Available online at www.sciencedirect.com
SCIENCE DIRECT®
Computer Networks 49 (2005) 119–146
Computer Networks
www.elsevier.com/locate/comnet

A survey of cross-layer performance enhancements for Mobile IP networks

Janise McNair [a,*], Tuna Tugcu [b,1], Wenye Wang [c,2], Jiang (Linda) Xie [d,3]

[a] Department of Electrical and Computer Engineering, University of Florida, P.O. Box 116130, Gainesville, FL 32611, United States
[b] Department of Computer Engineering, Bogazici University, TR-34342 Istanbul, Turkey
[c] Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27695, United States
[d] Department of Electrical and Computer Engineering, University of North Carolina at Charlotte, Charlotte, NC 28223, United States

Received 16 May 2005; received in revised form 7 June 2005; accepted 7 June 2005
Available online 27 June 2005

Responsible Editor: Dr. I.F. Akyildiz

Abstract

Among the characteristics of future wireless networks is the desire to support a wide range of wireless users and a diverse set of services from many different types of networks. One of the most often referenced networking protocols for diverse wireless and mobile networking is Mobile IP. Although recent focus has been on developing a micro-mobility architecture for Mobile IP, an emerging need is to enhance the unifying performance of Mobile IP by using a cross-layer, cross-technology, approach to protocol design, in order to serve a wide variety of users, services and networks. This paper provides a survey of recently proposed performance enhancements for Mobile IP and discusses the impact on network-level performance. After a review of the Mobile IP architecture, recent research on reducing handoff latency is discussed, including the use of layer 2 hints. Then, location registration is explored, including new techniques for authentication of mobile users. Finally, an overview of performance analysis models is provided to show the most recent approaches to determining the impact of mobility on a given network.

Keywords: Mobile IP; Mobility; Authentication; Handoff

* Corresponding author. Tel.: +1 352 392 2629; fax: +1 352 392 0044.
E-mail addresses: mcnair@ece.ufl.edu (J. McNair), tugcu@boun.edu.tr (T. Tugcu), wwang@eos.ncsu.edu (W. Wang), jxie1@uncc.edu (Jiang (Linda) Xie).
[1] Tel.: +90 212 359 7611; fax: +90 212 287 2461.
[2] Tel.: +1 919 513 2549; fax: +1 919 515 5523.
[3] Tel.: +1 704 687 4154; fax: +1 704 687 2352.

1389-1286/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.comnet.2005.06.001

1. Introduction

Among the characteristics of future wireless networks is the desire to support a wide range of wireless users and a diverse set of services from many different types of networks. A unifying technology for diverse wireless networking is Mobile IP, a protocol established by the Internet Engineering Task Force (IETF) that allows mobile nodes (MNs) to change their point of attachment to the Internet while still being able to maintain a connection to the network [1]. For example, the third generation partnership projects (3GPP and 3GPP2), which represent a wide range of standard organizations for the global wireless industry, are currently developing standards for integrated 3G and 4G wireless architectures that support seamless mobility between 3G, WLAN, and other networks over Mobile IP.

Research on Mobile IP and on the design of future wireless networks continues to focus on service optimization. Therefore, issues such as *registration*—including secure authentication and location management, and *handoff*—including reducing handoff delay and packet loss, continue to be at the forefront. In recent surveys on IP mobility, the focus has been to develop appropriate architectures for macro- and micro-mobility management. Ref. [2] outlines proposed solutions for future mobility architectures based on different layers of Mobile IP, and a new wireless network architecture is proposed that is based on network inter-working agents for inter-domain roaming among different types of networks. Ref. [3] describes and compares three alternatives for IP mobility management: (1) the traditional Mobile IP, (2) Migrate, which handles mobility on an end-to-end basis by allowing the MN to change its IP address, and (3) the host identity protocol, which handles mobility using a new name space by separating the host or process address from the interface addresses. A qualitative comparison is given, which focuses on the security, scalability, and robustness aspects of each approach. The focus of this survey is to explore the performance enhancements designed to improve service quality in Mobile IP, which includes reducing handoff delay and packet loss through layer 3 movement estimation and increasing security through updated authentication techniques. In Section 2, the Mobile IP architecture is reviewed. In Section 3, recent research on reducing handoff latency using layer 2 hints is discussed. In Section 4, the problem of location registration is explored, including new techniques for authentication of mobile users. Finally, Section 5 provides an overview of performance analysis models used to determine the impact of mobility on a given network. The paper concludes in Section 6.

2. Mobile IP

As mentioned previously, the Mobile Internet protocol (Mobile IP) was standardized by the Internet Engineering Task Force (IETF) to allow MN to change their point of attachment to the Internet while still being able to maintain a connection to the network [1]. The Mobile IP approach is illustrated in Fig. 1. Under the Mobile IP, an MN that is currently residing in its home subnetwork, is served by a *home agent* that forwards all incoming packets to the MN at its home IP address. When the MN moves away from its home subnetwork to a new location, the node must contact a *foreign agent* at the new subnetwork to obtain a new (temporary) IP address, called a *care-of-address*. The care-of-address is then used to identify and communicate with the MN at the local network. To connect the MN to its home network, a binding update must be performed to notify the home agent about the MN's new care-of-address. The home agent then forwards all incoming packets to the MN using a process referred to as *tunneling*, i.e., the home agent encapsulates the incoming packets for the MN and forwards them to the foreign agent, which in turn decapsulates them and delivers them to the MN. Meanwhile, the MN can continue to transmit packets directly to the correspondent node (CN).

The detailed procedures for Mobile IP designed for macro-mobility are as follows:

- *Agent discovery*: An MN is able to detect whether it has moved into a new subnet by periodically receiving unsolicited *agent advertise-*

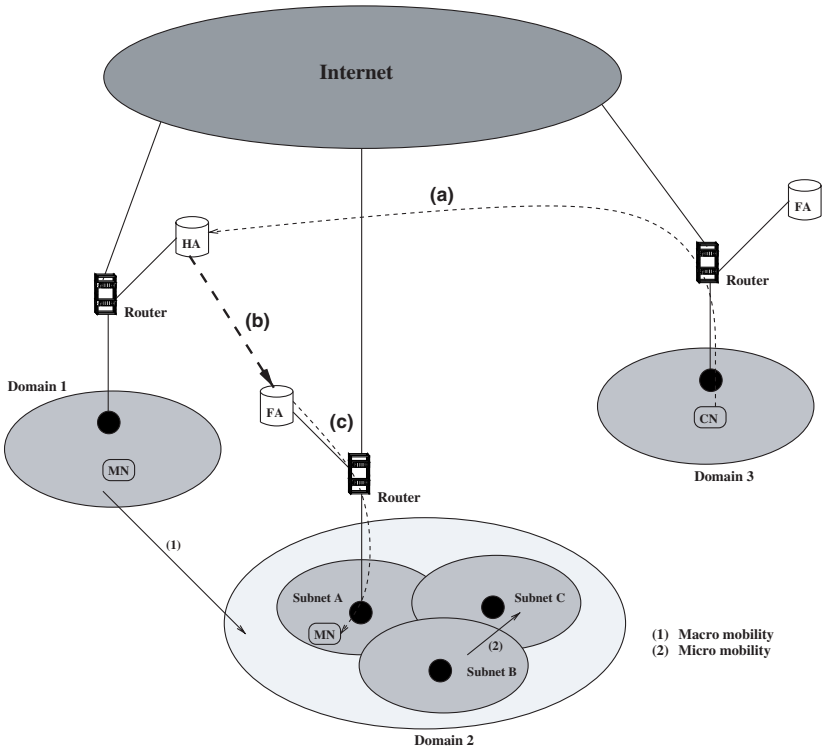


Fig. 1. Mobile IP architecture.

ment messages broadcasted from each foreign agent. An MN can also send *agent solicitation* messages to learn about the presence of any prospective mobility agent.

- *Registration*: When the MN discovers that it is in a foreign network, it obtains a new care-of-address (CoA). This CoA can be obtained by soliciting or listening for foreign agent advertisements (a foreign agent CoA), or contacting dynamic host configuration protocol (DHCP) or point-to-point (PPP) (a co-located CoA) [4]. The MN registers the new CoA with its home agent. Then home agent updates the mobility binding by associating the CoA of the MN with its permanent IP address.
- *Routing and tunneling*: Packets sent by a CN to the MN are intercepted by the home agent. The home agent encapsulates the packets and tunnels those to MN's CoA. In case of foreign agent CoA, the encapsulated packets reach the foreign agent serving the MN, which decapsulates the packets and forwards them to the MN. In case of co-located CoA, the encapsulated packets reach the MN, which then decapsulates them. Steps (a)–(c) of Fig. 1 show the routing and tunneling procedures when foreign agent CoA is used. For co-located CoA the steps are similar except that the tunneling (step (b)) ends at the MN instead of the foreign agent.

2.1. Mobile IP shortcomings and improvements

The Mobile IP approach has several shortcomings. First, the packets sent from a CN to an MN are first intercepted by the home agent, which then tunnels those to the MN. However, packets from the MN are sent directly to the CN. This triangular routing problem results in communication routes significantly longer than the optimal routes and introduces extra delay for packet delivery [5].

Second, when an MN moves from one subnet to another, there is no way that the new foreign agent can inform the old foreign agent about the movement of the MN. Hence, packets already tunneled to the old CoA and in flight cannot be delivered to the MN and are lost. Therefore, a smooth handoff is not possible. Third, Mobile IP is not a satisfactory solution for highly mobile users [6]. When an MN moves among subnets, its location and routes must be updated. Mobile IP requires that an MN sends a location update to its home agent whenever it moves from one subnet to another one. This location registration is required even though the MN does not communicate with others while moving. The signaling cost associated with location updates may become very significant as the number of MNs increases [7]. Moreover, if the distance between the visited network and the home network of the MN is large, the signaling delay for the location registration is long.

The problem of triangular routing can be solved by route optimization [8]. Route optimization in Mobile IP also defines procedure to notify the MN's old foreign agent when the MN changes its foreign agent. The basic idea behind route optimization is to use a direct route between MNs and their CNs to bypass the home agent. CNs maintain a binding cache of the CoAs of MNs. When a CN sends packets to an MN, it first checks if it has a binding cache entry for the MN. If yes, then the CN tunnels the packets directly to the CoA of the MN. If there is no binding cache entry available, then the CN sends the packets using the basic Mobile IP procedure, i.e., via MN's home agent. The CN learns about the most recent CoA of MNs in either of two ways:

- When the home agent intercepts and tunnels packets destined to an MN, it sends a *binding update* message to the source of the packets to inform it about the current CoA of the destination MN.
- When tunneled packets reach a foreign agent which no longer has the destination MN in its visitor list, the foreign agent sends a *binding warning* message to the home agent of the MN to ask the home agent to send a *binding update* message to the source node.

Route optimization also takes care of the packets already tunneled to the old CoA and in flight. When an MN moves and registers with a new foreign agent, it requests its new foreign agent to notify the previous foreign agent about the change in CoA. This ensures that the packets in flight to the old CoA are successfully forwarded. It also ensures that packets from the CN with out-of-date binding cache entries for the MN are successfully delivered to the MN's new CoA. Moreover, route optimization also ensures that any resources consumed by the MN at its old foreign agent are released immediately, rather than waiting for its registration time to expire [8].

Other improvements have been proposed and adopted for Mobile IP under the title of Mobile IP version 6 [1]. For example, Mobile IPv6 supports route optimization, which is not always available in Mobile IPv4. In addition, Mobile IPv4 suffers from a lack of security constructs for authorization, authentication, and accounting, as well as for source routing. Mobile IPv6 includes imbedded binding updates and care-of-address configuration for the execution of location updates and for processing the change in the MN's address. The newer version also includes authentication header processing to provide validation of MNs. Finally, IPv6 has a fourfold increase in the IP address space, which may be useful for developing new MN addressing schemes.

Thus, Mobile IP is able to support mobility across both homogeneous and heterogeneous systems, and is well suited for macro-mobility management, i.e., mobility across different network domains. However, Mobile IP is less suited for micro-mobility management, which has been the subject of research recent literature on Mobile IP improvements.

