 shows that the overhead-cost performance of the proposed

scheme can be further improved by increasing the WLAN3 ticket timeout. However, the amount of improvement depends on the symmetry of the roaming probabilities $(\alpha_1, \alpha_2, \alpha_3)$.

From the performance analysis, it is observed that the ticket greatly reduces the authentication latency. However, the ticket also allows the FA to link the MT's current PID with the previous PID stored in its database. On the other hand, the latency improvement becomes smaller and smaller when the ticket timeout increases. Therefore, it is neither suitable nor necessary to choose very large ticket expiration time so that the MT will do normal authentication to break the linage from time to time. The setting of the value reflects the balance between the latency performance and the identity intractability strength.

V. Conclusion

In this paper, a SA-based WLAN/cellular network integrated service model and the relevant authentication and event tracking for billing support schemes have been proposed. The service model does not require the inefficient peer-to-peer roaming agreements to support seamless user roaming between the WLAN hotspots and the cellular networks operated by the independent wireless network service providers. The proposed authentication and event-tracking schemes take both anonymity and intractability of the mobile users into consideration and operate independently so that the integrated billing service can be applied to the cellular networks even if they still use the traditional authentication scheme. Both security analysis and overhead evaluation have demonstrated that the proposed WLAN/cellular network integrated service model exhibits good security and efficiency.

Acknowledgment

This work was supported by the Natural Science and Engineering Research Council of Canada (NSERC) under a Postdoctoral Fellowship and a Strategic Project Grant.

References

[1] Cisco Systems Inc., *Wireless LAN Technologies, Products, & Trends*, Apr. 2004. Tech. Rep.
[2] M. Slocombe, "WiFi kit revenues hit record levels: Infonetics research," *Market Survey*, May 2005.
[3] M. Buddhikot, G. Chandranmenon, S. Han, Y. W. Lee, S. Miller, and L. Salgarelli, "Integration of 802.11 and third-generation wireless data networks," in *Proc. IEEE Infocom*, Apr. 2003, vol. 1, pp. 503–512.
[4] C. Perkins, *IP Mobility Support For IPv4*, 2002. IETF RFC 3344.
[5] D. Johnson, C. Perkins, and J. Arkko, *Mobility Support in IPv6*, 2004. IETF RFC 3775.
[6] G. Rose and G. M. Koien, "Access security in CDMA2000, including a comparison with UMTS access security," *IEEE Wireless Commun.—Special Issue on Mobility and Resource Management*, vol. 11, no. 1, pp. 19–25, Feb. 2004.
[7] G. M. Koien, "An introduction to access security in UMTS," *IEEE Wireless Commun.*, vol. 11, no. 1, pp. 8–18, Feb. 2004.
[8] W. Song, W. Zhuang, and A. Saleh, "Architectures for integrating wireless LAN and cellular networks," *Int. J. Wireless Mobile Comput.*, submitted for publication.
[9] V. W.-S. Feng, L.-Y. Wu, Y.-B. Lin, and W. E. Chen, "WGSN: WLAN-based GPRS environment support node with push mechanism," *Comput. J.*, vol. 47, no. 4, pp. 405–417, 2004.
[10] M. Shi, L. Xu, X. Shen, and J. W. Mark, "Fast vertical handoff for cellular and WLAN interworking," *Wireless Commun. Mobile Comput.*, submitted for publication.
[11] M. M. Buddhikot, G. Chandranmenon, S. Han, Y.-W. Lee, S. Miller, and L. Salgarelli, "Design and implementation of a WLAN/cdma2000 interworking architecture," *IEEE Commun. Mag.*, vol. 41, no. 11, pp. 90–100, Nov. 2003.
[12] M. Shi, X. Shen, and J. W. Mark, "IEEE802.11 roaming and authentication in wireless LAN/cellular mobile networks," *IEEE Wireless Commun.*, vol. 11, no. 4, pp. 66–75, Aug. 2004.
[13] 3GPP, *Universal Mobile Telecommunications System (UMTS); Characteristics of the Universal Subscriber Identity Module (USIM) Application*, 2006. 3GPP TS 31.102 Ver. 7.4.1 Rel. 7.
[14] M. Hazas, J. Scott, and J. Krumm, "Location-aware computing comes of age," *Computer*, vol. 37, no. 2, pp. 95–97, Feb. 2004.
[15] S. Patel, "Weakness of Northern American wireless authentication protocol," *IEEE Pers. Commun.*, vol. 4, no. 3, pp. 40–44, Jun. 1997.
[16] M. Burrows, M. Abadi, and R. Needham, "A logic of authentication," *ACM Trans. Comput. Syst.*, vol. 8, no. 1, pp. 18–36, 1990.
[17] B. Schneier, *Applied Cryptography: Protocols, Algorithms, and Source Code in C*, 2nd ed. Hoboken, NJ: Wiley, 1996.
[18] D. S. Wong and A. H. Chan, "Mutual authentication and key exchange for low power wireless communications," in *Proc. IEEE MILCOM*, 2001, vol. 1, pp. 39–43.
[19] K. Shim, "Cryptanalysis of mutual authentication and key exchange for low power wireless communications," *IEEE Commun. Lett.*, vol. 7, no. 5, pp. 248–250, May 2003.
[20] S. L. Ng and C. Mitchell, "Comments on mutual authentication and key exchange protocols for low power wireless communications," *IEEE Commun. Lett.*, vol. 8, no. 4, pp. 262–263, Apr. 2004.
[21] R. A. Howard, *Dynamic Probabilistic Systems—Volume I: Markov Models*. New York: Dover, 1971.

