



THE ATRIUM, SOUTHERN GATE, CHICHESTER, WEST SUSSEX P019 8SQ

**\*\*\*IMMEDIATE RESPONSE REQUIRED\*\*\***

Your article may be published online via Wiley's EarlyView® service (<http://www.interscience.wiley.com/>) shortly after receipt of corrections. EarlyView® is Wiley's online publication of individual articles in full-text HTML and/or pdf format before release of the compiled print issue of the journal. Articles posted online in EarlyView® are peer-reviewed, copy-edited, author-corrected, and fully citable via the article DOI (for further information, visit [www.doi.org](http://www.doi.org)). EarlyView® means you benefit from the best of two worlds - fast online availability as well as traditional, issue-based archiving.

Please follow these instructions to avoid delay of publication

**READ PROOFS CAREFULLY**

- This will be your only chance to review these proofs. **Please note that once your corrected article is posted online, it is considered legally published, and cannot be removed from the Web site for further corrections.**
- Please note that the volume and page numbers shown on the proofs are for position only.

**ANSWER ALL QUERIES ON PROOFS** (Queries for you to answer are attached as the last page of your proof.)

- List all corrections and send back via e-mail to the production contact as detailed in the covering e-mail, or mark all corrections directly on the proofs and send the scanned copy via e-mail. Please do not send corrections by fax or in the post.

**CHECK FIGURES AND TABLES CAREFULLY**

- Check size, numbering, and orientation of figures.
- All images in the PDF are downsampled (reduced to lower resolution and file size) to facilitate Internet delivery. These images will appear at higher resolution and sharpness in the printed article.
- Review figure legends to ensure that they are complete.
- Check all tables. Review layout, title, and footnotes.

**COMPLETE CTA (if you have not already signed one)**

- Please send a scanned copy with your proofs and post your completed original form to the address detailed in the covering e-mail. **We cannot publish your paper until we receive the original signed form.**

**OFFPRINTS**

- 25 complimentary offprints of your article will be dispatched on publication. Please ensure that the correspondence address on your proofs is correct for despatch of the offprints. If your delivery address has changed, please inform the production contact for the journal - details in the covering e-mail. Please allow six weeks for delivery.

**Additional reprint and journal issue purchases**

- Additional **paper reprints** (minimum quantity 100 copies) are available on publication to contributors. Quotations may be requested from [mailto:author\\_reprints@wiley.co.uk](mailto:author_reprints@wiley.co.uk). Orders for additional paper reprints may be placed in advance in order to ensure that they are fulfilled in a timely manner on publication of the article in question. Please note that offprints and reprints will be dispatched under separate cover.
- PDF files of individual articles may be purchased for personal use for \$25 via Wiley's Pay-Per-View service (see <http://www3.interscience.wiley.com/aboutus/ppv-articleselect.html>).
- Please note that regardless of the form in which they are acquired, reprints should not be resold, nor further disseminated in electronic or print form, nor deployed in part or in whole in any marketing, promotional or educational contexts without further discussion with Wiley. Permissions requests should be directed to <mailto:permreq@wiley.co.uk>
- Lead authors are cordially invited to remind their co-authors that the reprint opportunities detailed above are also available to them.
- If you wish to purchase print copies of the issue in which your article appears, please contact our Journals Fulfillment Department <mailto:cs-journals@wiley.co.uk> when you receive your complimentary offprints or when your article is published online in an issue. Please quote the Volume/Issue in which your article appears.

# A simulation study of IP-based vertical handoff in wireless convergent networks

Jung Kee Song and Wenye Wang<sup>\*,†</sup>

North Carolina State University, Raleigh, NC 27695, U.S.A.

## Summary

The advances in wireless networks and IP technologies has brought ubiquitous access to all-IP information into reality. For wireless IP services, IP-based handoff is a critical issue to the performance of application-level services. Although mobile IP (MIP) and its extensions, as network layer solutions, have been proposed as *de facto* standard, transmission throughput degradation due to packet loss, registration delay, and transport layer blocking are unavoidable because of MIP handoff mechanisms. In this paper, we evaluate the performance of a transport layer handoff approach, mobile stream control transmission protocol (mSCTP), and compare it with that of a network layer solution, MIP. mSCTP is based on stream control transmission protocol (SCTP), which is the third general purpose transport layer protocol from IETF. We investigate the use of mSCTP for *seamless vertical handoff* without any change in IP protocol stack by its multi-homing feature and dynamic address reconfiguration (DAR) extension. We evaluate the performance of mSCTP and MIP by introducing handoff delay, end-to-end transmission throughput, and packet loss, and verify our observations by a simulation study of the two protocols in UMTS/802.11b integrated networks using NS-2 network simulator. Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS: xxxxx<sup>Q2</sup>

Q2

## 1. Introduction

The Internet protocol suite has become one of the most essential and prevalent networking technologies in computer networks as well as in telecommunication networks. Based on its layered architecture and packet switching capability, IP technology offers unified, flexible, and scalable services to various applications over many heterogeneous networks. At the mean time, wireless technology erased the limitation of user mobility, which is originally ascribed to fixed wireline communications.