2.2. Mobile IP and micro-mobility

The binding updates and care-of-address exchanges that establish the MN at each new location cause an significant signaling load, as well as significant delays that may be detrimental to the service being received at the MN. For large-scale mobility using Mobile IP these delays are necessary to track a user over many networks or subnet-

works. However, if a user moves to a foreign domain and then stays within the foreign domain for some time, the long-distance signaling load to the home agent may not be necessary. Thus, the solutions proposed for providing IP mobility can be broadly classified into two categories: *macro-mobility* management solutions, which generally refers to Mobile IP itself, and *micro-mobility* management, which concerns the movement of mobile users between two network domains. (One domain usually refers to networks of the same administrative body.) The remainder of this section describes three protocols that are representative of micro-Mobile IP and the common methods to reduce the Mobile IP signaling load: *Hierarchical Mobile IP*, *Cellular IP*, and *HAWAII*.

2.2.1. Mobile IP regional registration/Hierarchical Mobile IP

Mobile IP regional registration aims to reduce the number of signaling messages to the home network, and also to reduce the signaling delay when an MN moves from one subnet to another by performing registrations locally in a regional network. (The detailed protocol specifications can be found in [9].) When an MN first arrives at a regional network, it performs a home registration with its home agent. During the home registration, the home agent registers the care-of-address of the MN, which is actually a publicly routable address of another mobility agent called gateway foreign agent (GFA). When an MN changes foreign agent within the same regional network, it performs only a regional registration to the GFA to update its CoA. When it moves from one regional network to another one, it performs a home registration with its home agent. The packets for the MN are first intercepted by its home agent, which tunnels those to the registered GFA. The GFA checks its visitor list and forwards the packets to the corresponding foreign agent in the visiting subnet of the MN. The foreign agent further relays the packets to the MN.

GFA introduces a layer of hierarchy between the home agent and foreign agents of the MN. The use of the GFA avoids any signaling traffic to the home agent as long as the MN is within a regional network. The structure can be extended to include multiple hierarchy levels of foreign agents beneath the GFA level. Such multiple hierarchy levels are discussed in Hierarchical Mobile IP [10].

2.2.2. Cellular IP

Cellular IP (CIP) [11,12] supports fast handoff and paging techniques in Cellular IP access networks. The protocol is intended to provide local mobility and handoff support for frequently moving hosts. For mobility between different Cellular IP networks, it can inter-work with Mobile IP to provide wide area mobility support. Cellular IP uses a *distributed paging cache* and a *distributed routing cache* for location management and routing, respectively. The distributed paging cache coarsely maintains the position of the idle MNs for efficient paging, whereas the routing cache maintains the position of an active MN up to the subnet level accuracy. When an MN performs handoff, the routing states in the routing cache are dynamically updated. This ensures that the data packets are routed properly to the MN after its handoff.

The Cellular IP architecture is shown in Fig. 2. It shows different wireless access networks connected to the Internet through a gateway (GW), which handles mobility within one domain. Packets destined to the MN reach the GW first. Then the GW forwards the packet to the MN using the host-specific routing path. During intra-domain mobility, when the strength of the beacon signal from the serving BS is lower than that of a neighboring BS, the MN initiates a handoff. The first packet which travels to the GW through the new BS configures a new path through the new BS. This results in two parallel paths from the GW to the MN: one through the old BS and the other through the new BS. If the MN is capable of listening to both BSs at the same time, the handoff is soft or else the handoff is hard [12]. These two paths might have some nodes in common and divert from each other at a crossover node. The path through the old BS will be active for a duration equal to the timeout of route caches. After timeout of the route cache, the entries corresponding to the MN in the nodes which belong only to the old path are deleted. Thereafter, only the new path exists

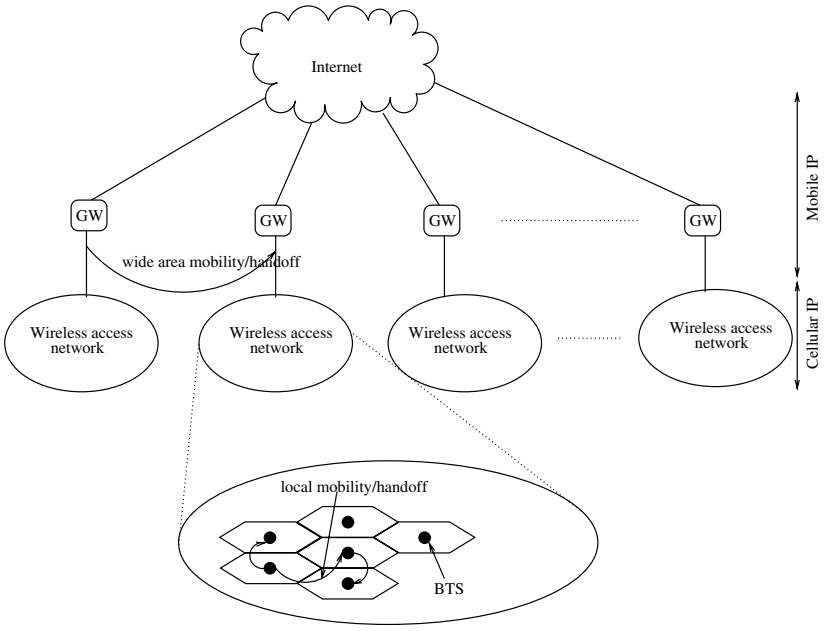


Fig. 2. Cellular IP architecture.

between the GW and the MN. Once the handoff is over, the MN communicates with its corresponding nodes through the new BS. This handoff process of CIP is automatic and transparent to the upper layers.

2.2.3. Handoff-aware wireless access Internet infrastructure (HAWAII)

As illustrated in Fig. 3, HAWAII is another domain-based approach for mobility support. In HAWAII, all of the issues related to mobility management within one domain are also handled by a gateway, in this case called the *domain root router*. When an MN roams within a particular domain, it maintains the connectivity using dynamically established paths set up from the domain root router to the MN. Once established, the IP packets destined to the MN are routed using typical IP routing when the MN is in its home domain. When the MN is in a foreign domain, it registers its CoA with its home agent upon the receipt of an acknowledgment from the domain root router. When the MN is in a foreign domain, all the packets for the MN are intercepted by its home agent first. Then the home agent tunnels the packets to the domain root router serving the MN. Finally, the domain root router routes the packets to the MN using the host-based routing entries. When the MN moves between different subnets of a particular domain, only the route from the domain root router to the BS serving the MN is modified and the remaining path remains the same. Thus, during an intra-domain handoff, the global signaling message load and the handoff latency are reduced significantly.

2.3. Comparison of IP mobility solutions

For mobility between different administrative domains, Mobile IP scheme is widely used. Comparisons of network layer micro-mobility solutions are conducted in [13–16] based on different criteria. In [14], a generic model is established and performance comparison between Cellular IP, HAWAII, and Hierarchical Mobile IP are evaluated based on the same model. A comprehensive comparison of handoff mechanisms of IP micro-mobility protocols is given in [16].

As stated in [14], despite the different design approaches of the proposed micro-mobility proto-

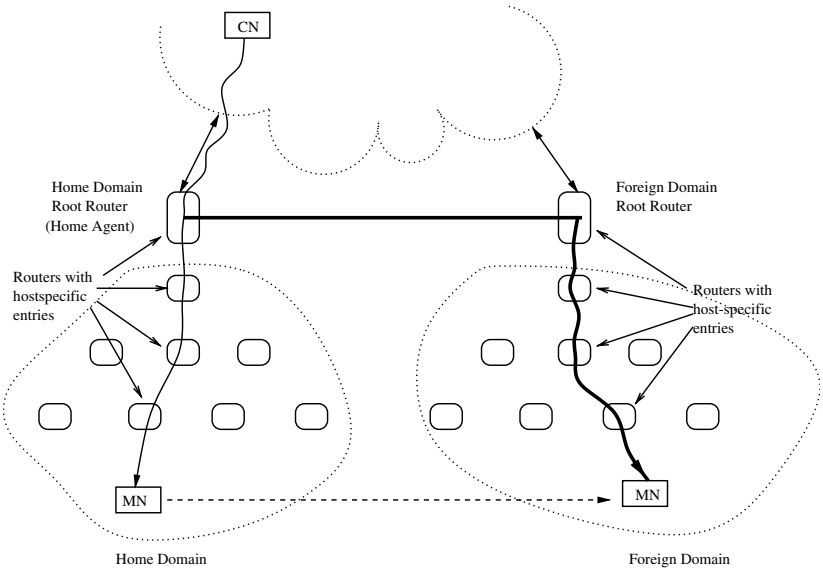


Fig. 3. Network architecture of HAWAII.

cols, the operational principles that govern them are largely similar. Domain root routers are designed in each protocol. All the solutions are trying to localize most of the signaling traffic into one domain to reduce global signaling. Based on different design goals, existing micro-mobility proposals are classified into two types: *routing-based* and *tunnel-based* schemes [17]. Routing-based schemes take advantage of robust IP forwarding. Mobile-specific address lookup tables are maintained by all the mobility agents within the network domain. Cellular IP and HAWAII fall into this category. Under the tunnel-based schemes, registration and encapsulation are performed in a local or hierarchical fashion. Mobile IP regional registration and Hierarchical Mobile IP belong to this category. Generally speaking, routing-based schemes avoid the tunneling overhead, but suffer from the high cost of propagating host-specific routes in all the routers within the domain. The root node of routing-based schemes constitutes a single point of failure [15]. Tunnel-based schemes enhances scalability by introducing hierarchies, but lead to additional costs and delays. Their reliability relies on the mobility agents at each hierarchy [18]. It is demonstrated in [14] that the basic handoff performance of the existing micro-mobility protocols depends only on the position of the crossover mobility agents. The choice of a micro-mobility protocol should be dictated more by deployment considerations.

3. Mobile IP handoff research—cross-layer optimizations

As mentioned previously, Mobile IP allows an MN to change its point of attachment to the Internet while still being able to maintain a connection to the network. Changing its point of attachment requires the MN to engage in handoff. Traditionally, handoff decision is performed based on a perception of channel quality reflected by the received signal strength and other measurements, and the availability of resources at the new cell. The base station usually measures the quality of the radio link channels being used by MNs in its service area. This is done periodically, so that degradations in signal strength below a prescribed threshold can be detected and handoff to another radio channel or cell can be initiated. Under network-controlled handoff (NCHO), or mobile-assisted handoff (MAHO), the network makes the decision for handoff, while under mobile-controlled handoff

(MCHO), the MN must take its own signal strength measurements and make the handoff decision on its own. While performing handoff, the MN's connection may be created at the target base station before the old base station connection is released. This is referred to as a "make before break" handoff. On the other hand, the new connection may be set up after the old connection has been torn down, which is referred to as a "break before make" handoff. In either case, the MN executes a *hard handoff*, which means that the MN can only communicate on a channel with one base station at time. In 3G CDMA networks, an MN is able to communicate on more than one coded channel, which enables it to communicate with more than one base station. Thus, CDMA networks allow a *soft handoff*, where the MN can listen to a set of candidate base stations at the same time before choosing one for its point of attachment [19].

3.1. Mobile IP handoff architecture

Consider the 3G wireless network and wireless local area network (WLAN) overlay structure shown in Fig. 4. The 3G wireless network, such as UMTS or GPRS, is assumed to cover a relatively larger area than the WLAN. Network connectivity is offered to the MN through an access point (AP) that connects to an access router (AR) which belongs to either the 3G wireless network or a WLAN.

When an MN moves out of the coverage area of its current AP, it may attach to a new AP. The roaming of MNs between APs is managed by the link-layer protocol and is known as layer 2, or link-layer handoff. The new AP can be connected to the same access network, or to a different access network. If the new AP is connected to the same subnet as the old AP, the MN can continue its IP communication through the new AP through a layer 2 handoff without any configuration change at layer 3. If the new AP is connected to a different subnet, then the MN needs to configure a new IP address that is valid for the new subnet and use some additional mechanism to maintain its ongoing communication sessions, such as a pre/post-registration protocol [20]. In this case, the layer 2 handoff will result in a layer 3 handoff. Furthermore, if the current AP and the new AP are connected to two different systems, e.g. 3G network and WLAN as in Fig. 1, the layer 2 handoff will result in a vertical handoff between two networks.

