WIRELESS COMMUNICATIONS AND MOBILE COMPUTING
Wirel. Commun. Mob. Comput. 2007; **7**:581–594
Published online 26 July 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/wcm.373

Air interface switching and performance analysis for fast vertical handoff in cellular network and WLAN interworking

Minghui Shi[1*,†], Liang Xu[2], Xuemin (Sherman) Shen[1], Jon W. Mark[1] and Aladdin Saleh[3]

[1]*Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario, Canada*
[2]*UTStarcom, Hangzou, Zhejiang, China*
[3]*Wireless Technology, Bell Canada, Canada*

Summary

The integration of wireless local area network (WLAN) hotspot and the 3G cellular networks is imminently the future mode of public access networks. One of the key elements for the successful integration is vertical handoff between the two heterogeneous networks. Service disruption may occur during the vertical handoff because of the IP layer handoff activities, such as registration, binding update, routing table update, etc. In this paper, the network interface switching and registration process are proposed for the integrated WLAN/cellular network. Two types of fast vertical handoff protocols based on bicasting and non-bicasting supporting real-time traffic, such as voice over IP, are modeled. The performance of a bicasting based handoff scheme is analyzed and compared with that of fast handoff without bicasting. Numerical results and the simulation are given to show that packet loss rate can be reduced by the bicasting during handoff scheme without increasing bandwidth on both wireless interfaces. Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS: WLAN; cellular; interworking; vertical handoff; VoIP; bicasting; packet loss rate; registration

1. Introduction

IEEE 802.11b, or Wi-Fi, compatible products have become the de facto standard component in mobile devices. There has been an increasing number of wireless Internet access services appearing in airports, cafes, book stores, etc. The annual industry revenue already exceeds one billion dollars and is expected to pass four billion dollars by 2007 [1,2]. In addition, the availability of mobile devices having both cellular phone and Wi-Fi capability makes preeminent demand for integrating multiple mobile computing services into a single entity.

Depending on the inter-dependence between the wireless local area network (WLAN) and 3G cellular network, there are tight coupling and loose coupling architectures. In the tightly coupled approach, the 802.11 network appears to the 3G core network as another 3G access network. In the loosely coupled approach, the 802.11 gateway connects to the Internet and does not have any direct link to 3G network elements such as packet data serving nodes (PDSNs),

*Correspondence to: Minghui Shi, Department of Electrical and Computer Engineering, University of Waterloo, 200 University Avenue W., Waterloo ON N2L 3G1, Canada.
†E-mail: mshi@bbcr.uwaterloo.ca

Contract/grant sponsors: Natural Science and Engineering Council (NSERC) Postgraduate Scholarship; Bell University Laboratories, Canada.

Copyright © 2006 John Wiley & Sons, Ltd.

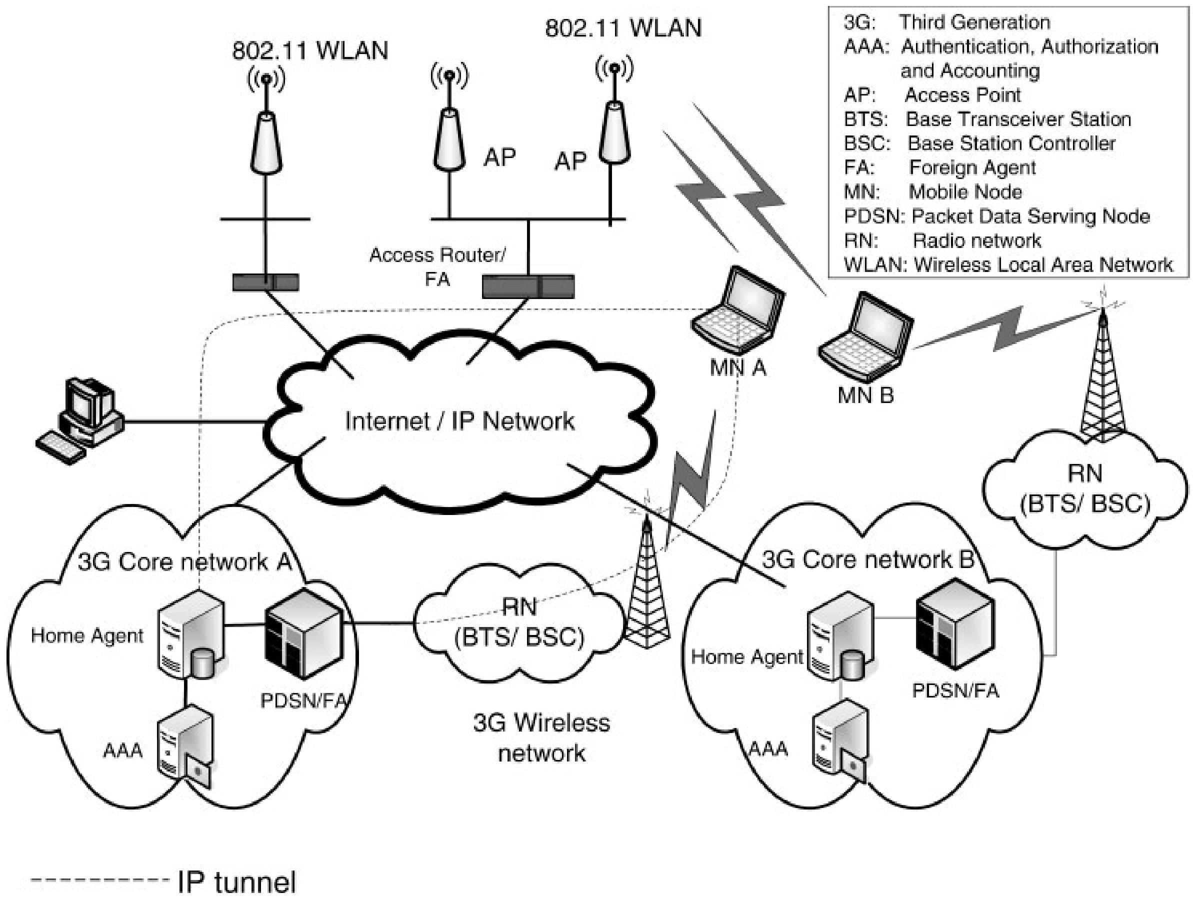


Fig. 1. Loosely coupled 802.11 wireless local area network (WLAN) and 3G cellular networks.

gateway GPRS supporting nodes (GGSNs), or 3G core network switches. Generally a loosely coupled network architecture, which is shown in Figure 1, is preferred [3] due to the flexible deployment and it is the network topology considered in the paper. In the figure, WLAN and 3G cellular networks work in a complementary way, and the interworking between WLAN and 3G cellular network is through an IP network. Mobile IP (MIP) protocol [4,5] is applied to support IP mobility for mobile nodes (MNs). The MNs are provided with MIP clients and can support both 802.11 and 3G cellular access technologies. Data can be transported between two different paths, shown as dashed lines in the figure, depending on the access interface that the MN is associated with. With this loosely coupled architecture, the same WLAN can be shared by different cellular service providers who own different 3G cellular networks. The main parts of a 3G cellular network owned by each service provider are radio access network (RAN) and core network. The network has two new functional entities, the home agent (HA) and the foreign agent (FA), specifically added to support MIP. The PDSN is modified to act as a FA for the cellular network in addition to its original intended functionality. 802.11 WLANs are connected through an IP network (or the Internet) to the 3G cellular core networks. Each WLAN has its own gateway which serves as a FA for MNs within its coverage. The authentication, authorization, and accounting (AAA) server is used for authentication and authorization of WLAN users as well as 3G cellular network users. Users registered with a cellular service provider is authenticated by the AAA/ RADIUS (remote authentication dial-in user service) server at the home network, whether accessing the network through the 3G cellular network section or the WLAN section.