Among various wireless access technologies, *wireless local area networks* (LANs) and *cellular networks*

have turned out to be the most widely deployed infrastructures providing mobile access to voice and data services upon users' needs. The two technologies have complimentary characteristics in terms of physical specifications. For instance, the third generation (3G) universal mobile telecommunication systems (UMTS) are designed to support up to 384 Kbps data rate for pedestrian users and 2 Mbps rate for vehicular users within a few kilometers coverage areas. In this reason, exploiting their complimentary advantages, wireless LAN and 3G cellular networks can offer good quality of convergent services in a complimentary form.

Due to the advantages of IP protocol and various complimentary wireless technologies, wireless IP

\*Correspondence to: Wenye Wang, North Carolina State University, Raleigh, NC 27695, U.S.A.

†E-mail: wwang@eos.ncsu.edu

platform has been considered for many mobile applications. However, in the evolution of wireless IP networks, we encounter quite a few problems. IP mobility is an importunate topic among them. The inherent properties of wireless communication networks require inevitable handoff between two independent physical networks. Handoff procedures affect the overall performance of application level services. Since wireless IP convergent networks provide IP applications and services to mobile users in heterogeneous environments, IP-based vertical handoff solution is a fundamental issue for seamless roaming.

To support IP-based handoff, quite a few solutions have been proposed as network layer solutions including MIP, HAWAII [1], Cellular IP [2], and IDMP [3]. HAWAII, a domain-based mobility solution, Cellular IP, and IDMP are micro-mobility solutions proposed to solve out the signaling overhead problem of macro-mobility when a node frequently moves in foreign domain.

Although the proposed IP-based handoff solutions, such as mobile IP (MIP) which is a *de facto* standard to support network layer mobility, solved the fundamental problem of node movement in the Internet, they still incur significant handoff, delays, which affect the quality of applications and services, especially in mobile environments. Hence, seamless IP-based handoff solutions are definitely required for future wireless IP networks.

Meanwhile, a new method of IP-based handoff was proposed in transport layer using *stream control transmission protocol* (SCTP) [4]. SCTP provides a multi-homing feature with which a mobile node can hold multiple IP connections at a given moment. Based on multi-homing feature of SCTP, an extension, called *dynamic address reconfiguration* (DAR) [5], has been proposed as an Internet draft. DAR consists of pairs of request and response messages to update IP address information between two end nodes. The multi-homing feature of SCTP and DAR extension enables a transport layer seamless handoff by eliminating registration delay and tunneling overhead.

In other words, MIP, as a network layer solution, has been an Internet standard and most widely deployed but also incurs significant handoff delay. Mobile SCTP (mSCTP) for transport layer handoff, on the other hand, does not incur significant handoff delay based on the multi-homing feature of SCTP and DAR extension. However, there is a very limited study on the performance of these two solutions, neither comparison between two individual protocols, nor integrative study. Therefore, we focus on the performance evaluation of the two IP-based handoff solutions, MIP and mSCTP, a

new transport layer solution, in wireless IP convergent networks such that we verify the feasibility of the transport layer handoff comparing to MIP in future wireless IP networks.

In this paper, we introduce three performance metrics to compare the performance of MIP and mSCTP, including *total handoff delay*, *end-to-end throughput*, and *packet loss*. With the three performance metrics, we conduct an analysis and a simulation study using NS-2 network simulator [6] to evaluate the performance of MIP and mSCTP in UMTS/802.11b-integrated networks. We have observed the following characteristics of mSCTP over MIP:

- Total handoff delay of mSCTP is constant regardless of handoff rate while that of MIP increases proportional to handoff rate and residential time of a mobile node in foreign networks.
- End-to-end throughput of mSCTP is constant irrespective of handoff rate while that of MIP decreases proportional to handoff rate.
- mSCTP does not incur any packet loss based on its zero handoff delay while packet loss of MIP increases proportional to handoff rate such that transport layer transmission behavior can have negative effects.
- mSCTP does not require any third party agent while MIP needs additional efforts with the integration of agents into legacy architectures. This implies that mSCTP is easier to be deployed in heterogeneous networks.