Thus, each time the MN has a handoff, it is possible that it will initiate a change in IP-layer configurations, such as its IP address and default gateway information. In order to make these

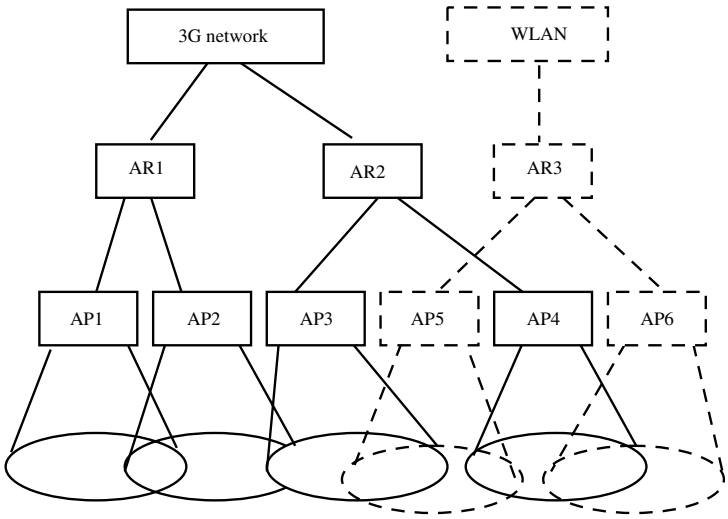


Fig. 4. 3G/WLAN integration.

changes, the IP module has to detect the new network attachment, realize that the old configuration is no longer valid, and obtain the new configuration parameters. In Mobile IPv4 and IPv6, the network detection phase traditionally uses network-layer movement detection, such as a change in the advertised subnet prefixes [4,1]. However, network-layer indications are not readily available upon a link change. Thus, general reliance on the network layer may introduce unnecessary delays due to layer 3 signaling for a simple layer 2 handoff. If information could be gathered at layer 2 to determine the need for layer 3 signaling, then both the delay and signaling load could be greatly improved over current standards. Thus, recent proposals for MIP enhancement include the use of layer 2 information, or *hints* and address three areas: (1) determining a catalogue of available layer 2 hints, (2) describing procedures for evaluating layer 2 hints, and (3) effective use of layer 2 hints in MIP handoffs.

3.2. Catalogue of available layer 2 hints

Two sets of parameters have been proposed to gather layer 2 information: (1) the MN set, which gathers information available at the MN, and (2) the AR set. In each set, parameters are grouped into three distinct categories: static parameters, which are related to the hardware implementation of the interface, configuration parameters, which are managed through interface configuration, and status parameters, which are highly varying in order to provide the current link environment of the interface. In [21], a non-exhaustive catalogue of link-layer hints from well-known link-layer technologies is provided, and possible link-layer hints indicating link status were discussed specifically for GPRS, 3GPP2, and WLAN. A high-level abstract categorizes such hints into *link identifiers*, *link-up hints*, and *link-down hints*. In [22], the following categories of layer 2 hints are provided:

- *Link-type hint*: Characteristics that describes the type of the technology from which the layer 2 trigger was generated. Examples include:
  - *MN measured bandwidth*: current available bandwidth measured by the MN over the link.
  - *MN bit error rate*: current measured bit error rate.
  - *MN packet error rate*: current measured packet error rate.
  - *MN current data rate*: current rate at which the MN link layer is transmitting/receiving packets.
- *Link identifier*: For example, in GPRS networks, the relevant link identifiers are the transaction identifier (TI), which includes the network-layer service access point identifier (NSAPI). The NSAPI can be used as the link identifier since it can uniquely identify the associated policy decision point (PDP) context (the soft state maintained between the MN, the SGSN and the GGSN for guaranteeing a negotiated quality of service in a GPRS network). In the WLAN, the link identifier used by the MN is the basic service set identification (BSSID), where the BSSID is the MAC address of the AP. However, several service set identifiers (SSIDs) can be configured on a single AP. So it is possible that an MN can switch between two SSIDs and change its network-layer configuration while remaining connected to the same AP.
- *IP address*: IP address identifiers which may need to be resolved to IP addresses using methods that may be specific to the wireless network. For example, if the old foreign agent (oFA) or MN determines that the IP address of the new foreign agent (nFA) is equal to the oFA's address, then the layer 3 handoff does not need to be initiated [20].
- *Subnet prefix*: Subsequent to a layer 2 handoff, an MN detects a change in an on-link subnet prefix that would require a change in the primary care-of-address. For example, a change of AR typically results in a layer 3 handoff [1].

3.3. Evaluating layer 2 hints for layer 3 movement detection

Because certain layer 2 technologies are capable of providing various link status information to the IP module, such as connected or disconnected, the link identifier can help the IP module make intelligent decisions regarding configuration changes. In

order to evaluate the catalogue of hints available, some procedure must be employed to determine the likelihood that an MN is making a layer 3 movement and requires the initiation of a layer 3 signaling protocol. In [22], a likelihood function is defined and used to evaluate a wide variety of layer 2 hints for layer 3 movement detection. The estimate of layer 3 movement occurs on two levels: (1) movement between subnets and (2) movement between networks, where the MN requires a vertical, or inter-system, handoff. The likelihood function is evaluated as follows:

$$L = E \sum_i w_i L_i, \tag{1}$$

where i is an index representing the particular layer 2 hint parameter as outlined in the list above, L_i is the probability that a layer 3 handoff is required, given a change in the corresponding layer 2 hint parameter, and the weighting factor, w_i is the probability that the layer 2 parameter change would indicate that a handoff is needed. E is a parameter added for more efficient processing of the likelihood function. For example, if the subnet address has not changed, then the E parameter is set to zero, which reduces the function to a zero likelihood of layer 3 movement. Two thresholds are set for the evaluation of the result of Eq. (1). If the result passes the first threshold, a layer 3 movement is assumed. If a second threshold is passed, then an vertical movement is also assumed. Otherwise, a layer 2 handoff is assumed. The equivalent layer 3 movement estimation is shown in Fig. 5.

The layer 3 movement estimation starts with layer 2 hints collection. When any parameter, labeled E_i, is available, a quick judgment will be deployed to decide if an layer 3 handoff will be triggered. Otherwise, further steps will be implemented to decide if the layer 3 handoff is an intra- or inter-system handoff based on the value of Eq. (1). If no E parameters are available, the layer 3 movement estimation will be based only on likelihood factors L_i (e.g., the bandwidth and BER measured at the MN). When the value of Eq. (1) is larger than the first (horizontal handoff) thresh-

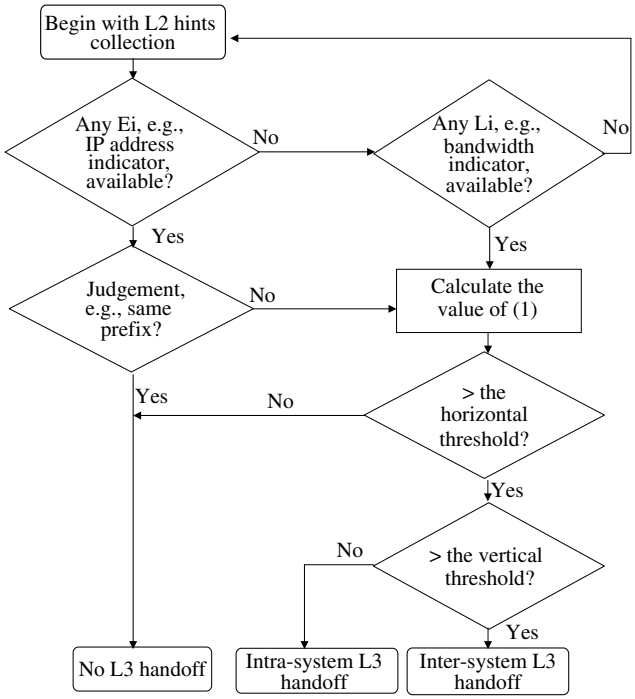


Fig. 5. Layer 3 movement estimation process.

old, then a layer 3 handoff with the same access network is considered to be necessary. If the value is also larger than the second (vertical handoff) threshold, it indicates that the MN is moving into a new subnet of different access network. Consider again the subnet prefix hint. If the prefix from the new AP is exactly the same as the prefix from the old AP, the MN can draw the conclusion that it is not moving to a new subnet, so that the probability of layer 3 movement is minimized in Eq. (1), resulting in only a layer 2 handoff being initiated, and avoiding all of the higher layer signaling. On the other hand, if the prefix has changed, the likelihood function is evaluated to see if the layer 3 signaling may be necessary.

3.4. Use of layer 2 hints in Mobile IP handoff

Researchers have recently begun to explore protocols that implement the use of layer 2 hints when executing Mobile IP handoff. An IETF draft in 2004 [23], presented a mechanism that extends Mobile IPv6 by including link events information to optimize layer 3 movement detection. The work considers smooth handoffs for MNs that are equipped with multiple interfaces moving across different and heterogeneous links. In particular, the use of link-up, link-down, and link-type hints were recommended for Mobile IP nodes moving between 802.11 and GPRS.

Pre- and post-registration techniques were described in [20] to achieve low latency Mobile IP handoffs and to allow greater support for real-time services. The purpose of pre-registration is to reduce handoff delay by anticipating the need for handoff, and beginning the layer 3 handoff signaling before the handoff begins. Optimally, the pre-registration signaling is complete even before the first packet is redirected to the new location of the MN. The post-registration handoff method uses tunnels to perform low latency changes in the MN's layer 2 point of attachment without requiring any involvement from the MN. Following a successful registration between the MN a foreign agent, when the MN moves to a new foreign agent, it can defer the layer 3 handoff and continue to communicate via the old foreign agent. This technique minimizes the period of time when an MN is unable to send or receive IP packets due to the delay in the Mobile IP registration process.

Finally, the layer 2 hint is used as an input to a handoff decision process in [24]. An algorithm for handoff initiation and decision is developed, based on the policy-based handoff framework introduced by the IETF. A cost function is designed to allow networks to judge handoff targets based on a variety of user- and network-valued metrics. These metrics include layer 2 hint parameters, as well as other quality of service metrics. The optimal handoff target is decided by evaluating the cost function. A performance analysis demonstrates that significant gains in quality of service and a more efficient use of resources can be achieved from the proposed technique.

Traditionally, quality of service has included the treatment of data traffic. An emerging area of quality of service is the security protections that are available in a network, where security includes securing the communication channel, the equipment, and authenticating and authorizing the mobile users. In the next section, we outline the current research issues for implementing authentication for Mobile IP networks.

4. Mobile IP registration research—secure authentication

Registration is the network management process that authenticates mobile users [25]. Traditionally, registration research is concerned with tracking the mobile user's current location. Part of the tracking problem is addressed by the Mobile IP address discovery and hierarchical routing techniques discussed in Section 2. However, the authentication problem is currently significant due to the increasingly distributed nature of wireless networks. Familiar research on location registration is based on a global cellular system, which has a controlled, centralized authentication architecture.

However, future Mobile IP based authentication architectures may be centralized, distributed, or locally centralized. A typical example of centralized architecture is that an authentication center (AuC) is used in UMTS networks for security

support specified by European Telecommunications Standards Institute (ETSI). The UMTS standard claims that it provides security protection from the aspects of network access security, network domain security, user domain security, application security, security visibility and reconfigurability. In an ad hoc network, distributed architecture is used because a mobile user only registers with a local network for multi-hop communications [26,27]. In a locally centralized architecture, a number of autonomous networks may share an authentication center, such as in Mobile IP networks. In this paper, we focus on the architecture design of authentication in Mobile IP networks.

4.1. Mobile IP authentication

Authentication in wireless networks is defined as a security technique to protect the network against acceptance of a fraudulent transmission by establishing the validity of a transmission, a message, or an originator [28]. During the authentication process, a user must provide verifiable credentials to access a network. In particular, when an MN requires a service from a network other than its home network from which it subscribes the service, it must provide sufficient individual information for authorization and register its location to the home network for subsequent services. This process of authentication and registration plays an important role in protecting the confidentiality and integrity of wireless networks since it is the first step in denying an unauthorized transmission and preventing intrusions [29–33].

In the basic Mobile IP architecture, an *authentication extension* (AE) is defined for registration messages, comprising a *security parameter index* (SPI) referring to some previously defined mobile security association (MSA), and an authenticator, which is calculated using a keyed-hash function. *Mobile IPv4 challenge/response extensions* (MICRE) [31] is developed for basic Mobile IP architecture. This protocol provides replay protection for all messages exchanged with Mobile IP protocol by defining two new types of message extensions: *A challenge extension* for foreign agent advertisement messages and a *mobile challenge response extension* for registration messages.

When an MN wants to authenticate itself, it must send an authentication request message with the challenge value received from the foreign agent advertisement. By checking the challenge value to see if it has already been used, the foreign agent can avoid a malicious relay attack from an MN. The verification of the challenge value depends on the security association between the MN and its home agent, while the security association between the foreign agent and the MN may not exist. The authentication of this protocol must rely on an external authentication system in assumption that a secure notification can be returned. From the procedure described above, we can see that this protocol cannot protect the messages exchanged between foreign agent and MN, and between foreign agent and home agent, although it can avoid replay attacks.