In loosely coupled WLAN/cellular networks, seamless service continuity is desired. Seamless service continuity requires that data loss and service break time during the vertical handoff between 3G cellular network and 802.11 WLAN should be minimized.

Copyright © 2006 John Wiley & Sons, Ltd.

Multimedia or voice over IP (VoIP) sessions should be maintained without noticeable interruption, by means of a fast vertical handoff protocol.

Several algorithms based on the received signal strength or received power have been employed or investigated to make efficient and effective handoff decision [6–10]. The algorithms employ thresholds to compare the values of metrics from different points of attachment and then decide on when to make the handoff. Primarily, the received signal strength measurements from the serving point of attachment and neighboring points of attachment are used. Also, information derived from the path loss, carrier-to-interference ratio, signal-to-interference ratio, bit error rate, power budgets, and cell ranking has been used as metrics in certain mobile voice and data networks. In order to avoid a ping-pong effect, where the MN oscillates between the two networks due to variable radio signal strength, additional parameters, such as hysteretic margin, dwell times, and averaging window, are employed in the algorithms. However, the relative signal strength method [9] initiates too many unnecessary handovers. A timer-based algorithm [10] may not be able to efficiently utilize the higher bandwidth when it becomes available. When the MN enters the service area of a WLAN hotspot and requires higher transmission rate for the multimedia streaming applications, it cannot be satisfied using timer-based algorithms. To make more intelligent decisions, additional parameters need to be applied [11].

During vertical handoff between the 3G cellular network and WLAN, service may be disrupted due to the air interface switching. Packet delay and loss due to the service disruption need to be minimized, in order to maintain a satisfactory QoS for IP traffic. In general, the handoff procedure consists of two stages: Layer 2 (link layer or L2) handoff and Layer 3 (IP layer or L3) handoff. While L2 handoff can be fast (in the order of tens of milliseconds), L3 handoff usually takes longer time than that can be tolerated by real-time traffic. Therefore, fast handoff protocol is necessary to support real-time services such as VoIP [12]. In order to further reduce the packet loss, a fast hanoff scheme with bicasting is proposed in Reference [13]. A bicasting scheme sends duplicated packets to both air interfaces (cellular network and WLAN) during the interface switching period to minimize the packet loss. However, there is little literature on modeling the packet loss characteristic of bicasting scheme in WLAN and cellular network air interface switching.

In this paper, the network interface switching decision and registration issues, and the fast vertical handoff protocols supporting real-time traffic are investigated. The performance of a bicasting-based handoff scheme is analyzed and compared with that of the one without bicasting. Numerical results show that bicasting is an effective approach to reduce the packet loss and maintain satisfactory QoS for VoIP traffic. The rest of the paper is organized as follows. The registration process for WLAN and cellular network vertical handoff is presented in Section 2. Section 3 discusses fast MIP handoff schemes with and without bicasting. Performance analysis for the two fast handoff schemes is presented in Section 4. Section 5 concludes the paper.

2. WLAN and Cellular Network Registration for Vertical Handoff

The MN associates itself with the service network for exchanging identity information and security credential through the registration process. The registration procedure of the integrated cellular network and WLAN is AAA based because of its popularity and scalability. The identity of the MS is formatted as a network address identifier (NAI) and has the structure username@realm, where realm is an Internet domain name as specified in Reference [14]. AAA protocols [15,16] are protocols used to convey data for AAA purposes. There are local certificate authority (CA) servers in the network to authenticate the MN. Those local CA servers can be seen as sub CAs, which are associated by a higher level CA server. Figure 2 shows the considered network structure for the MN registration. The two FAs and the HA are connected to a CA. The CA in the figure may be comprised of several local CA servers with a higher level CA in

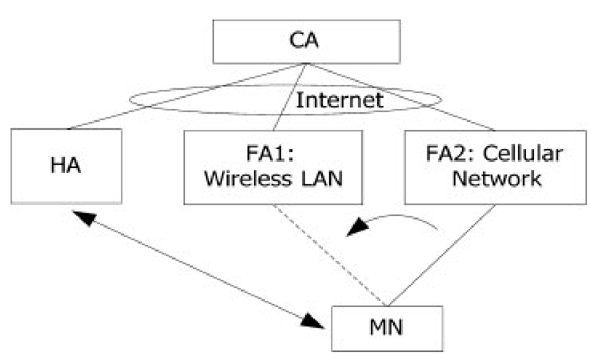


Fig. 2. Hierarchical view of network structure.

Copyright © 2006 John Wiley & Sons, Ltd.

order to handle high-network load and different administration realms. The MN associated with its HA is connected with WLAN hotspot associated with FA1. The networks can be verified by the certificates issued by the CA server upon request. The identity verification can be done by using the public-key encryption and digital signature algorithms. Since the CA servers is responsible for large amount of certificate issuing and verification, the MN never contacts the root CA server. It is also not practical for the CA server to record the certificate information of all MNs because of their enormous population. The certificate of each MN is stored in its home network. Thus, each home network server can be considered as a CA server of its MNs.

2.1. Multiple Network Interface Structure

In order to access both the WLAN and the cellular network, the MN should be equipped with the corresponding network access interfaces. Figure 3 shows a structure for MN with two network access interfaces using MIP. The data packets from the corresponding server are routed to the MN through the corresponding HA. When the MN roams to the foreign network, the two network cards are assigned temporary care of address (CoA) by the FA. The interface switching agent, which is a software-based engine switching the data source for the common network protocol stack, isolates the interface changes to the Internet applications and makes decision of when and how to switch the network interface.

The switching of the two interfaces can be considered as the change of CoA in MIP. The access points in the WLAN hotspots and the base station in the cellular network send the beacon signals periodically, so that the MN is able to sense that it is approaching a point of attachment. When the MN decides to switch the interface, it informs the HA by updating its current CoA to the IP address of the other network access card. The HA forwards the data to the new IP address. This will ensure that the process of network access card switching is dealt with using the switching process in MIP.

2.2. Registration Process for Vertical Handoff

The handoff process has two scenarios: an upward handoff, which is a network switching to a wireless overlay with a larger cell size and lower bandwidth density, that is, cellular network, and a downward handoff, which is a network switching to a wireless overlay with a smaller cell size and higher bandwidth density, that is, WLAN.