Minghui Shi received the B.S. degree in electrical and computer engineering from Shanghai Jiao Tong University, Shanghai, China, in 1996 and the M.A.Sc. and Ph.D. degrees in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2002 and 2006, respectively.

He is currently with McMaster University, Hamilton, ON, as a Natural Sciences and Engineering Research Council of Canada Postdoctoral Fellow. He is also with the Centre for Wireless Communications, Department of Electrical and Computer Engineering, University of Waterloo, as a Research Associate. His current research interests include network security and mobility management in wireless LAN/cellular network integration, vehicular communications networks and delay tolerant networks.

Humphrey Rutagemwa (S'03) received the B.Sc. degree in electronics and communications (with first class honors) from the University of Dar es Salaam, Dar es Salaam, Tanzania, in 1998 and the M.Sc. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2002. He is currently working toward the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Waterloo.

From 1999 to 2000, he was a System Engineer with CRDB Bank, Ltd. Since 2003, he has been working as a Research Assistant with the Centre for Wireless Communications, Department of Electrical and Computer Engineering, University of Waterloo. His current research interests include modeling and performance evaluation, cross-layer design and optimization, and vehicular communication networks.

Xuemin (Sherman) Shen (M'97–SM'02) received the B.Sc. degree in electrical engineering from Dalian Maritime University, Dalian, China, in 1982 and the M.Sc. and Ph.D. degrees in electrical engineering from Rutgers University, Camden, NJ, in 1987 and 1990, respectively.

He is with the Centre for Wireless Communications, Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, where he is a Professor and the Associate Chair for Graduate Studies. His research focuses on mobility and resource management in interconnected wireless/wireline networks, ultrawideband (UWB) wireless communication systems, wireless security, and *ad hoc* and sensor networks. He is a coauthor of two books, and he has published more than 200 papers and book chapters in wireless communications and networks, control, and filtering.

Dr. Shen serves as the Technical Program Committee Chair for IEEE Globecom'07, General Cochair for Chinacom'07 and QShine'06, and the Founding Chair for IEEE Communications Society Technical Committee on P2P Communications and Networking. He also serves as the Founding Area Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the Editor-in-Chief for *Peer-to-Peer Networking and Application*, and as an Associate Editor for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, *KICS/IEEE Journal of Communications and Networks*, *Computer Networks*, *ACM/Wireless Networks*, and *Wireless Communications and Mobile Computing* (Wiley), etc. He has also served as a Guest Editor for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, *IEEE Wireless Communications*, and *IEEE Communications Magazine*. He received the Excellent Graduate Supervision Award in 2006 and the Outstanding Performance Award in 2004 from the University of Waterloo, the Premier's Research Excellence Award in 2003 from the Province of Ontario, and the Distinguished Performance Award in 2002 from the Faculty of Engineering, University of Waterloo. He is a Registered Professional Engineer in Ontario.

Jon W. Mark (M'60–SM'80–F'88–LF'03) received the Ph.D. degree in electrical engineering from the McMaster University, Hamilton, ON, Canada, in 1970.

Since then, he has been with the Department of Electrical Engineering (currently Electrical and Computer Engineering), University of Waterloo, Waterloo, ON, and became a Full Professor in 1978. He served as the Department Chairman from July 1984 to June 1990. In 1996, he established the Centre for Wireless Communications, University of Waterloo, and has since been serving as the Founding Director. He was on a sabbatical leave at the IBM Thomas Watson Research Center, Yorktown Heights, NY, as a Visiting Research Scientist from 1976 to 1977, at AT&T Bell Laboratories, Murray Hill, NJ, as a Resident Consultant from 1982 to 1983, at the Laboratoire MASI, Université Pierre et Marie Curie, Paris, France, as an Invited Professor from 1990 to 1991, and at the Department of Electrical Engineering, National University of Singapore, Singapore, as a Visiting Professor from 1994 to 1995. He is a coauthor of the textbook *Wireless Communications and Networking* (Prentice-Hall, 2003). His current research interests are in wireless communications and wireless/wireline interworking, particularly in the areas of resource management, mobility management, and end-to-end information delivery with QoS provisioning.

Dr. Mark has served as a member of a number of editorial boards, including the IEEE TRANSACTIONS ON COMMUNICATIONS, *ACM/Baltzer Wireless Networks, Telecommunication Systems*, etc. He was a member of the Inter-Society Steering Committee of the IEEE/ACM TRANSACTIONS ON NETWORKING from 1992 to 2003 and a member of the IEEE COMSOC Awards Committee from 1995 to 1998.

Aladdin Saleh (M'03–SM'04) received the Ph.D. degree in electrical engineering from London University, London, U.K.

Since March 1998, he has been with the High-Speed Wireless Access, Wireless Technology Department, Bell Canada, Montreal, QC, Canada, which is the largest service provider of wireless, wire line, and Internet in Canada. He primed several projects in the data group; among them are the wireless application protocol and location-based services. Later, he led work on several key projects in the broadband wireless access strategy group, including planning of the IEEE 802.16/Wimax, the IEEE 802.11/WiFi, and the integration of these technologies with the 3G cellular network. He also led or actively participated in several key projects related to push e-mail, IP multimedia subsystem (IMS), and multimedia application. He has also been an Adjunct Full Professor with the Department of Electrical and Computer Engineering, Waterloo University, Waterloo, ON, Canada, since January 2004. He is currently conducting several joint research projects with the University of Waterloo and the University of Toronto, Toronto, ON, on IEEE 802.16/Wimax, multiple-input multiple-output technology, interworking of IEEE 802.11 WLAN and 3G cellular networks, and next-generation wireless networks. He has also worked as a faculty member with different universities and was the Dean and the Chairman of the Department for several years.