Alternatively, *secure scalable authentication* (SSA) is aimed at providing Mobile IP with a strong, scalable authentication mechanism based on public key cryptography! [34]. In SSA, when an MN is moving close to a foreign agent, it receives an advertisement with authentication extension and certificate extension broadcast by the foreign agent. The MN then extracts and validates the certificate with a public key issued by a certificate authority. After the verification, the MN uses the public key of the foreign agent from the certificate to verify the digital signature in the foreign agent authentication extension, which is created using the foreign agent's private key. After that, the MN will obtain the secret key of the foreign agent; thus, the communication between the MN and the foreign agent will be protected.

In 2002, IETF published another draft using diameter protocol in Mobile IP networks [29]. A Diameter server is defined as an authority center, which is able to authenticate, authorize and collect accounting information for Mobile IPv4 services rendered to an MN. The Diameter is intended to provide an AAA framework for applications, such as network access or IP mobility, and work in both local AAA and roaming situations. Based on the standard security association model in wireless networks, DIAMETER is being deployed as a more flexible successor to the widely-deployed RADIUS protocol for authentication, authorization and accounting.

The reasons for the continuing development of new schemes and technologies rather than using available mobility support technologies, include (1) developing a distributed architecture, composed of many authentication servers for delivering authentication messages; (2) ensuring scalability, that is, authentication mechanisms for wireless networks must be scalable enough to adapt to various user densities and roaming patterns; (3) increasing security services, including information secrecy, data integrity, and confidentiality; and (4) reducing the overhead for the negotiation of encryption/decryption algorithms, encryption/decryption of messages, transmission of messages and credential verification.

Many new schemes and protocols have been proposed to satisfy the constraints of these factors in recent years [29,30,35–38].

4.2. Design of authentication architecture for Mobile IP networks

In order to provide security in wireless and mobile networks, *Mobile IP with AAA extension* is proposed in which basic authentication, authorization and accounting requirements are introduced. Some of them are highlighted as follows and more details can be found in [37].

- Each local attendant must have a security relationship with a local AAA server (AAAL). An attendant is a node designed to provide the service interface between a client and the local domain. An AAAL is an identity which can authenticate the local MN and is expected to be configured with enough information to negotiate the verification of client credentials with external authorities.
- Since the MN's credentials need to remain unforgeable, intervening nodes (e.g., neither the attendant nor the AAAL or any other intermediate nodes) must not be able to learn any (secret) information which may enable them to reconstruct and reuse the credentials.
- Attendants should be configured to obtain authorization from a trusted AAAL for QoS requirements placed by the client.
- The local authority must share, or dynamically establish, security relationships with external authorities that are able to check client credentials.

In the proposed basic architecture shown in Fig. 6(A), each AAAL should share security association with a home AAA server (AAAH) of the roaming MN in current domain so that the AAAL can securely transmit MNs' credentials. In this configuration, the local and the home authority share the trust relationship, mutually. Depending on the security model used, this configuration can cause a quadratic growth in the number of trust relationships, as the number of AAA authorities (AAAL and AAAH) increases. This has been identified as a problem by the roamops working group [35], and any AAA proposal must solve this problem. Using brokers is a possible solution to

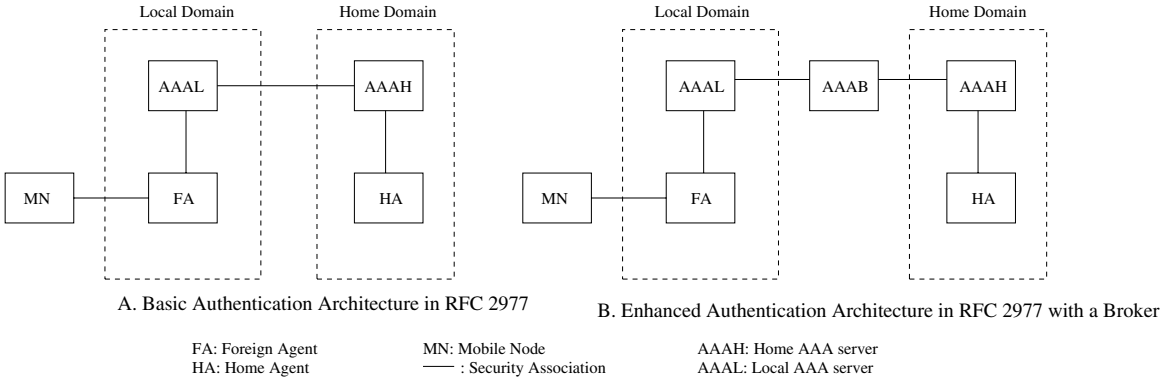


Fig. 6. Basic and enhanced authentication architecture in Mobile IP.

the scalability problems associated with requiring direct business/roaming relationships between every two administrative domains. In order to provide scalable networks in many service providers and large numbers of private networks, multiple layers of brokers should be supported like the broker model described in Fig. 6(B).

A broker may play the role of a proxy between two administrative domains, which have security associations with the broker, and be able to relay AAA messages back and forth securely. Though this mechanism may reduce latency in the transmission of messages between the domains after the broker has completed its involvement, there may be a lot of overhead messages as a result of additional copies of authorization and accounting to the brokers. There may also be additional latency for the initial access to the network, especially when a new security association needs to be created between AAAL and AAAH. These delays may become important factors for latency-critical applications.

AAA extension with mobility support is proposed for seamless roaming and mobility support in combination with AAA extensions for inter-domain roaming among heterogeneous networks [38,39]. When an MN moves out of the coverage of its home network, the network address which was assigned previously, such as active IP session, is useless. To efficiently solve these problems, a common architecture is developed for handling inter-system terminal mobility with Mobile IP authentication architecture. In this architecture, mobility support is integrated with AAA functions through carefully designed signaling messages. In other words, before a foreign agent confirms the registration of a visiting node, it contacts a foreign AAA server with an access request message. As a result, AAA functions are completed along with the registration, thus reducing the number of packets exchanged.

In [38], a new authentication entity, security gateway (SGW), is added to provide a secure communication segment between the roaming MN and its home agent without requiring that the foreign network be trusted or participate in the process. Based on this authentication architecture, IPSec tunnel mode is enabled between the SGW and its MH by which the SGW sets up security association for each MH in its network. The home agent is only responsible for Mobile IP registration and relaying packets to MH's care-of-address (CoA). The MH is protected by the IPSec tunnel between the SGW and MH. Therefore, IPSec is applied to Mobile IP for providing security services with mobility support.

4.3. A local authentication control scheme for Mobile IP networks

In order to reduce authentication cost, an authentication, authorization and accounting (AAA) architecture is proposed for wireless networks [37]. In this architecture, an AAA server is a central server in one autonomous network with hop-by-hop SAs between AAA servers for authentication. Based on this architecture, the credentials are delivered from a local AAA server (LAS) to the home authentication server (HAS) for authentication when the MN is roaming in foreign networks, regardless of the traffic and mobility patterns of the MN as well as the distance between the MN and its HAS. These operations deteriorate the QoS with expensive authentication cost and long latency, especially for two networks far from each other [37,29,40]. In addition, remote authentication imposes heavy cost burden on servers because hop-by-hop encryption/decryption is applied due to the lack of a direct SA in the AAA architecture.

In this section, we introduce a local authentication protocol, which is able to securely establish a local SA for inter-domain roaming MNs and produce challenge/response values for local authentication [41]. This scheme not only reduces authentication delay and cost significantly, and it is also feasible in any wireless networks using the AAA architecture.

4.3.1. Security association and challenge/response based authentication

As defined in IP security architecture (IPsec), security association (SA) is a trust relationship between a sender and a receiver for secure data transmission. It has many parameters, such as security parameters index (SPI), key and lifetime, all of

which can be used to serve for encryption and authentication [42]. An SA can be established and modified with protocols such as Internet security association and key management protocol (ISAKMP), secure socket layer (SSL) or transport layer security (TLS). In these protocols, SSL and TLS are two protocols commonly used in mobile networks. SSL is a standard for encrypted client/server communication between network devices. It works by using a public key to encrypt and transfer data. TLS is an IETF standard with the goal to produce an Internet standard version of SSL [43].

In order to facilitate the authentication in mobile networks, secret key based authentication is widely adopted [31]. In particular, challenge/response authentication requires the roaming MN to submit a response value for authentication *each time*, which is encrypted from a *challenge value*, a random value, with an SA shared between the MN and its home network. The challenge and response values are delivered to the home network of the MN for verification. An authentication approval message will be returned if the authentication is granted. However, when an MN initiates a service request or crosses the boundaries of subnetworks, authentication will be triggered. Thus, frequent authentication requests impose a great burden to deliver the authentication messages between networks, which are related with the mobility and traffic patterns of MNs.

4.3.2. Overview of local authentication control scheme

The framework of the proposed scheme is illustrated in Fig. 7. When an inter-domain authentication request from a visiting MN comes to a local authentication server (LAS), the LAS first checks if a local SA exists for the MN. If the local SA exists, the LAS authenticates the roaming MN with this SA. Otherwise, the LAS checks if the residence time of the MN will be greater than a threshold value. There are many methods to estimate the residence time of an MN [44]. In our paper, we assume that the estimation result of the residence time exists. Then, if the residence time of the MN is greater than a threshold value, the LAS will authenticate the MN through the AAA architecture and generate a local SA for it. Otherwise, the LAS simply authenticates the MN though the AAA architecture and does not generate a local SA for it.

If a local SA is generated, we assign the value of residence time of the MN to the life time of the SA. The sequential authentication requests that will arrive within the life time of the local SA can be

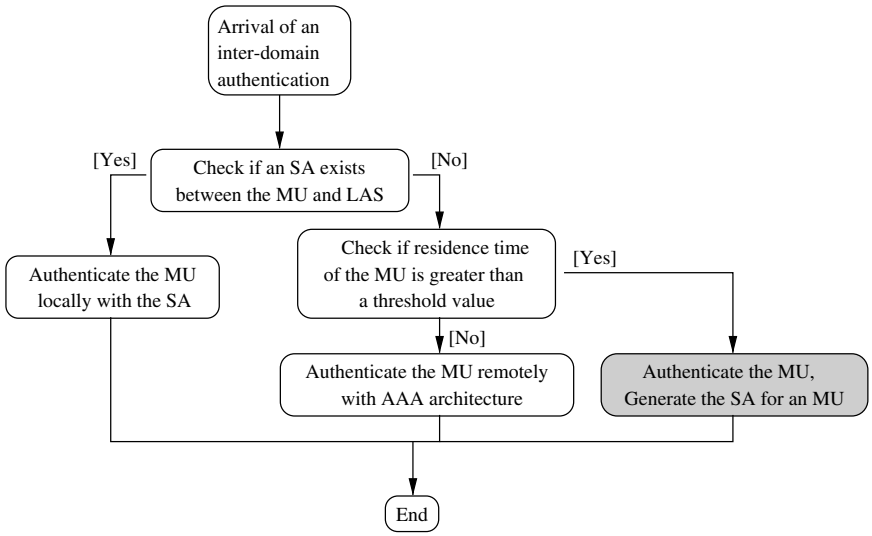


Fig. 7. Overview of local authentication control scheme.

processed efficiently with the local SA without transmitting the credentials to the HAS of the MN. We focus on the establishment of the local SA in a mobile network for the roaming MN, which is highlighted in Fig. 7. The establishment of a local SA involves with two problems. One is how to distribute the key securely and efficiently; the other is how to determine the threshold value of residence time, i.e., the lifetime of an local SA, which is used to trigger proposed scheme.

4.3.3. Authentication and local SA establishment protocol

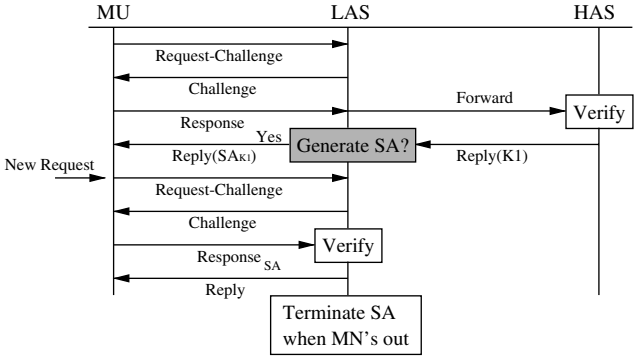
The signaling diagram of the protocol to authenticate a roaming MN and establish a local SA for sequential authentication requests is shown in Fig. 8. As we can see in this diagram, when a foreign MN is requesting services in the local network, an authentication request is sent out to the LAS. The LAS replies a challenge, i.e., a random value, to the MN. The MN encrypts the challenge value with an SA shared with the HAS. The result of the value is a response value that is returned to the LAS. Because the LAS has no SA shared with the MN, the LAS relays the response value to the HAS of the roaming MN through the AAA architecture. The HAS of the MN decrypts the response value and compares the result with the challenge value transferred by the LAS. If these two values are matched, the MN is authenticated.