2.2.1. Upward handoff

Let the MN be connected to the WLAN. The typical transmission rate of WLAN is mostly greater than that of cellular network. Since the cellular network is lying on top of the WLAN hotspot, cellular network service is always available no matter if the MN is in the WLAN hotspot or not. So the MN should attach to the WLAN as long as possible. When the MN moves out of the service area of a WLAN hotspot, it has to make the upward switching to maintain the consistent network connection. The condition of upward switching is always that the WLAN signal

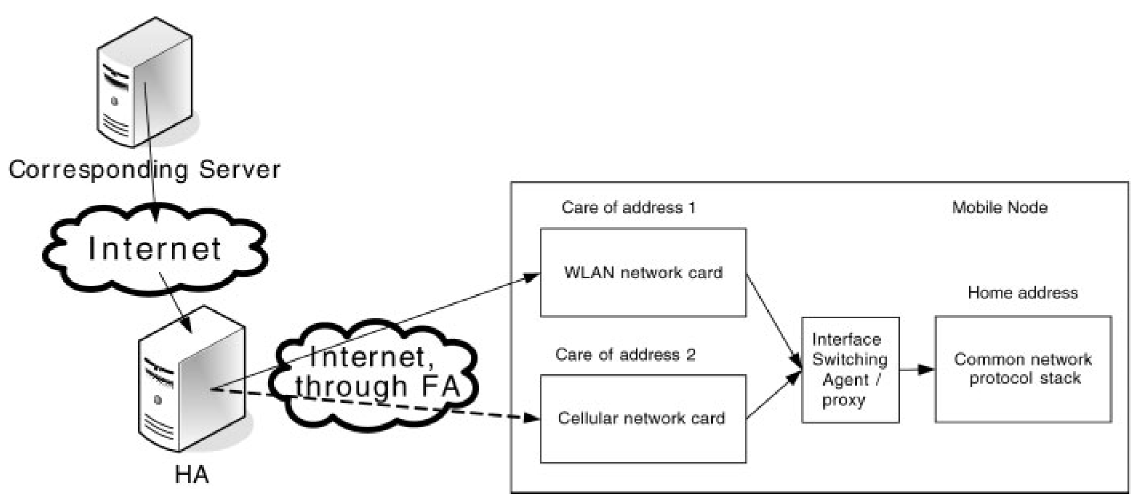


Fig. 3. Dual interface mobile node.

Copyright © 2006 John Wiley & Sons, Ltd.

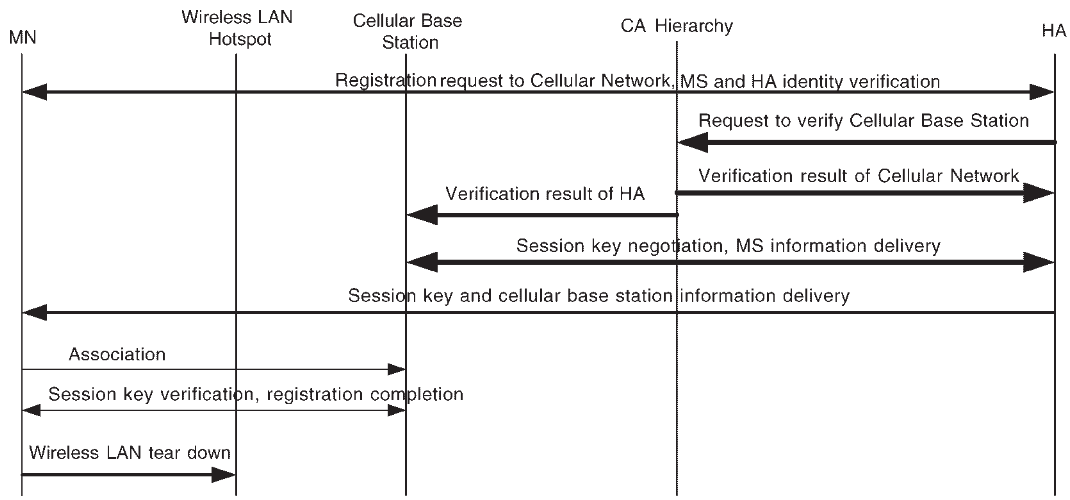


Fig. 4. Upward registration signaling.

strength drops below its lower threshold. When the upward switching occurs, the network registration process is involved. Registration process modification may help to reduce the period in this scenario because the MN is switching to a lower rate network from a higher one and the registration process is traditionally processed in the network that the terminal is switching to. It is preferred that the registration process can be mostly done in the higher rate network.

The registration procedure, as shown in Figure 4, can be divided into two steps: pre-switching and post-switching. In pre-switching registration, the MN receives the beacon signals from the cellular base stations and identifies the cellular network it will register. The MN then sends the registration request, which includes the identity of the cellular base station to the corresponding HA. The HA verifies the cellular base station via the CA server. The CA server also sends the verification result of the HA to the base station. If the base station can admit the MN according to its resource allocation policy, the base station and the HA will start to negotiate a session key. Then the HA sends the identity of the MN to the base station, and the session key to the MN. When the pre-switching phase completes, the post-switching phase begins. The MN associates itself with the base station, and both of them perform the session-key verification and finish the registration process. The MN terminates the session with the WLAN hotspot. The traffic between the HA and the MN is encrypted by the pre-set secret key shared by the MN and the corresponding HA to assure privacy and integrity. The information exchanges between the fixed sites, such as WLAN hotspot, cellular network base station, CA and HA, are tunneled. In the registration process, the HA registers the MN to the cellular network on behalf of the MN. The benefits of the proposed registration procedure are as follows:

- Most of the registration procedure is completed when the MN is still connected to the higher transmission rate network.
- On the MN side, only local symmetric encryption/decryption is involved so that the requirement of computation power and wireless channel bandwidth is low.
- The registration infrastructure is suitable for the integration of heterogeneous networks.

2.2.2. Downward handoff

Let the MN be connected to the cellular network. When it moves into a WLAN hotspot service area, unlike the previous case, the MN can have two choices: stay with the current base station or switch to the hotspot. The interface switching decision can be made either remotely or locally. However, when interface switching decision is to be made, there are one or more of the following factors involved, such as signal quality, the priority of each interface, network throughput, network load change caused by the switching, mobility of the MN, and user experience, etc. If the switching decision is made by the HA, some of those data need be transmitted to the HA so that it is able to monitor the status of the MN. This may add the traffic load to the already tight wireless bandwidth

Copyright © 2006 John Wiley & Sons, Ltd.

resources, which is not desirable. Therefore, although the HA has much more computational power than the MN, a local decision making mechanism is preferred.

The goal of the downward interface switching scheme is to introduce a hysteretic effect to make better use of the wireless bandwidth and reduce power consumption. Using the mobility as the primary parameter to make the decision is because the WLAN hotspot is relatively small and non-adjacent. For instance, if the MN is moving fast, it should not switch to a WLAN hotspot when it passes the hotspot. Although a timer can be used, by which the MN switches to WLAN after a fixed time if it is still within WLAN area, it may affect the overall throughput. The typical mobility pattern suitable for the WLAN hotspot is high-speed network connection for the MN at very low to zero velocity. Since mobility is a realtime available parameter and does not require accumulation computation, using mobility related information can overcome the shortcoming of timer based algorithm. The scheme also tries to avoid making the switching during the ongoing session if it is unnecessary and not urgent, such as when the MN receives low bandwidth contents.