The rest of the paper is organized as follows. In Section 2, we provide an overview of two IP-based handoff solutions, MIP and mSCTP. In Section 3, the vertical handoff architecture between UMTS and wireless LAN networks is described for in MIP and mSCTP, respectively. A detailed performance analysis of MIP and mSCTP operations is presented in Section 4. The simulation methodology, protocol suite, and simulation modules are presented in Section 5. Simulation results are demonstrated in Section 6 in which we compare the results from simulations with the analytical results. In Section 7, our observations and conclusions are summarized.

## 2. IP Mobility: Handoff Solutions

In this section, we describe two IP-based handoff solutions: a network layer approach, MIP and, a transport layer approach, mSCTP. First, we describe the basic

mechanism of MIP: agent discovery, registration, and tunneling. Then, SCTP and DAR extension of SCTP are covered as basic features of mSCTP.

## 2.1. Mobile IP: A Network Layer Approach

### 2.1.1. Overview

MIP is a network layer mobility solution for IP networks. MIP defines three basic components: a mobile node (MN) that wanders within MIP network; an home agent (HA), which is a special agent sitting on a router located in MN's home link and a foreign agent (FA), which is yet another special agent built in a router residing in foreign links. These three components cooperate to locate and register the current IP address of an MN as it moves across different IP subnets. MIP is also designed to provide mobility transparent packet transmission service, called *tunneling*, to upper layer protocols.

MIP handoff consists of two phases: *agent discovery* and *registration*. Agent discovery is a period in which an MN detects its movement from one subnet to another and obtains a new IP address, called care-of-address (CoA). Registration is a procedure in which an MN informs the HA its CoA, and the HA updates the binding information according to the registration request.

Meanwhile, MIP is designed to provide mobility-transparent packet forwarding to MN regardless of its location in foreign links. The mobility support is based on *tunneling* capability. From the information given by agent discovery, an HA sets up a virtual tunnel, which is a particular route, to the CoA of MN (either an FA's CoA or a collocated CoA). The HA forwards the packets, originally destined to the home address of the MN, to the CoA of the MN. MIP provides three tunneling options: *IP in IP encapsulation*, *minimal encapsulation*, and *generic routing encapsulation* (GRE) [7].

### 2.1.2. Problems

Although MIP resolves host mobility in IP layer, additional agents and tunneling overheads degrade performance of data transmission to/from mobile host when handoff occurs. Such problems include:

- *High handoff delay* MIP handoff delay is ascribed to agent discovery and registration periods as we discussed in this section. As we will discuss in Section 4, total MIP handoff delay increases proportional to handoff rate and the period of time MN stays in for-

eign links. Hence, MIP handoff can cause significant performance degradation, especially in large-scale mobility environments.

- *Tunneling overhead* [8] Tunneling mechanism of MIP offers mobility transparent routing service to upper layer protocols. However, two major drawbacks exist in tunneling. First, an additional IP header should be attached to an original IP datagram. The 20-byte overhead decreases end-to-end transmission throughput. Second, since tunneling generates triangular routing path, additional network delay cannot be avoided. Other issues such as conflict with network security solutions in MIP is beyond the scope of this paper [9].

### 2.1.3. Enhancement efforts

In order to reduce handoff delay and tunneling overhead incurred by MIP, quite a few enhancement efforts have been made. There have been many other extensions and drafts in terms of more general mobility architecture and performance, but we focus on MIP extensions regarding handoff performance. We discuss two major internet drafts *MIP with route optimization* and *hierarchical MIP*.

- *Optimized routing* Optimized routing [10] is an MIP extension to reduce tunneling overhead. In this extension, MN informs its CoA to CN directly so that CN can send data packets to the CoA directly without tunneling by HA. In order to realize the optimized routing, CN maintains a binding cache in which CoAs of MN are being updated as MN moves. For example, MIP version 6 [11] supports route optimization, which allows 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 MIP procedure, that is, via MN's home agent. MIP v6 includes imbedded binding updates and CoA configuration for the execution of location updates and for processing the change in the MN's address.
- *Regional registration and hierarchical MIP* MIP *regional registration* [12] and HMIP [13] are another MIP extension to reduce signaling overhead. Hierarchical MIP is categorized as a micro-mobility solution. When the frequency of an MN movement inside a subnet increases, signaling overhead for the

MN's registration of its CoA with the HA increases and cause high handoff delay. In order to reduce the signaling overhead, a hierarchical mobility agent, called *gateway foreign agent* (GFA) [12,14], is defined. Whenever a handoff is triggered from an MN movement, the MN registers its CoA with the most nearby local GFA. As the HA had already been noticed about the tunneling information to the GFA in the previous (or the first) registration, the tunneling of packets are processed in a hierarchical manner from the HA to the GFA, and the GFA to the FA, and the FA to the MN.