After the above operations are finished, a local SA can be established at the visiting MN and the LAS as follows:

$$\mathrm{SA} ::= \{\mathrm{UID};\ \mathrm{SPI};\ \mathrm{ALGORITHM};\ \mathrm{DIRECTION};\ \mathrm{KEY};\ \mathrm{LIFETIME}\},$$

where UID is the *unique user identification*, which indicates the user for whom the local SA is used. In the local SA at the LAS, UID is the identification of the MN. In the local SA at the MN, UID is the identification of the LAS. SPI (Security Parameter Index) is the identification number of the association, which is used to differentiate the SAs uniquely. ALGORITHM is a description on a specific algorithm that should be used with this local SA. DIRECTION specifies the association used for packets arriving or leaving, KEY provides the encoding and decoding key for the authentication, which is $K1$ in our proposed protocol. LIFETIME is a time period to keep the SA, which is determined and transferred by the LAS.

From the protocol, we can see that the security to distribute the key for the local SA between the MN and the LAS, is guaranteed because the messages transmitted between the AAA servers are encrypted with a pair of SAs with nonce technique. Thus, information secrecy and data integrity are provided and replay attack can be defeated. Second, the transmission of this key to the visiting MN is protected through a random value with



LAS: Local Authentication Server SA: Security Association

HAS: Home Authentication Server MU: Mobile User

Fig. 8. Authentication and local SA establishment protocol.

an SA shared between the MN and its HAS, which avoids direct key distribution on the unprotected medium and guarantees secure transmission from the HAS to the MN. Thus, by establishing a local security association with concern of traffic pattern, mobility pattern, and number of hops between the mobile user and its home authentication server, the proposed scheme becomes a promising alternative for secure and efficient authentication approach in various wireless networks.

4.4. Authentication protocol comparisons

For different authentication architectures, we compare aforementioned protocols in Table 1 according to following criteria for suitability and acceptance in wireless IP networks:

- Security: An authentication protocol should be able to restrict services to authorized users only. In addition, it should protect the network against internal and external security threats.
- Efficiency: The authentication protocols should induce little overhead and computing requirements.
- Scalability: The authentication protocols should also be deployable in large networks with many mobile users and frequent handoffs.
- Transparency: The authentication protocols should require as little change as possible to existing systems, especially to MNs and CNs.
- Manageability: In massive networks, it is crucial that protocol mechanisms can be efficiently managed. For security protocols, this includes key management, policy management, and access control.

Future wireless networks will suffer from diverse standards that limit the authentication for roaming users. There are many challenging issues related to authentications in mobile networks. One of them is the *integration of WLAN and WWAN*. Most of the existing authentication mechanism are inappropriate for supporting inter-domain, especially, roaming in heterogeneous systems such as the integration of WLANs and WWANS, which is clearly needed in future trends to provide public access to information systems. The ability to maintain a secured connection without re-association and re-authentication needs to be investigated and provided.

5. Mobile IP research—mobility modeling

The evolution scenario to an all-IP wireless system both obligates and enables a strict performance evaluation for wireless networks, which has been the driving force for the search for better mobility models. The mobility patterns of the mobile users in a wireless network directly affects data and signaling traffic due to location registration, paging and handoffs. Mobility of the users also affects the call holding and sojourn times [45]. All these factors are vital in the performance evaluation of any scheme in wireless networks. The mobility models in the literature can be analyzed according to several criteria, including location information and mobile independence. In this section, we survey most of the frequently used mobility models in the literature that fall into these categories.

Table 1
Comparison of authentication protocols

Protocol	Security	Efficiency	Scalability	Transparency	Manageability
IP mobility support	(−)	(+)	(−)	(+)	(−)
MICRE	(−)	(+)	(−)	(+)	(−)
SSA	(+)	(−)	(+)	(−)	(+)
DIAMETER	(+)	(+)	(+)	(−)	(+)
IPSec	(+)	(−)	(+)	(+)	(+)
SCPCS	(+)	(+)	(+)	(−)	(−)
(+)	Advantage		(−)	Disadvantage	

5.1. Location information

One method for obtaining the location information is observation and recording of traces of user activity in existing wireless networks [46–51]. Although a trace provides a realistic model, it also has some drawbacks. A large number of mobiles must be tracked for long observation periods, resulting in huge amounts of data to be stored. The observations should also be repeated several times, and the results should be averaged to provide statistical data and compensate for any anomalies. Furthermore, an existing system is required for the observation to be performed. However, traces that belong to the old type of wireless networks do not necessarily apply to a new network. For example, the mobility patterns and call holding times from the traces of a 2G voice network does not hold for a 3G network, which supports both voice and data connections including multimedia. Moreover, the available location information is generally at the cell level for active users and location area level for inactive users. The location information also depends heavily on the cellular layout of the existing system and is not helpful for calculating signal interference or cell radius analysis. A final drawback for traces is the fact that most service providers prefer to keep trace information confidential, forcing researchers to devise alternative models.

On the other hand there are *synthetic models*, which includes analytical and simulation based models, which use mathematical and computational methods but not trace data. *Analytical mobility models* allow derivation of mathematical expressions for the performance of the system. However, the simplifying assumptions made on the mobility patterns make these assumptions unrealistic. *Simulation-based models*, alternatively, do not generally allow derivation of mathematical expressions, but provide more detailed and realistic mobility patterns. To obtain statistically dependable results, the results of several simulation runs must be averaged and confidence intervals should be analyzed. The mobility models studied in this paper are given in Fig. 9.

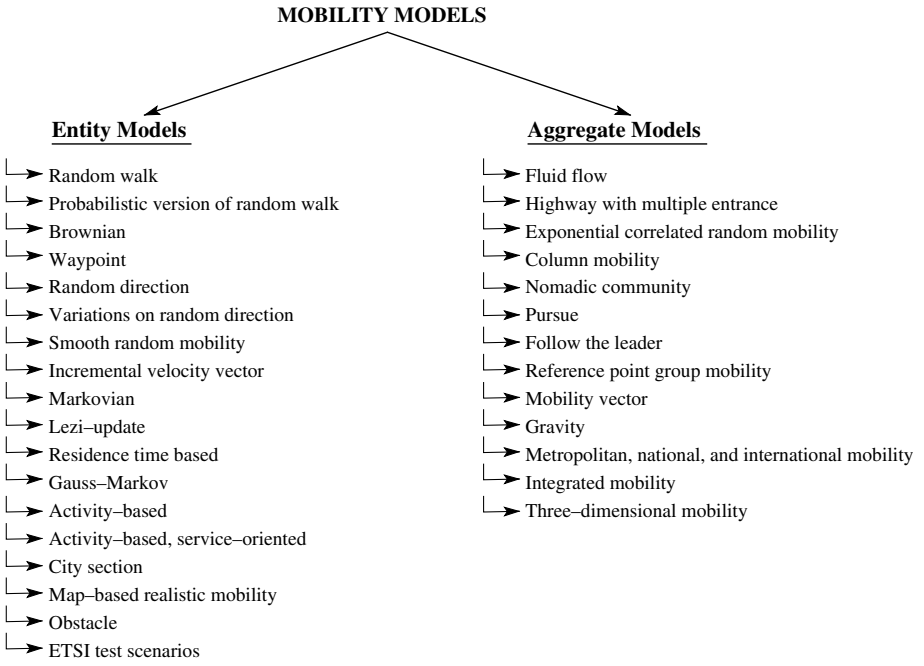


Fig. 9. Classification of mobility models.

5.2. Mobile independence

The independence of mobile users determines the relationship between the mobility patterns of different users. *Entity mobility models* determine the mobility pattern of each user individually. The location, speed, direction, and call generation process of each mobile is managed exclusively. This approach allows handling and monitoring the service provided to each user. Some entity models allow the mobiles to display similar but independent patterns (e.g., [52]) while most entity models disregard this point. The mobiles in such models do not demonstrate real life features like *moving-in-groups*. Alternatively, *aggregate (group) mobility models* determine the mobility patterns for groups of users. This class of models display moving-in-groups behavior inherently and provide statistics like mean number of users in area or mean number of users that cross a cell or location area boundary. Aggregate mobility models cannot be used for cases that require per user information, for example, to calculate actual signal interference.

Finally, it is important if a mobility model is *memoryless* or not. In a memoryless model, the next direction and speed of a mobile is independent of its current direction and speed. This approach results in mobiles making sudden turns at high speeds, a mobility pattern that is not observed in real life. With the introduction of the *inertial behavior*, mobiles tend to refrain from sudden changes in speed and direction. Inertial behavior together with moving-in-groups feature and considering physical structures provides *conscious travelling*.

In the following sections, we describe some of the most frequently used mobility models in the research literature.

5.3. Random walk model

Random walk is the simplest and the most frequently used mobility model with slight modifications [45,53–55]. Unfortunately, the term is often used incorrectly instead of Markovian or even any probabilistic mobility model. In its most general form, random walk selects the speed and direction for each user from uniform distributions over $[v_{\min}, v_{\max}]$ and $[0, 2\pi]$, respectively, where $v_{\min}$ is typically zero. The mobile travels with the selected speed and direction for a fixed time interval of Δt, and then selects new speed and direction.

In [45], Zonoozi and Dassanayake also develop a mathematical formulation for systematic tracking of the random movement of the mobile. The authors show that the cell residence time can be described by the gamma distribution, and the channel holding time by the negative exponential distribution.

5.3.1. Probabilistic version of random walk model

In [56], Chiang uses the matrix

$$\mathbf{P} = \begin{bmatrix} P(0,0) & P(0,1) & P(0,2) \\ P(1,0) & P(1,1) & P(1,2) \\ P(2,0) & P(2,1) & P(2,2) \end{bmatrix}$$

to determine the x and y coordinates of the next location of the mobile where states 0, 1, and 2 represent the current, previous, and next positions of the mobile, respectively. When the matrix is set properly, this model produces probabilistic mobility patterns rather than random. Chiang uses the following values for the **P** matrix:

$$\mathbf{P} = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 0.3 & 0.7 & 0 \\ 0.3 & 0 & 0.7 \end{bmatrix}.$$

These values ensure a mobile does not move from previous to next location without visiting current location. The state transition diagram corresponding to these values is given in Fig. 10 [56].

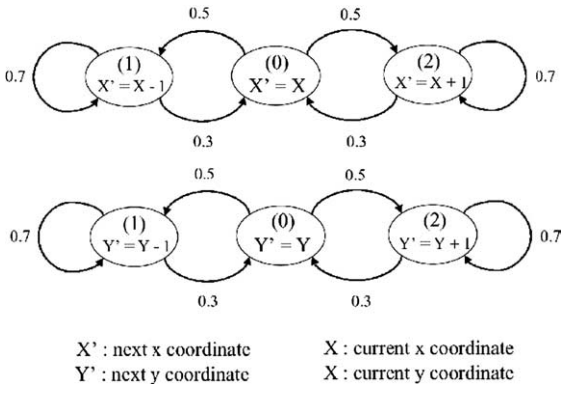


Fig. 10. Example state transition diagram.

5.3.2. Brownian model

Brownian motion is closely related to random walk. The roots of Brownian motion date back to 1785, but the first mathematical theory was developed by Albert Einstein in 1905, bringing him the Nobel prize [57]. The theory explains the erratic and constant movement of tiny particles when they are suspended in a fluid or gas, but it can also be used to generate mobility patterns for wireless users. Given the location of a user at time t_0, Brownian motion allows us to calculate the probability distribution of the physical location of the mobile at time $t > t_0$ [58,59].

5.3.3. Waypoint model

Waypoint model is also a variation on random walk model [60–67]. In the waypoint model, a mobile pauses for some time before selecting new speed and direction. Waypoint model causes accumulation of mobiles at certain parts of the service area. These gatherings are called *density waves* [68]. The density waves generally occur close to the center of the service area. In [69], Boleng notes that the average number of neighbors in the beginning of a simulation may be significantly different from that in the rest. Therefore, the use of warmup period is suggested.