The proposed scheme can be executed as follows. The WLAN network card checks the WLAN hotspot beacon signal periodically. When the beacon signal is detected, the fading characteristics of the signal, such as signal fluctuation and strength increase rate, etc., are examined. If fast fading is the major component of the fluctuation or signal strength increase rate is high, it indicates that the MN is moving at relatively high speed. The mobile switch agent predicts that the MN will pass the WLAN quickly. Since a typical WLAN hotspot usually covers an area with a diameter at the around 200 m, it is not suitable for the MN to switch to the WLAN. If the majority of the signal variation is found to be slow, the MN checks the quality of the signal using metrics such as SNR, etc. If the signal is good, the MN checks the current network traffic on the cellular network interface card. If the short term average traffic rate does not saturate the cellular network interface, it means that the Internet application is currently working well with the connected cellular network and switching is not urgent. The MN schedules the interface switching when the cellular network card is in an idle state. Otherwise, the MN will engage the interface switching immediately to solve the current network traffic bottleneck. Figure 5 shows the flow chart of the downward network interface switching scheme.

Similar to the upward registration process, Figure 6 illustrates the corresponding registration process. In the figure, the MN associates with the WLAN hotspot first to utilize its higher bandwidth and sends the registration traffic through the hotspot. Since the

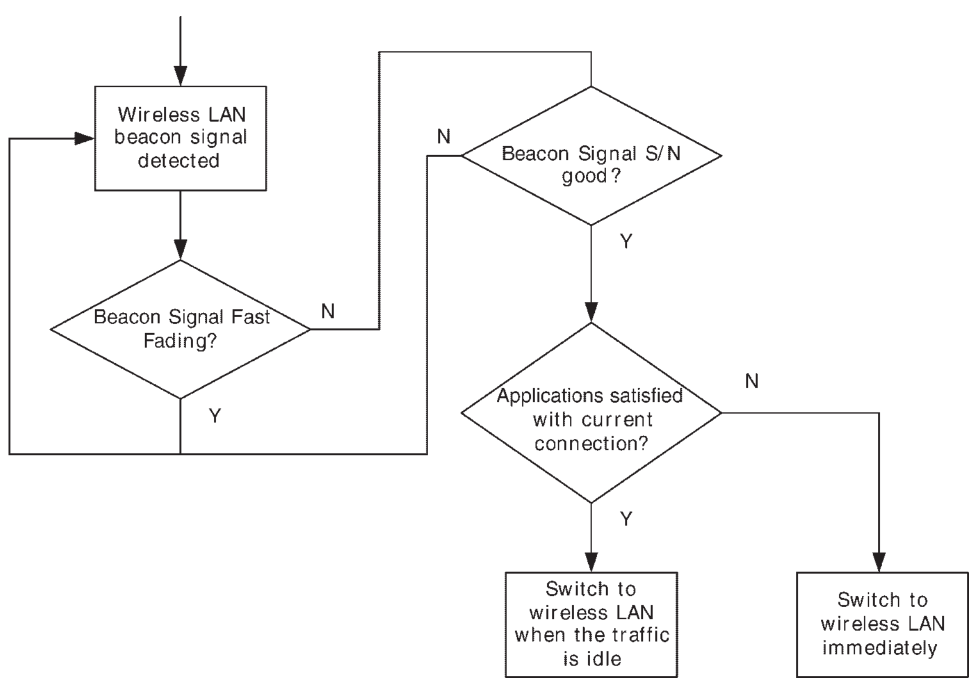


Fig. 5. Downward network interface switching scheme.

Copyright © 2006 John Wiley & Sons, Ltd.

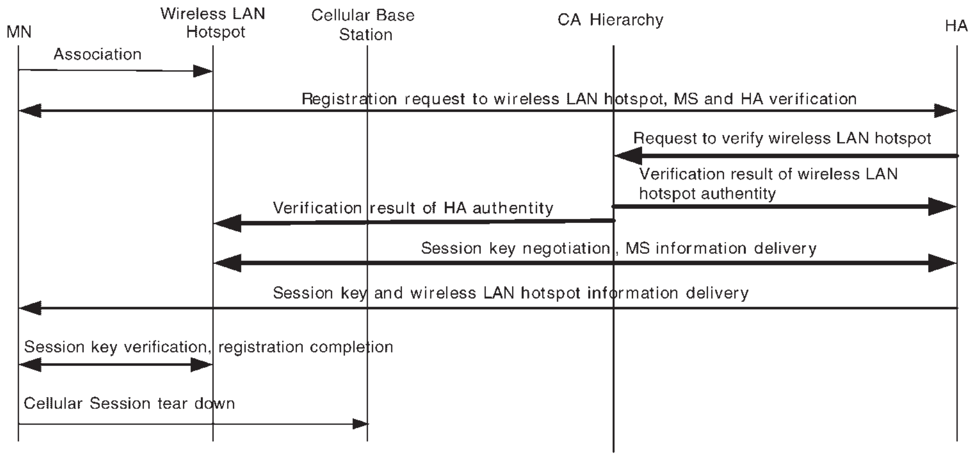


Fig. 6. Downward registration signaling.

outbound data traffic is required before the MN is validated, the security policy of WLAN hotspot should be properly set to accommodate this modification.

3. Fast Vertical Handoff With and Without Bicasting

In general, the handoff procedure consists of two levels: L2 handoff and L3 handoff. The L2 handoff can be triggered by the handoff decision engine based on statistics of received signal strength/status (RSS) measured at the MN and interface switching policy discussed in the previous section, and L3 handoff can be triggered by the L2 handoff. After L3 handoff, the CoA of the MN is changed, and the MN is detached from the previous access router (PAR), and associated with a new access router (NAR). While L2 handoff can be fast (in the order of tens of milliseconds), L3 handoff usually takes longer time than that can be tolerated by real-time traffic. Therefore, a fast handoff protocol is necessary to support real-time services such as VoIP. In order to satisfy the QoS requirement in terms of maximum end-to-end delay, we assume that the packets for real-time services are not buffered at the routers during the handoff.

3.1. A Fast Handoff Protocol Based on Mobile IP

The protocol is designed to minimize the amount of service disruption when performing L3 handoff. This mechanism involves the use of L2 triggers which allow the L3 handoff to be anticipated rather than traditionally to be performed after the L2 handoff completion. Fast handoff is required to ensure that the L3 (MIP) handoff delay is minimized, thus also minimizing and possibly eliminating the period of service disruption which normally occurs when an MN moves between two access routers (ARs). The following is a short summary of the fast handoff mechanism described in Reference [12].

While the MN is connected to its PAR and is about to move to an NAR, fast handoffs in Mobile IPv6 requires:

- the MN to obtain a new care-of address at the NAR while connected to the PAR;
- the MN to send a binding update (BU) to its old anchor point (e.g., PAR) to update its binding cache with the MN new care-of address;
- the old anchor point (e.g., PAR) to start forwarding packets destined for the MN to the NAR.