Though frequently used in the literature due to its simplicity, the waypoint model has a significant drawback. Mobile speeds, which are randomly drawn from the $[v_{\min}, v_{\max}]$ interval, are assumed to have an average that remains the same during the simulation. However, in [70] Yoon et al. show that the increasing number of slow moving mobiles will dominate the mobile society. Thus, the average mobile speed will decay until it converges to the *steady state average speed* ($\overline{V}$) bounded by $\overline{V} < \frac{v_{\min}+v_{\max}}{2}$. The authors also prove that $\overline{V} \to 0$ as $v_{\min} \to 0$. The decay period is longer for small $v_{\min}$ values and it takes longer for the system to stabilize. For the case $v_{\min} = 0$, the results of the simulation should be treated with suspicion. In [71], it is shown that any statistical model that draws destination independent of speed suffers from the speed decay problem. To solve this problem, Yoon et al. suggest that the initial speeds are drawn from the steady state speed distribution.

5.3.4. Gauss–Markov model

In Gauss–Markov model, the initial values assigned for the speed and direction of each mobile is updated with fixed time intervals [72,73]. The speed and direction at the nth instance are calculated as

$$s_n = \alpha s_{n-1} + (1-\alpha)\bar{s} + \sqrt{(1-\alpha^2)}s_{x_{n-1}},$$
$$d_n = \alpha d_{n-1} + (1-\alpha)\bar{d} + \sqrt{(1-\alpha^2)}d_{x_{n-1}},$$

where s_n and d_n are the speed and direction of the mobile at time n and $\alpha \in [0,1]$ is the tuning parameter to adjust randomness. Since the previous speed and direction are considered while calculating the new ones, the sudden turns and accelerations in the random walk and waypoint models are not observed in the Gauss–Markov model.

5.4. Random direction model

In the random direction model, after selecting a speed and direction, the mobile travels until the border of the service area [68]. Thus, instead of accumulating at the center of the service area, the mobiles travel longer distances. When the mobile reaches the border, it pauses for some time, and then selects a new speed and direction for the next step. Although the random direction model solves the density waves problem, it results in the mobiles pausing only at the boundaries of the service area.

In [68], a modified version of random direction is also proposed. Royer et al. propose allowing the mobiles to change speed and direction at any point on the way to the destination. However, this approach is actually a kind of waypoint mobility model. In [74–76], a simplified version of random direction in which mobiles travelling at constant speed are reflected from the boundary of the service area with an angle of $-\varphi_0$ or $(\pi - \varphi_0)$ where φ_0 is the angle of arrival.

In [77], the smooth random mobility model, enhances the random direction model to make the movement of the mobiles look more realistic. The model is composed of two stochastic processes, one for determining when to change the speed and the other for changing the direction. To obtain smooth mobility patterns, the previous

and current values of the speed and direction are autocorrelated. The speed is changed incrementally by the current acceleration of the mobile. Furthermore, the direction change is performed in several steps until the new target direction is achieved.

In [78], the velocity vector is updated according to the formula

$$v(t+\Delta t) = \min[\max(v(t)+\Delta v, 0), V_{\max}],$$

where $V_{\max}$ is the maximum mobile velocity (assumed to be 65 mph) and Δv is the velocity change. Then, the direction and new coordinates are determined by

$$\theta(t+\Delta t) = \theta(t) + \Delta\theta,$$
$$x(t+\Delta t) = x(t) + v(t)\cdot\cos\theta(t),$$
$$y(t+\Delta t) = y(t) + v(t)\cdot\sin\theta(t).$$

The distribution from which Δv and $\Delta\theta$ are drawn are specified with respect to maximum acceleration and maximum angular change to provide a kind of inertial behavior.

5.5. Cell tracking

In the Markovian model, the current cell of the mobile rather than the exact location is of interest. The model is represented as a state transition diagram. In the *one-dimensional Markovian model* [55], a user is in one of the three states: (1) stationary state; (2) right-move state; (3) left-move state. The link weights represent the state transition probabilities. In the case of *two-dimensional Markovian model*, there are two possible approaches:

- One state for each neighboring cell. This approach results in a very complex Markovian chain [79].
- Divide the neighbor cells into two groups with respect to a dividing line [80].

In [81], Akyildiz et al. show how to reduce the two-dimensional model for hexagonal and mesh planes so that the simplified model behaves exactly the same as the original model. In [82], Tonguz et al. propose a biased Markovian motion in which mobiles tend to move towards some of the paging areas.

In the residence time based mobility models, the mobile stays in once cell until the cell residence time expires, and then moves to another cell [83,84]. There are two methods for finding the cell residence time: at the end of each time slot the decision to leave the cell can be made based on a given probability or cell residence time can be drawn from a random variable. In the literature, exponential or gamma distribution is assumed for cell residence time.

5.6. Activity-based model

The activity based model developed in [49] is based on the concept of activities, i.e., trips with specific purposes. Activities are defined according to the travel surveys conducted by organizations such as regional traffic [85]. Given the user group, time of the day, and the previous activity, the next activity of a mobile is determined by stochastic lookups in an *activity transition matrix*, which contains the cumulative probability of transition from one activity to another. The duration of the activity is found in a similar way, looking up from an *activity duration matrix*. The path is constructed using the path and distance lookup tables. After the activity duration expires, the whole processes is repeated.

Another activity-based model in the literature is developed for service-oriented ad hoc networks [86]. The approach Breyer et al. propose is actually a collection of four models:

- *Activity model* determines transforms an abstract list of non-networking tasks into a schedule of activities. An activity is defined by its starting time and duration. The activities are weighed by a set of priorities in case of conflicts.
- *Environment model* determines where the activities take place and provides the paths constrained by the obstacles according to the motion model.
- *Motion model* breaks down the high-level movement commands to fine-grained micro-movements and constructs the mobility patterns with the help of a multi-graph based mobility model.
- *Service model* derives the type of service to be performed during the activity.

5.7. Physical structure based mobility model

Some mobility models consider the physical structures while generating the mobility patterns while others allow the mobiles to visit any point in the service area. Though the latter class of models are simpler to implement, the former class of models give more realistic results since the mobiles move on the streets rather than across buildings as in real life.

5.7.1. Three-dimensional model

In [87], Kim et al. propose a three-dimensional analytical model considering vertical motion in buildings with multiple floors. The movement of the mobile is restricted by the walls, and the mobiles are allowed to have vertical motion only at the staircases. A mobile can change his direction when he reaches a new floor.

5.7.2. City section model

The city section model [88,89] simulates users walking or driving on a grid of streets where each street has a different speed limit. For each mobile, a starting point and a destination is selected randomly. The mobile moves along the path that gives the shortest travel time, subject to the street structure and speed limits. When the destination is reached, the mobile pauses before selecting the new destination. New features such as hot and dead spots with various population densities, acceleration in case of speed changes, and additional pauses along the path can be introduced to improve the model.

5.7.3. Map-based realistic mobility model

The realistic mobility model in [52] uses a given map that is color-coded with each type of physical structure represented with a different color. The mobility patterns of the individual users are created according to a given matrix, which represents the user flow rates between the physical structures. The initial distribution of the users, speed and connection patterns are determined according to the types of the physical structures. Thus, a mobile is not allowed to drive over the buildings, or the population over the sea is significantly less than the land (or non-existent at all). Entering and leaving physical structures such as highways are allowed only at selected points through connection roads, allowing a smooth transition between streets and highways. A user on a specific type of structure prefers staying on the same type of structure and tends to keep its direction towards the destination. The important features of the model are:

- moving-in-groups behavior,
- conscious travelling towards a destination,
- inertial behavior, and
- respecting the *non-pass-through* feature of some physical structures.

The model also calculates the individual interference disseminating from each active user rather than using an approximation.

5.7.4. Obstacle model

In the obstacle model, the mobiles move towards selected destinations over the shortest paths, walking around blocking obstacles in the service area [90]. Jardosh et al. utilize *Voronoi diagrams* to construct the service area and mobility patterns. A Voronoi diagram, also known as *Dirichlet tessellation*, partitions a plane with n points into n convex polygons such that each polygon contains exactly one point and every point in a given polygon is closer to its central point than to any other [91]. The buildings are represented as rectangles of random size at random points. The pathways are constructed halfway in-between the buildings and the doorways along the sides of the buildings allow movement through the buildings. Each mobile selects a buildings as the destination, walks the shortest path to the destination along the pathways, pauses for some time, and then selects another destination. Jardosh et al. assume the obstacles completely block the signals and consider only direct path for signal reception.

5.8. Aggregate mobility models

5.8.1. Fluid flow model

The fluid flow model mimics the flow of a fluid and defines mobility in terms of the mean number of mobiles crossing the boundary of a given area [92,93]. The volume of traffic from region i to j is

formulated in proportion to the population in region i, average velocity, and the length of the boundary between two regions. Assuming a circular region with a population density of ρ, average velocity of v, diameter of L, the average number of mobiles that cross the region boundary, N, is given by

$$N = \rho \pi L v.$$

In [94], Rose developed a more sophisticated model based on diffusion processes.

In [95], Leung et al. develop a deterministic fluid model that models traffic on a semi-infinite highway. Entrance and exists to/from the highway traffic are allowed only at fixed points. The proposed model captures the overall dynamics of the system rather than the behavior of the individual mobiles. The model treats the number of non-calling and calling mobiles in location $(0, x]$ at time t, $N(x,t)$ and $Q(x,t)$, respectively, as continuous fluids. Thus, call traffic load in the cells along the highways and handoff rates can be obtained. Leung et al. have also developed a related stochastic traffic model that captures the stochastic fluctuations.

5.8.2. Exponential correlated random mobility model

Bergamo et al. propose [96] an aggregate mobility model in which the new position of a mobile or group of mobiles, $\vec{b}(t+1)$, is calculated based on the previous position, $\vec{b}(t)$ and a random deviation, $\vec{r}$, as

$$b(t+1) = b(t) \cdot \mathrm{e}^{-1/\tau} + \left(\sigma \cdot \sqrt{1 - (\mathrm{e}^{-1/\tau})^2}\right) \cdot r,$$

where $b(t) = (r, \theta)$ is defined for a group or mobile at time t, τ adjusts the rate of change from the previous position of the mobile to the new position (small τ implies large change), and r is a random Gaussian variable with variance σ. Thus, the mobiles in a group exhibit similar patterns subject to some random deviation $\vec{r}$. The drawback of this model is that generating a specific mobility pattern by adjusting τ and σ is not an easy task.

5.8.3. Column mobility model

In the column mobility model, a group of mobiles move in a line or column in a given reference grid [57]. Each mobile moves in restricted vicinity of its associated reference point according to one of the entity mobility models described above. The reference mobility points are displaced by an advance vector, which moves the reference grid by a random distance and a random angle. Thus, the reference points of the mobiles are stationary relative to each other. This model can be used to represent a group of mobiles with the same destination or purpose, such as a squad of soldiers moving together.

5.8.4. Nomadic community model

This model mimics the mobility patterns of the ancient nomadic societies [57]. Each mobile in the group roams around the reference point using an entity mobility model. All mobiles in the same group share the same reference point as opposed to the column mobility model where each mobile has its own reference point. As a consequence, the mobiles in the nomadic community model are less constrained compared to those in the previous model. The column mobility model can be used for ad hoc systems where the mobiles move with a strict formation, such as military applications, and nomadic community model can be used for systems with a more relaxed formation, such as civilian applications.

5.8.5. Pursue model

The pursue model mimics the mobility pattern of several mobiles tracking the same mobile target [57], such as flies tracking a person with an apple-pie in hand or several police cars following a suspect. The mobiles try to reach the target by making random moves. The new position of each mobile is calculated by using a random vector and an acceleration function as

$$\begin{aligned} new\_pos = {} & old\_pos + accel(target - old\_pos) \\ & + random\_vector, \end{aligned}$$

where $accel(target - old\_pos)$ is information on the movement of the mobile under pursuit and $random\_vector$ is a random offset obtained from an entity model for each mobile [88]. The movement of a mobile is limited from above and below to enforce a maximum speed limitation and to maintain a certain random orbit around the target point when it is reached.

5.8.6. Follow the leader model

Follow the leader model is based on car following theory, a.k.a. GM model [97]. Considering a sequence of drivers $1,2,\ldots,n,\ldots$, the speed and acceleration of the nth driver are denoted by $\dot{x}_n(t) = \mathrm{d}x_n/\mathrm{d}t$ and $\ddot{x}_n(t) = \mathrm{d}^2x_n/\mathrm{d}t^2$, respectively. The relationship between the mobility patterns of the nth driver and his leader, $(n-1)$th driver, is expressed using a sensitivity coefficient λ as

$$\ddot{x}_n(t) = \lambda(\dot{x}_{n-1}(t) - \dot{x}_n(t)), \tag{2}$$

$$\lambda = \lambda_0 \frac{(\dot{x}_n(t))^m}{(x_{n-1}(t) - x_n(t))^l}. \tag{3}$$

Eq. (2) says if the nth driver is driving at the same speed with his leader, he will not accelerate or decelerate. Eq. (3) says that the sensitivity coefficient, λ, depends on both the speed of the driver and the distance with the leader. The values of m and l are typically non-negative, and generally assumed to be either 0 and 1, or 1 and 2, respectively.

5.8.7. Reference point group mobility model

In the reference point group model, each group has a logical center that represents the trajectory of the whole group [96]. Each mobile in the group has its own reference point. The group motion vector, $\overrightarrow{GM}$, defines the movement of the group, i.e., the reference points. Each mobile wanders randomly around its reference point using a random motion vector $\overrightarrow{RM}$. The waypoint model can be used for the mobility of both the logical centers and the individual mobiles. Typically, the logical centers have pause times between the steps while the individual mobiles move without pause times. This model can be used to depict convention, disaster recovery, and battlefield ad hoc scenarios.

The mobility vector model is more general than the reference group mobility model [98] and can describe scenarios with different types of mobiles with different mobility patterns. The mobility vector of an individual mobile, $\overrightarrow{M}$, is calculated as

$$\overrightarrow{M} = \overrightarrow{B} + \alpha \cdot \overrightarrow{V},$$

where $\overrightarrow{B}$ is the base vector that defines the primary velocity component, $\overrightarrow{V}$ is the deviation vector that defines the deviation from $\overrightarrow{B}$, and α is the acceleration factor.

5.8.8. Gravity model

Gravity model is a class of models in transportation theory rather than a single model [99,100,50]. Inspired by Newton's Law of Gravity, the volume of traffic from region i to region j, T_{ij}, is defined as

$$T_{i,j} = K_{i,j} P_i P_j,$$

where P_i and P_j represent the sizes of mobile population in regions i and j, respectively, and $K_{i,j}$ is a parameter that should be calculated (or measured) for these two regions. The drawback of these models is the requirement for calculating the $K_{i,j}$ parameters for all $\{i,j\}$ pairs.

Metropolitan, national, and international mobility models constitute a set of models at three different levels [50]. The metropolitan model includes the Markovian model as a special case. It generates movement trips for different mobility behavior classes: stationary, simple, round-trip, and return home. Conditional probabilities of moving from one site to the other is specified in a movement connectivity matrix. The other models are also based on similar concept. The national model describes the movement behaviors between metropolitan areas in USA. The gravity model used in the national model is

$$T^*_{i,j} = \frac{m_i m_j P_i P_j}{d_{i,j}^{\gamma_i+\gamma_j}},$$

where $T^*_{i,j}$ is equal to $0.5(T_{i,j} + T_{j,i})$, $d_{i,j}$ is the distance between regions i and j, $\{m_i\}$ and γ_i are parameters that should be calibrated. The international model describes the movement behaviors between USA and 10 other countries. The gravity model used in the international model is

$$T^*_j = K_j P_{\mathrm{USA}} P_j.$$

As mentioned previously, the modeling techniques chosen for the performance evaluation of a mobile network protocol affect the realistic nature of the signaling load and quality of service effects. The ability to capture the performance without physically implementing a mobile system distinguishes mathematical analysis, network simulators and network emulators from experimental approaches. Currently, the latter can realize the physical time and/or frequency selective behavior, but they are

costly and are not perfectly repeatable. On the other hand, mathematical analysis and related network simulators avoid the high experiment cost and are perfectly repeatable. However, to enable tractability, they are oftentimes based on simplifying assumptions that considerably affect the accuracy of the performance results. Therefore, although the models provided above, give an accurate picture of the current state of mobility modeling, future analyses may require the use of network emulators to combine the advantage of tractability and more realistic physical propagation channels.

6. Conclusion

As the unified, ubiquitous global wireless system continues to develop, the need for inter-operability as well as service optimization continues to grow and evolve. New techniques for integration and optimization must be developed for each layer of the network, and unifying networking protocols, such as Mobile IP must be expanded to operate under different types of networks. Finally, the problems of wireless networks, such as registration, handoff, security, and reliable performance evaluation techniques must be addressed to be able to create and produce a wide range of services to the user at the expected wireline-level service quality.

Acknowledgment

Authors McNair and Wang would like to acknowledge the US National Science Foundation (Award #0322893) in support of research presented in this article.

References

[1] C. Perkins, J. Arko, IP Mobility Support for IPv6, Tech. rep., Internet Engineering Task Force. Available from: <http://www.ietf.org/rfc.html> (June 2003).
[2] I. Akyildiz, J. Xie, S. Mohanty, A survey of mobility management in next generation all-ip based wireless systems, IEEE Wireless Communications 11 (4) (2004) 16–28.
[3] T. Henderson, Host mobility for ip networks: a comparison, IEEE Network 17 (6) (2003) 18–26.
[4] C. Perkins, IP Mobility Support for IPv4, Tech. Rep. RFC 3220, Internet Engineering Task Force. Available from: <http://www.ietf.org/rfc.html> (August 2002).
[5] J. McNair, I. Akyildiz, M. Bender, Handoffs for real-time traffic in Mobile IP version 6 networks, in: Proceedings of IEEE GLOBECOM 2001, vol. 6, 2001, pp. 3463–3467.
[6] R. Caceres, V.N. Padmanabhan, Fast and scalable handoffs for wireless Internetworks, in: ACM Mobicom, 1996, pp. 56–66.
[7] C. Castelluccia, Extending Mobile IP with adaptive individual paging: a performance analysis, in: IEEE Symposium on Computer and Communications, 2000, pp. 113–118.
[8] C.E. Perkins, D.B. Johnson, Route optimization in Mobile IP, Tech. rep., Internet Draft, Internet Engineering Task Force. Available from: <draft-ietf-mobileip-optim-11.txt> (September 2001).
[9] E. Gustafsson, A. Jonsson, C.E. Perkins, Mobile IPv4 regional registration (work in progress), Tech. rep., Internet Draft, Internet Engineering Task Force. Available from: <draft-ietf-mobileip-reg-tunnel-07.txt> (October 2002).
[10] H. Soliman, C. Castelluccia, K. El-Malki, L. Bellier, Hierarchical Mobile IPv6 mobility management (HMIPv6) (work in progress), Tech. rep., Internet Draft, Internet Engineering Task Force. Available from: <draft-ietf-mobileip-hmipv6-07.txt> (October 2002).
[11] A.T. Campbell, J. Gomez, S. Kim, A.G. Valko, C.-Y. Wang, Z.R. Turanyi, Design, implementation, and evaluation of Cellular IP, IEEE Personal Communications Magazine (2000) 42–49.
[12] A. Valko, Cellular IP: a new approach to Internet host mobility, ACM SIGMOBILE Computer Communication Review 29 (1) (1999) 50–65.
[13] R. Ramjee, K. Varadhan, L. Salgarelli, S.R. Thuel, S.-Y. Wang, T.L. Porta, HAWAII: a domain-based approach for supporting mobility in Wide-area Wireless Networks, IEEE/ACM Transactions on Networking 10 (3) (2002) 396–410.
[14] A.T. Campbell, J. Gomez, S. Kim, C.-Y. Wan, Z.R. Turanyi, A.G. Valko, Comparison of IP micromobility protocols, IEEE Wireless Communications (2002) 72–82.
[15] F.M. Chiussi, D.A. Khotimsky, S. Krishnan, Mobility management in third-generation All-IP networks, IEEE Communications Magazine (2002) 124–135.
[16] P. Reinbold, O. Bonaventure, IP micro-mobility protocols, IEEE Communications Surveys & Tutorials (2003) 40–57.
[17] A.T. Campbell, J. Gomez-Castellanos, IP micromobility protocols, ACM SIGMOBILE Mobile Computing and Communication Review 4 (4) (2001) 45–54.
[18] H. Omar, T. Saadawi, M. Lee, Supporting reduced location management overhead and fault tolerance in Mobile IP systems, in: IEEE Symposium on Computer and Communications, 1999, pp. 347–353.
[19] K. Pahlavan, P. Krishnamurthy, Principles of Wireless Networks, Prentice-Hall, 2002.

[20] K.E. Malki, Low latency Handoff in Mobile IPv4, Tech. rep., Internet Engineering Task Force. Available from: <http://www.ietf.org/rfc.html> (June 2004).
[21] A. Yegin, E. Njedjou, S. Veerepalli, T. Noel, Link-layer hints for detecting network attachments, Tech. Rep. RFC3344, Internet Engineering Task Force. Available from: <http://www.ietf.org/rfc.html> (October 2003).
[22] F. Zhu, J. McNair, Cross-layer mobility for Mobile IP networks, in: IEEE Spring Vehicular Technology Conference (VTC), 2005.
[23] S.D. Park, L2 triggers optimized Mobile IPv6 vertical handoff: The 802.11/GPRS example, Tech. rep., Internet Engineering Task Force. Available from: <http://www.ietf.org/rfc.html> (January 2004).
[24] F. Zhu, J. McNair, An optimized vertical handoff decision algorithm, in: IEEE Wireless Communications and Networking Conference (WCNC), 2004.
[25] I. Akyildiz, J. McNair, J. Ho, H. Uzunalioğlu, W. Wang, Mobility management in next generation wireless systems, Proceedings of the IEEE 87 (8) (1999) 1347–1384.
[26] S. Capkun, J.-P. Hubaux, L. Buttyan, Mobility helps security in ad hoc networks, in: ACM International Symposium on Mobile Ad Hoc Networking and Computing, 2003.
[27] S. Levijoki, Authentication, authorization and accounting in ad hoc networks, in: Proceedings of the Helsinki University of Technology Seminar on Internetworking spring 2000: Ad Hoc Networking, 2000.
[28] Telecom Glossary 2000. Available from: <http://www.atis.org/tg2k/t1g-forw.html#refs>.
[29] P. Calhoun, J. Loughney, E. Guttman, G. Zorn, J. Arkko, Diameter base protocol. Available from: <draft-ietf-aaa-diameter-17.txt>.
[30] S. Jacobs, Security and authentication in Mobile IP, in: IEEE PIMRC'99, vol. 3, 1999, pp. 1103–1108.
[31] C. Perkins, P. Calhoun, Mobile IPv4 challenge/response extensions, IETF RFC 3012.
[32] M. Xu, S. Upadhyaya, Secure communication in PCS, in: IEEE Vehicular Technology Conference, 2001, vol. 3, 2001, pp. 2193–2197.
[33] IEEE 802.11 Working Group.
[34] S. Jacobs, Mobile IP public key based authentication. Available from: <draft-jacobs-mobileip-pki-auth-02.txt>.
[35] B. Aboba, G. Zorn, Criteria for evaluating roaming protocols, RFC 2477.
[36] J.E. Katz, P. Aspden, Mobile communications: theories, data, and potential impacts a longitudinal analysis of U.S. National Surveys, in: Advances in Wireless Communications, Kluwer Academic Publishers, 1998.
[37] S. Glass, T. Hiller, S. Jacobs, C. Perkins, Mobile IP authentication, authorization and accounting requirements, Tech. Rep. RFC2977, Internet Engineering Task Force. Available from: <http://www.ietf.org/rfc.html> (October 2000).
[38] M. Barton, D. Atkins, J. Lee, S. Narain, D. Ritcherson, K. Tepe, K. Wong, Integration of IP mobility and security for secure wireless communications, in: 2002 IEEE International Conference on Communications, 2002, pp. 1045–1049.
[39] M. Cappiello, A. Floris, L. Veltri, Mobility amongst heterogeneous networks with AAA support, in: IEEE ICC 2002, vol. 4, 2002, pp. 2064–2069.
[40] H. Kim, H. Afifi, Improving mobile authentication with new AAA protocols, in: IEEE International Conference on Communications, vol. 1, 2003, pp. 497–501.
[41] W. Liang, W. Wang, A local authentication control scheme for efficient authentication in wireless networks, in: Proceedings of IEEE VTC'04, 2004.
[42] W. Stallings, Network Security Essentials, Applications and Standards, Prentice-Hall, 2002.
[43] T. Dierks, C. Allen, The TLS protocol, rfc2246.
[44] I. Chen, N. Verma, Simulation study of a class of autonomous host-centric mobility prediction algorithms for wireless cellular and ad hoc networks, in: 36th Annual Simulation Symposium, 2003, pp. 65–72.
[45] M. Zonoozi, P. Dassanayake, User mobility modeling and characterization of mobility patterns, IEEE Journal on Selected Areas in Communications 15 (7) (1997) 1239–1252.
[46] D. Kotz, K. Essien, Analysis of a campus-wide wireless network, in: ACM International Conference on Mobile Computing and Networking (MOBICOM), 2002, pp. 107–118.
[47] R. Hutchins, E.W. Zegura, Measurements from a campus wireless network, in: IEEE International Conference on Communications (ICC), 2002, pp. 107–118.
[48] D. Tang, M. Baker, Analysis of a local-area wireless network, in: ACM International Conference on Mobile Computing and Networking (MOBICOM), 2000, pp. 1–10.
[49] J. Scourias, T. Kunz, An activity-based mobility model and location management simulation framework, in: ACM International Workshop on Modeling, Analysis, and Simulation of Wireless and Mobile Systems (MSWiM), 1999.
[50] D. Lam, D.C. Cox, J. Widom, Teletraffic modelling for personal communication services, IEEE Communications Magazine 35 (2) (1997) 79–87.
[51] D. Tang, M. Baker, Analysis of a metropolitan-area wireless network, ACM/Kluwer Wireless Networks 8 (2–3) (2002) 107–120.
[52] T. Tugcu, C. Ersoy, How a new realistic mobility model can effect the relative performance of a mobile networking scheme, Wiley Journal on Wireless Communications and Mobile Computing 4 (2) (2004).
[53] J.J. Garcia-Luna-Aceves, E. Madruga, A multicast routing protocol for ad-hoc networks, in: IEEE INFOCOM, 1999, pp. 784–792.
[54] I. Rubin, C. Choi, Impact of the location area structure on the performance of signaling channels in wireless cellular networks, IEEE Communications Magazine (1997) 108–115.
[55] A. Bar-Noy, I. Kessler, M. Sidi, Mobile users: to update or not to update? in: IEEE Conference on Communications (INFOCOM), 1994, pp. 570–576.