The MN or the PAR may initiate the fast handoff procedure by using wireless link-layer information or link-layer triggers, which inform that the MN will soon be handed off between two wireless access points, respectively, attached to the PAR and the NAR. Such link-layer information can be obtained by measuring the RSS of the corresponding access point (AP) of 802.11 WLAN. Upon receipt of the link-layer trigger, the MN will initiate the L3 handoff process by sending a proxy router solicitation (RtSolPr) message to the PAR. The RtSolPr message contains the information of the NAR that the MN wants to attach to. In response to this solicitation message, The PAR sends back a proxy router advertisement (PrRtAdv) message containing information

Copyright © 2006 John Wiley & Sons, Ltd.

from the NAR, based on which the MN may obtain a new CoA. The MN then updates the PAR with its new CoA using a fast-binding update (F-BU) message. The PAR will validate the MN new CoA by sending a handoff initiate (HI) message to the NAR. Based on the response generated in the handoff acknowledge (HAck) message, the PAR will either generate a tunnel to the MN new CoA (if the address was valid) or generate a tunnel to the NAR address (if the address was already in use on the new subnet). After the tunnel is generated, packets that arrive at the PAR will be forwarded to the NAR.

3.2. Fast Handoff With Bicasting

The fast-handoff protocol described above allows the anticipation of the L3 handoff such that data traffic can be redirected to the MN new location before it moves there. However, it is not simple to determine the correct time to start forwarding between the PAR and the NAR, which has an impact on how smooth the handoff will be. Packet loss will occur if the handoff is performed too late or too early with respect to the time in which the MN detaches from the PAR and attaches to the NAR. Moreover, when the MN moves quickly back-and-forth between two ARs (i.e., ping-pong effect), it is difficult for the L3 handoff protocol to follow.

To address these issues, one solution is to bicast packets destined to the MN for a short period from the old anchor point (e.g., PAR) to one or more potential future MN locations (e.g., NARs) before the MN actually moves there. This means that the handoff procedure described previously should be enhanced by having the PAR send one copy of packets to the MN current CoA and another copy of the packets to the MN new CoA connected to the NAR. This is called bicasting based 'soft handoff,' in contrast to hard handoff, which maintains only one traffic flow for each user during the handoff process. Traffic for the MN is bicast for a short period to its current location and to one or more locations where the MN is moving to. Since the duplicated packets are sent from the PAR and received by the two air interfaces, respectively, and the traffic between the PAR and the NAR is transferred via a wired link, the bicasting scheme will not over occupy the bandwidth of any wireless channel interface at either the sender or the receiver. The mechanism removes the timing ambiguity regarding when to start sending traffic for the MN to its new point of attachment following a fast handoff, which also allows the decoupling of L2 and L3 handoff and removes the periods of service disruption in the case of MN ping-pong movement.

The bicasting-based soft vertical handoff can assure a satisfactory QoS in terms of packet loss and smooth L3 handoff [13]. However, bicasting involves two communication links and IP addresses assigned to one user, which may cause duplication of received packets and excess total wireless bandwidth allocation. In order to minimize the redundant bandwidth and the number of duplicated packets, the timing for bicasting should be carefully designed. To achieve zero service disruption, it is necessary for the time period between starting the IP handoff procedure and the MN completing the L2 handoff to be greater than or equal to the time it takes for traffic to reach the MN at its new link (through NAR). In other words, the following inequality should be satisfied:

$$A + h \geq D_1 + D_2 + P$$

where A is the anticipation time between the time when the MN initiates the L3 handoff and when it starts L2 handoff, h is the L2 handoff time, D_1 is MN to HA delay (through PAR), D_2 is MN to the new AP delay (through NAR), and p is routing table processing time in the HA and MN. In practice, we will reduce A to minimize the bicasting time, as long as service is not disrupted. However, since D_1 and D_2 are random variables, it is difficult to find an optimal A to guarantee seamless handoff with minimal excessive resource. In the following, an analytical model is presented to evaluate the performance of the fast handoff protocols, and the impact of A.

4. Performance Analysis

In this section, we analyze and compare the performance of fast handoff with and without bicasting. Consider a simplified network model for cellular/WLAN interworking shown in Figure 7. During the vertical handoff, real-time VoIP traffic are forwarded by the HA to the PDSN. With bicasting, the PDSN will forward packets to both 3G base station (BS) and the AR of WLAN, after the binding update is completed. For handoff without bicasting, PDSN will only forward the received packets to the AR after binding update. As mentioned before, we assume packets are not buffered at the BS (or the AR) for real-time traffic if no wireless link is established between the MN and the BS (or the AR). In the following, we analyze packet loss of VoIP traffic

Copyright © 2006 John Wiley & Sons, Ltd.

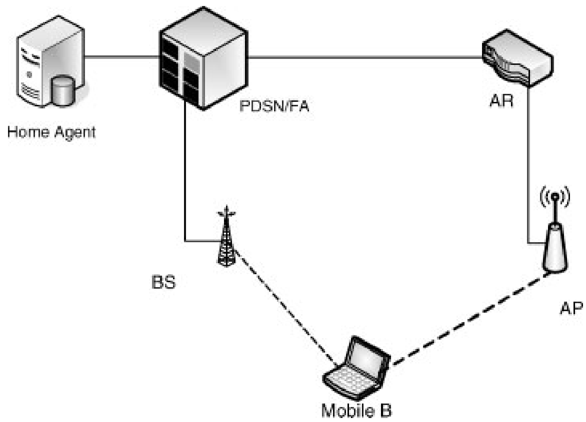


Fig. 7. Simplified network model for vertical handoff.

during vertical handoff from the cellular network to the WLAN.

Vertical handoff is initiated by the MN when it enters a WLAN hotspot and decides to switch to the WLAN. After obtaining a new CoA through the PAR (i.e., the PDSN it attached to), the MN initiates a L3 handoff procedure by sending out the F-BU message. Let t_0 be the time that the MN sends out F-BU, t_1 be the time when the PDSN completes binding update and starts to forward packets to the new CoA. The MN will start the L2 handoff at time $t_2 = t_0 + A$. The MN will attach to the AR and the AP of WLAN at time $t_3 = t_0 + A + h$. t_1 is a random variable that depends on the processing time for HI and HAck messages and the link delay between the PDSN and the AR, that is,

$$t_1 = t_0 + D_\mathrm{p} + D_\mathrm{AR} + D'_\mathrm{p} + L_1 \tag{1}$$

where D_P is the delay due to the processing of F-BU message at the PDSN, D_AR is the delay due to the processing of HI message at the AR, D'_P is the delay due to the processing of HAck message at the PDSN, L_1 is the link delay due to transmission of the L3 messages over wireless and wired links. Poisson and exponential processes are reasonable models for describing the incoming and service time of VoIP packets, respectively. In the following analysis, we model the PDSN and the AR as M/M/1 queues. Therefore, D_P and D_AR can be considered as exponentially distributed random variables. The link delay L_1 is assumed constant.

Consider a VoIP packet flow arriving at the PDSN and then being forwarded to the MN either through the cellular system or through the WLAN, during the vertical handoff. We assume that the packets of the VoIP flow arrive at the PDSN at a constant rate and the interarrival time is T_v. Let t_k be the time when the kth packet arrives at the PDSN. For the fast-handoff protocol without bicasting, we have the following cases:

Case 1: $t_k < t_1$, the kth packet arrives before the PDSN completes binding update, and is routed to the MN through the BS. The following two events may occur:

- Event 1.a: the kth packet is forwarded to the MN by time t_2.
- Event 1.b: the kth packet is lost because it arrives at the BS after t_2.

Case 2: $t_k > t_1$, the kth packet arrives after the PDSN completes binding update, and is routed to the MN through the WLAN AR. The following two events may occur:

- Event 2.a: the kth packet is forwarded to the MN after t_3.
- Event 2.b: the kth packet is lost because it arrives at the WLAN access point before t_3.