[56] C. Chiang, Wireless Network Multicasting, PhD thesis, University of California, Los Angeles.
[57] M. Sanchez, P. Manzoni, Anejos: a java based simulator for ad-hoc networks, Elsevier Future Generation Computer Systems Magazine 17 (5) (2001) 573–583.
[58] Z. Lei, C. Rose, Wireless subscriber mobility management using adaptive individual location areas for PCS systems, in: IEEE ICC, 1998.
[59] Z. Lei, C. Rose, Probability criterion based location tracking approach for mobility management of personal communication systems, in: IEEE GLOBECOM, 1997, pp. 977–981.
[60] S.R. Das, C.E. Perkins, E.M. Royer, M.K. Marina, Performance comparison of two on-demand routing protocols for ad hoc networks, IEEE Personal Communications Special Issue on Advances in Mobile Ad Hoc Networking 8 (1) (February 2001).
[61] J. Broch, D.A. Maltz, D.B. Johnson, Y.-C. Hu, J. Jetcheva, A performance comparison of multi-hop wireless ad hoc network routing protocols, in: ACM MOBICOM, 1998, pp. 85–97.
[62] D. Johnson, D. Maltz, Dynamic source routing in ad hoc wireless networks, in: Mobile Computing, Kluwer Academic Publishers, 1996.
[63] E.M. Royer, C.E. Perkins, Multicast operation of the ad-hoc on-demand distance vector routing protocol, in: ACM MOBICOM, 1999, pp. 207–218.
[64] G. Holland, N.H. Vaidya, Analysis of TCP performance over mobile ad hoc networks, in: ACM MOBICOM, 1999, pp. 219–230.
[65] O. Avellaneda, R. Pandya, G. Brody, Traffic modelling of a cellular mobile radio system, ITC-II 1 (1985) 2-4B-1-1–2-4B-4-7.
[66] P. Johansson, T. Larsson, N. Hedman, B. Mielczarek, M. Degermark, Scenario-based performance analysis of routing protocols for mobile ad-hoc networks, in: ACM MOBICOM, 1999, pp. 195–206.
[67] C.E. Perkins, E.M. Royer, Ad-hoc demand distance vector routing, in: IEEE Workshop on Mobile Computing Systems Applications, 1999.
[68] E.M. Royer, P. Melliar-Smith, L. Moser, An analysis of the optimum node density for ad hoc mobile networks, in: IEEE ICC, 2001.
[69] J. Boleng, Normalizing Mobility characteristics and enabling adaptive protocols for ad hoc networks, in: IEEE Local and Metropolitan Area Networks (LANMAN), 2001, pp. 9–12.
[70] J. Yoon, M. Liu, B. Noble, Random waypoint considered harmful, in: IEEE INFOCOM, 2003, pp. 1312–1321.
[71] J. Yoon, M. Liu, B. Noble, Sound mobility models, in: ACM MOBICOM, 2003, pp. 205–216.
[72] B. Liang, Z. Haas, Predictive distance-based mobility management for PCS networks, in: IEEE INFOCOM, 1999, pp. 1377–1384.
[73] V. Tolety, Load Reduction in Ad hoc Networks Using Mobile Servers, Master's thesis, Colorado School of Mines.
[74] Z.J. Haas, M.R. Pearlman, The performance query control schemes for the zone routing protocol, in: ACM SIGCOMM, 1998.
[75] M.R. Pearlman, Z.J. Haas, P. Sholander, S.S. Tabrizi, On the impact of alternate path routing for load balancing in mobile ad hoc networks, in: ACM MOBIHOC, 2000.
[76] J.-H. Ryu, Y.-W. Kim, D.-H. Cho, A new routing scheme based on the terminal mobility in mobile ad-hoc networks, in: IEEE Vehicular Technology Conference (VTC), 1999, pp. 1253–1257.
[77] C. Bettstetter, Mobility modeling in wireless networks: categorization, smooth movement, and border effects, ACM Mobile Computing and Communications Review 5 (3) (2001) 55–67.
[78] Z. Haas, A new routing protocol for reconfigurable wireless networks, in: IEEE International Conference on Universal Personal Communications (ICUPC), 1997, pp. 562–565.
[79] A. Bhattacharya, S.K. Das, LeZi-Update: an information-theoretic approach to track mobile users in PCN networks, in: ACM MOBICOM, 1999, pp. 1–12.
[80] A. Burulitisz et al., On the accuracy of mobility modelling in wireless networks, in: IEEE International Conference on Software, Telecommunications and Computer Networks (SoftCOM), 2002.
[81] I.F. Akyildiz, Y.-B. Lin, W.R. Lai, R.-J. Chen, A new random walk model for pcs networks, IEEE Journal on Selected Areas in Communications 18 (7) (2000) 1254–1260.
[82] O.K. Tonguz, S. Mishra, R. Josyula, Intelligent paging in wireless networks: random mobility model and grouping algorithms for locating subscribers, in: IEEE VTC, vol. 2, 1999, pp. 1177–1181.
[83] I.F. Akyildiz, J.S.M. Ho, Y.-B. Lin, Movement-based location update and selective paging for pcs network, IEEE/ACM Transactions on Networking 4 (4) (1996) 629–638.
[84] Y.-B. Lin, Reducing location update cost in a pcs network, IEEE/ACM Transactions on Networking 5 (1) (1997) 25–33.
[85] T. Associates, Waterloo region travel survey 1987: an overview of the survey findings, Regional Municipality of Waterloo, Department of Planning and Development.
[86] T. Breyer, M. Klein, P. Obreiter, B. König-Ries, Activity-based user modeling in service-oriented ad-hoc networks, in: First Working Conference on Wireless On-demand Network Systems, 2004, pp. 43–56.
[87] T.S. Kim, M.Y. Chung, D.K. Sung, Mobility and traffic analysis in three-dimensional pcs environments, IEEE Transactions on Vehicular Technology 47 (2) (1998) 537–545.
[88] T. Camp, J. Boleng, V. Davies, A survey of mobility models for ad hoc network research, Wireless

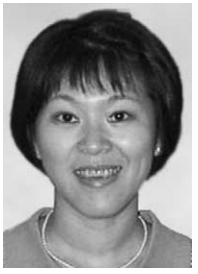
Communication and Mobile Computing Journal 2 (5) (2002) 483–502.
[89] V. Davies, Evaluating Mobility Models Within an Ad Hoc Network, Master's thesis, Colorado School of Mines.
[90] A. Jardosh, E. Royer, K. Almeroth, S. Suri, Towards realistic mobility models for mobile ad hoc networks, in: ACM MOBICOM, 2003, pp. 217–229.
[91] E.W. Weisstein, oronoi diagram. Available from: <http://mathworld.wolfram.com/VoronoiDiagram.html>.
[92] R. Thomas, H. Gilbert, G. Mazziotto, Influence of the moving of the mobile stations on the performance of a radio mobile cellular network, in: Nordic Seminar on Digital Land Mobile Radio Communications, 1988.
[93] V.S. Frost, B. Melamed, Traffic modeling for telecommunications networks, IEEE Communications Magazine 32 (3) (1994) 70–81.
[94] C. Rose, Minimizing the average cost of paging and registration: a timer-based method, ACM Journal of Wireless Networks 2 (2) (1996) 109–116.
[95] K. Leung, W.A. Massey, W. Whitt, Traffic models for wireless communications networks, IEEE Journal on Selected Areas in Communications 12 (8) (1999) 1353–1364.
[96] X. Hong, M. Gerla, G. Pei, C. Chiang, A group mobility model for ad hoc wireless networks, in: ACM International Workshop on Modeling and Simulation of Wireless and Mobile Systems (MSWiM), 1999.
[97] D.C. Gazis, R. Herman, R. Rothery, Non-linear follow-the-leader models of traffic flow, Operations Research 19 (1961) 545–567.
[98] X. Hong, T. Kwon, M. Gerla, D. Gu, G. Pei, A mobility framework for ad hoc wireless networks, in: ACM 2nd International Conference on Mobile Data Management, 2001.
[99] R.J. Bouchard, C.E. Pyers, Use of gravity model for describing urban travel, Highway Research Record 88 (1965) 1–43.
[100] P.B. Slater, International migration and air travel: global smoothing and estimation, Applied Mathematics and Computation 53 (2–3) (1993) 225–234.

Janise McNair received her B.S. and M.S. in electrical engineering from the University of Texas at Austin in 1991 and 1993, respectively, and received her Ph.D. in electrical and computer engineering from the Georgia Institute of Technology in 2000.

She is currently an Assistant Professor in Electrical and Computer Engineering at the University of Florida, where she conducts research in wireless and mobile networks, including handoff management, mobile user authentication, and medium access control. McNair has served on the committees of several ACM and IEEE workshops and currently serves on the Editorial Board of the Elsevier Ad Hoc Networks Journal and the IEEE Transactions on Mobile Computing.

Tuna Tugcu received his B.S. and Ph.D. degrees from Bogazici University, Istanbul, Turkey, in 1993 and 2001, and M.S. degree from New Jersey Institute of Technology, Newark, NJ, in 1994. He worked as a post-doc for 1.5 years and a visiting professor for 2 years at Georgia Institute of Technology.

He is now an Assistant Professor with the Department of Computer Engineering, Bogazici University. His research interests are in mobile systems, quality of service (QoS), and real-time systems. He has served on program committees for IEEE VTC'03, HWISE'05, and in organizing committee of ISCIS'05. He has been a member of IEEE and ACM since 1996 and 2004, respectively.



Wenye Wang received the B.S. and M.S. degrees from Beijing University of Posts and Telecommunications, Beijing, China, in 1986 and 1991, respectively. She also received the M.S.E.E. and Ph.D. degrees from the Georgia Institute of Technology, Atlanta, Georgia, in 1999 and 2002, respectively.

She is now an Assistant Professor with the Department of Electrical and Computer Engineering, North Carolina State University. Her research interests are in mobile and secure computing, quality of service (QoS) sensitive networking protocols in single- and multi-hop networks. She has served on program committees for IEEE INFOCOM, ICC, ICCCN in 2004. She has been a member of the Association for Computing Machinery since 2002.

Jiang Xie received her B.E. degree from Tsinghua University, Beijing, China, in 1997, 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. Her current research interests include resource and mobility management of wireless networks, QoS provisioning, and next-generation Internet. She is a member of IEEE and ACM.