The probability that the kth packet could be lost is

$$\mathrm{Prob}\{\mathrm{loss}, t_k\} = \mathrm{Prob}\{1.b\} + \mathrm{Prob}\{2.b\} \tag{2}$$

We have

$$\mathrm{Prob}\{1.b\} = \mathrm{Prob}\{t_k < t_1\}\mathrm{Prob}\{t_k + D_\mathrm{p} + L_2 > t_2\} \tag{3}$$

where L_2 is the total link delay due to the transmission of the packet over the wired link between the PDSN and the BS, and over the wireless link between the BS and the MN. L_2 is assumed to be a constant in the analysis. And

$$\mathrm{Prob}\{2.b\} = \mathrm{Prob}\{t_k > t_1\}\mathrm{Prob}\{t_k + D_\mathrm{p} + D_\mathrm{AR} + L_3 < t_3\} \tag{4}$$

where L_3 is the total link delay due to the transmission of the packet over the wired link between the PDSN and the BS and over the wireless link between the BS and the MN. L_3 is assumed to be a constant in the analysis.

Let T_P be the average queuing time at the PDSN, and T_A the average queuing time at the AR. Let $Y = D_\mathrm{p} + D_\mathrm{AR} + D'_\mathrm{p}$. Since D_p, D_AR, and D'_p are three

Copyright © 2006 John Wiley & Sons, Ltd.

independent exponential random variables with the following probability density functions (*pdf*):

$$f_{D_p}(x) = f_{D'_p}(x) = \frac{1}{T_P} e^{-x/T_P}, x \geq 0$$

$$f_{D_{AR}}(x) = \frac{1}{T_A} e^{-x/T_A}, x \geq 0$$

We can obtain the *pdf* of Y as follows:

$$f_Y(y) = \frac{1}{(T_P - T_A)T_A} y e^{-x/T_A} - \frac{T_A}{(T_P - T_A)^2} (e^{-y/T_P} - e^{-y/T_A}), y \geq 0$$

Without loss of generality, let $t_0 = 0$. Then

$$\text{Prob}\{t_k < t_1\} = \text{Prob}\{y > t_k - L_1\} = 1 - \int_0^{t_k - L_1} f_Y(y)\text{d}y$$

$$\text{Prob}\{t_k + D_p + L_2 > t_2\} = \text{Prob}\{D_p > t_2 - L_2 - t_k\} = \int_{t_2 - L_2 - t_k}^{\infty} f_{D_P}(x)\text{d}x$$

By integrating over the *pdf* of Y and D_p, we have

$$\text{Prob}\{1.b\} = \left\{ \frac{y_k}{T_P - T_A} e^{-y_k/T_P} - \frac{2T_P T_A - T_P^2}{(T_P - T_A)^2} e^{y_k/T_P} + \frac{T_A^2}{(T_P - T_A)^2} e^{y_k/T_A} \right\} e^{-\frac{t_2 - L_2 - t_k}{T_P}} \quad (5)$$

where $y_k = t_k - L_1$.

$$\text{Prob}\{2.b\} = \left[1 - \frac{y_k}{T_P - T_A} e^{-y_k/T_P} + \frac{2T_P T_A - T_P^2}{(T_P - T_A)^2} e^{y_k/T_P} - \frac{T_A^2}{(T_P - T_A)^2} e^{y_k/T_A} \right] \cdot \left[1 - \frac{T_A}{T_A - T_P} e^{-\frac{t_3 - L_3 - t_k}{T_A}} + \frac{T_P}{T_A - T_P} e^{-\frac{t_3 - L_3 - t_k}{T_P}} \right] \quad (6)$$

For the fast handoff protocol with bicasting, we have the following cases:

Case B1: $t_k < t_1$, the kth packet arrives before the PDSN completes binding update, and is routed to the MN through the BS. The following two events may occur:

- Event B1.a: the kth packet is forwarded to the MN by time t_2.
- Event B1.b: the kth packet is lost because it arrives at the BS after t_2.

Case B2: $t_k > t_1$, the kth packet arrives after the PDSN completes binding update, and is bicast to the MN through both the cellular network and the WLAN. The following three events may occur:

- Event B2.a: the kth packet is forwarded to the MN via the BS by time t_2.
- Event B2.b: the kth packet is lost because one copy of the packet arrives at the BS after t_2 and the other copy arrives at the WLAN AP before t_3.
- Event B2.c: the kth packet is forwarded to the MN via the AR and the AP after t_3.

The probability that the kth packet could be lost during the fast handoff with bicasting is

$$\text{Prob}\{\text{loss}, t_k\} = \text{Prob}\{B1.b\} + \text{Prob}\{B2.b\} \quad (7)$$

Since Event B1.b is equivalent to Event 1.b, we have $\text{Prob}\{B1.b\} = \text{Prob}\{1.b\}$. Observing the relationship between Event B2.b and Event 2.b, we have

$$\begin{aligned}\text{Prob}\{B2.b\} &= \text{Prob}\{t_k > t_1\}\text{Prob}\{t_k + D_p + D_{AR} + L_3 < t_3\}\text{Prob}\{t_k + D_p + L_2 > t_2\} \\ &= \text{Prob}\{2.b\}\text{Prob}\{t_k + D_p + L_2 > t_2\} \\ &= \text{Prob}\{2.b\} e^{-\frac{t_2 - L_2 - t_k}{T_P}}\end{aligned}$$

Figure 8 shows examples of loss probabilities calculated from Equations (5)–(7) and the corresponding simulation results. The parameters used in the numerical example are given in Table I. In the example, we investigate the loss probabilities as a function of the anticipation time A, of four packets arriving at instants $t = 50$ ms, $t = 80$ ms, $t = 100$ ms, and $t = 130$ ms, respectively. It can be observed that, for packets arriving at different instants, the probability of loss due to vertical handoff can be different.

Copyright © 2006 John Wiley & Sons, Ltd.

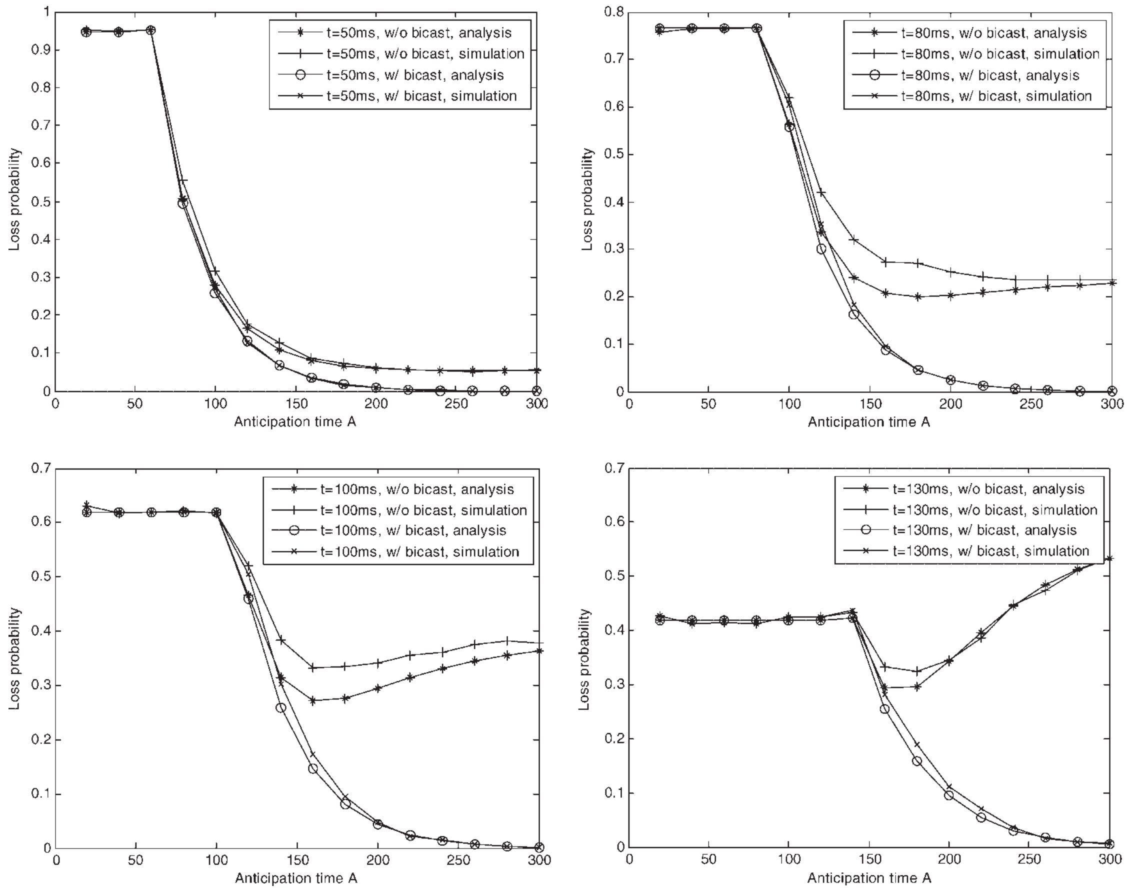


Fig. 8. Loss probability for packets arriving at different times versus A (ms) for time instance at 50, 80, 100, and 130 ms, respectively.

Table I. Parameters for the numerical example.

Parameter	Value (ms)
t_0	0
T_P	30
T_A	50
L_1	20
L_2	10
L_3	15
h	10

The parameter A of fast handoff protocols has an impact on the packet loss probability, and the impact can be different for packets arriving at different time. In general, when A is small, a packet arriving at an earlier time has a higher loss probability, and when A is large, the loss probability of this packet can be higher than that of a packet arriving later. It can also be seen that loss probabilities in handoff with bicasting can be lower than those in handoff without bicasting. When A increases (i.e., the bicasting time also increases correspondingly), loss probabilities can decrease towards zero in fast handoff with bicasting. For handoff without bicasting, however, there exists a floor for the loss probability. One could observe that the packet-loss probability without bicasting increases after the anticipation time exceeds a critical point of time. The reason is that when the anticipation time is too large, packets arriving after the L3 handoff are more likely to be forwarded to the WLAN AR before the L2 handoff is completed, which results in increased packet-loss probability.

Figure 9 shows an example of expected number of lost packets for a VoIP packet flow in vertical handoff versus anticipation time. In this example, we assume the inter-arrival time of packets is $T_v = 5$ ms. The expected number of lost packets can be estimated as follows:

$$N_{\text{loss}} = \sum_{k=0}^{N} \text{Prob}\{\text{loss}, t_k\} \tag{8}$$

Copyright © 2006 John Wiley & Sons, Ltd.

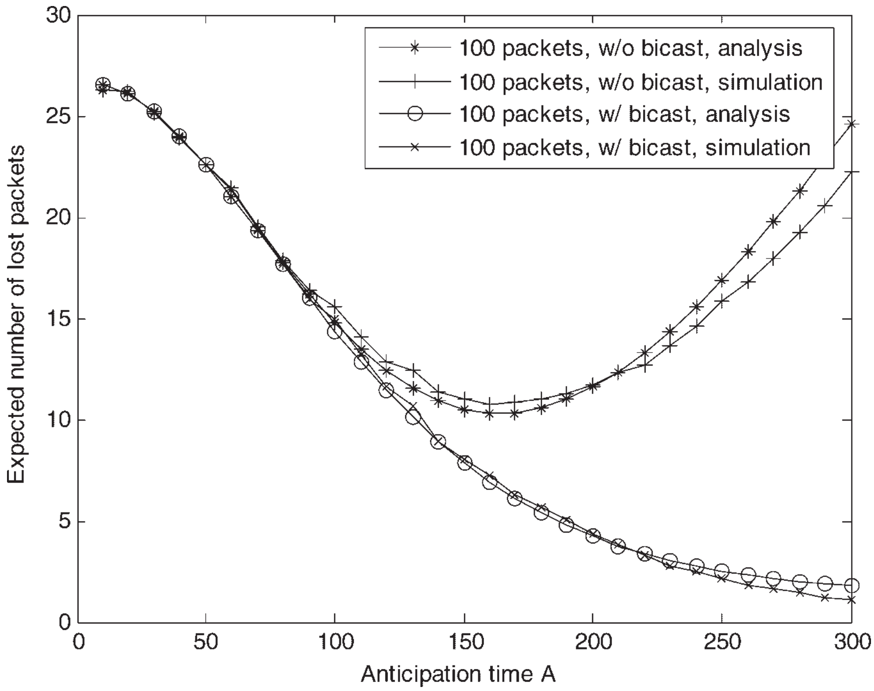


Fig. 9. Expected number of lost packets versus Anticipation time (ms).

where $t_k = t_0 + kT_v$, N is a large number. The handoff protocol with bicasting achieves lower packet loss than the protocol without bicasting. When the anticipation time A increases, the packet loss can be eliminated with bicasting. Without bicasting, there is an optimal A that corresponds to a minimal packet-loss rate, and there is a floor to the expected number of lost packets.

5. Summary

The mobile node registration and faster network interface switching for vertical handoff in integrated 3G cellular and WLAN networks have been investigated. An analytical model has been presented to examine the effect of fast handoff with and without bicasting on packet-loss rate of real-time traffic. Numerical results show that the proposed fast handoff protocol with bicasting can achieve lower packet loss rate without increasing bandwidth on both wireless interfaces.

References

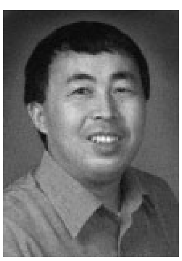
1. Loredo CS, deGrimaldo SW. Wireless LANs: global trends in the workplace and public domain. *The Strategies Group*, Feb. 2002.
2. Axiotis DI, Al-Gizawi T, Peppas K *et al*. Services in interworking 3G and WLAN environments. *IEEE Wireless Communications* 2004; **11**: 14–20.
3. Buddhikot M, Chandranmenon G, Han S *et al*. Integration of 802.11 and third-generation wireless data networks. *Proceedings of IEEE Infocom 2003* 2003; **1**: 503–512.
4. Perkins C. IP mobility support for IPv4. *IETF RFC 3220*, Jan. 2002.
5. Johnson D, Perkins C, Arkko J. Mobility support in IPv6. *IETF RFC 3775*, 2004.
6. Li J, Pan Y, Jia X. Analysis of dynamic location management for PCS networks. *IEEE Transactions on Vehicle Technology* 2002; **51**: 1109–1119.
7. Pollini GP. Trends in handover design. *IEEE Communications Magazine* 1996; **34**(3): 82–90.
8. Caceres R, Iftode L. Improving the performance of reliable transport protocols in mobile computing environments. *IEEE Journal on Selected Areas in Communications* 1995; **13**(5): 850–857.
9. Gudmundson M. Analysis of handover algorithms. *Proceedings of Vehicular Technology Conference '91* May 1991; 537–542.
10. Chuah C-N, Yates RD. Evaluation of a minimum power handoff algorithm. *Proceedings of IEEE International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC) '95*, September 1995; 814–818.
11. Pahlavan K, Krishnamurthy P, Hatami A *et al*. Handoff in hybrid mobile data networks. *IEEE Personal Communications Magazine* 2000; **7**: 34–47.
12. Koodli R, *et al*. Fast handoffs for mobile IPv6. *draft-ietf-mobileip-fast-mipv6–07.txt, IETF Draft*, September 2003.
13. Malki KE, Soliman H. Simultaneous bindings for mobile IPv6 fast handoffs. *draft-elmalki-mobileip-bicasting-v6-03, IETF Draft*, October 2003.
14. Mockapetris P. Domain names—implementation and specification. *IETF RFC 1035*, November 1987.
15. Rigney C, *et al*. Remote authentication dial in user service (RADIUS). *RFC 2865*, June 2000.
16. Aboda B, Calhoun P. RADIUS (remote authentication dial in user service) support for extensible authentication protocol (EAP). *RFC3579*, September 2003.

Copyright © 2006 John Wiley & Sons, Ltd.

Authors' Biographies

Minghui Shi received his B.S. degree in 1996 from Shanghai Jiao Tong University, China, and his M.S. degree in 2002 from the University of Waterloo, Ontario, Canada, both in electrical engineering. He is currently working towards his Ph.D. degree at the University of Waterloo. His current research interests include wireless LAN/cellular network integration and network security.

Liang Xu received his B.Eng. degree in electronic engineering from Tsinghua University, China, in 1994, M.Eng. degree in communications and electronic systems from Southeast University, China, in 1997, his Ph.D. in electrical and computer engineering from the University of Waterloo, in 2003. He is currently with UTStarcom, Inc. His research interests include radio resource management, media access control, WLAN/cellular interworking, and 3G/4G wireless networks.



Xuemin (Sherman) Shen (SM) received his B.Sc. (1982) degree from Dalian Maritime University (China) and his M.Sc. degree (1987) and Ph.D. (1990) from Rutgers University, New Jersey (U.S.A.), all in electrical engineering. From September 1990 to September 1993, he was first with the Howard University, Washington D.C., and then the University of Alberta, Edmonton (Canada). Since October 1993, he has been with the Department of Electrical and Computer Engineering, University of Waterloo, Canada, where he is a professor and the associate chair for graduate studies. Dr. Shen's research focuses on mobility and resource management in interconnected wireless/wireline networks, UWB wireless communications systems, wireless security, and *ad hoc* and sensor networks. He is a coauthor of two books, and has published more than 200 papers and book chapters in wireless communications and networks, control, and filtering. Dr. Shen was the technical program co-chair for *IEEE Globecom'03 Symposium on Next Generation Networks and Internet, ISPAN'04, IEEE Broadnet'05, QShine'05*, and is the special track chair of *2005 IFIP Networking Conference*. He serves as the associate editor for *IEEE Transactions on Wireless Communications; IEEE Transactions on Vehicular Technology; ACM/Wireless Networks; Computer Networks; Wireless Communications and Mobile Computing* (Wiley); and *International Journal of Computers and Applications*. He also serves as guest editor for *IEEE JSAC, IEEE Transactions Vehicular Technology, IEEE Wireless Communications*, and *IEEE Communications Magazine*. Dr. Shen received the Premier's Research Excellence Award (PREA) from the Province of Ontario, Canada for demonstrated excellence of scientific and academic contributions in 2003, and the Distinguished Performance Award from the Faculty of Engineering, University of Waterloo, for outstanding contribution in teaching, scholarship, and service in 2002. Dr. Shen is a registered Professional Engineer of Ontario, Canada.

Jon W. Mark (LF) received his Ph.D. in electrical engineering from McMaster University in 1970. In September 1970 he joined the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario, where he is currently a distinguished professor emeritus. He served as the department chairman during the period July 1984–June 1990. In 1996 he established the Centre for Wireless Communications (CWC) at the University of Waterloo and is currently serving as its founding director. Dr. Mark had been on sabbatical leave at the following places: IBM Thomas J. Watson Research Center, Yorktown Heights, NY, as a visiting research scientist (1976–77); AT&T Bell Laboratories, Murray Hill, NJ, as a resident consultant (1982–83): Laboratoire MASI, Université Pierre et Marie Curie, Paris France, as an invited professor (1990–91); and Department of Electrical Engineering, National University of Singapore, as a visiting professor (1994–95). He has previously worked in the areas of adaptive equalization, image and video coding, spread spectrum communications, computer communication networks, ATM switch design, and traffic management. His current research interests are in broadband wireless communications, resource and mobility management, and cross-domain interworking. He is a co-author of the text *Wireless Communications and Networking*, Prentice-Hall 2003. A life fellow of IEEE, Dr. Mark is the recipient of the 2000 Canadian Award for Telecommunications Research and the 2000 Award of Merit of the Education Foundation of the Federation of Chinese Canadian Professionals. He was an editor of *IEEE Transactions on Communications* (1983–1990), a member of the Inter-Society Steering Committee of the IEEE/ACM Transactions on Networking (1992–2003), a member of the IEEE Communications Society Awards Committee (1995–1998), an editor of *Wireless Networks* (1993–2004), and an associate editor of *Telecommunication Systems* (1994–2004).

Aladdin Saleh earned his Ph.D. in electrical engineering from London University, England. Since March 1998, Dr. Saleh has been working in the Wireless Technology Department of Bell Canada, the largest service provider of wireless, wire-line, and Internet in Canada. He worked as a senior application architect in the wireless data group working on several projects among them the wireless application protocol (WAP) and the location-based services. Later, he led the work on several key projects in the broadband wireless access planning group including planning of the IEEE 802.16/ Wimax, the IEEE 802.11/ WiFi, and the integration of these technologies with the 3G cellular network.

Copyright © 2006 John Wiley & Sons, Ltd.

Dr. Saleh also holds the position of adjunct full professor at the Department of Electrical and Computer Engineering of Waterloo University, Canada since January 2004. He is currently conducting several joint research projects with the University of Waterloo and the University of Toronto on IEEE 802.16-Wimax, MIMO technology, interworking of IEEE 802.11 WLAN and 3G cellular networks, and next generation wireless networks. Prior to joining Bell Canada, Dr. Saleh worked as a faculty member at different universities and was Dean and Chairman of Department for several years. Dr. Saleh is a fellow of IEE and a senior member of IEEE.

Copyright © 2006 John Wiley & Sons, Ltd.