## 2.2. mobile SCTP: A Transport Layer Approach

mSCTP is a transport layer handoff solution based on SCTP [4] and DAR extension [5]. Originally designed to support telephony signaling messages, SCTP has been adopted as the third general purpose transport protocol<sup>‡</sup> by IETF, inheriting main features of TCP including the concept of flow control and congestion control. SCTP, in addition, provides two novel features, *multi-homing* and *multi-streaming* [15–17]. In addition, DAR extension is used to deliver messages of add-ip, delete-ip, and set-primary-ip requests. All the mSCTP handoff procedures are processed between two end-to-end hosts without involving any third party agent.

As discussed in Subsection 2.1, one of the major problems of MIP is that an MN cannot keep communicating with a CN while it has to deal with its registration of its CoA with the HA during handoff period. However, mSCTP can provide a seamless handoff based on its multi-homing feature and DAR extension.

### 2.2.1. SCTP association and multi-homing

A transport endpoint in TCP/IP network is canonically defined as a pair of IP address and port number. In TCP, a connection is established between two (IP address, port number) endpoints, and a TCP connection is always one-to-one relationship of a single IP address from each endpoint. Unlike in TCP, an SCTP endpoint can multiplex multiple IP addresses on a multi-homed (interfaced) host. An SCTP association is defined as [a set of IP addresses at A]+[Port-A]+[a set of IP ad-

<sup>‡</sup> Stream Control Transmission Protocol (SCTP) is the third general purpose transport protocol standardized by IETF, after TCP and UDP.

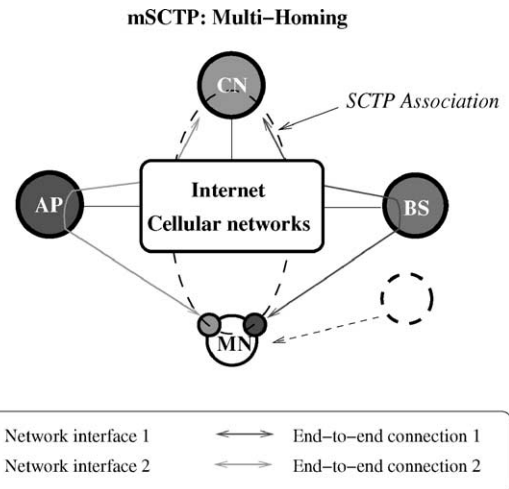


Fig. 1. mSCTP multi-homing.

resses at Z]+[Port-Z] [17]. That is, two SCTP endpoints, having set of IP addresses, have an SCTP *association*.

Meanwhile, an SCTP endpoint can establish multiple sessions (associations) to other endpoints at the same time [15]. That is, an SCTP endpoint can utilize its multi-homed network interfaces to signal to more than two peer nodes at a given moment. Based on association and multi-session features, an SCTP node can support multi-homing capability as shown in Figure 1.

In Figure 1, the MN and the CN establish an SCTP association. With the association, a multi-homed MN can utilize multiple-path communications by signaling to the AP and the BS at the same time while staying in the coverage of both networks, and notice the CN to use either of the MN's IP addresses as a destination IP address.

Since SCTP is a connection-oriented transport protocol, two endpoints should establish an *association* before exchanging data chunks. Unlike TCP, SCTP uses four-way handshake to setup an association [4]. Four-way handshake can prevent TCP SYN-flooding-attack, which is ascribed to the three-way handshake of the connection initialization of TCP. As SCTP is a reliable transport protocol, it has a procedure to close an existing association, which is a three-message handshake process. One of major differences between SCTP association close and TCP connection close process is that SCTP does not allow half-closed state that TCP does [4].

Once an association has been established, two SCTP endpoints can transfer user data. An SCTP packet is composed of an SCTP common header and number

of *chunks*. A *chunk* is a unit of information within an SCTP packet [15]. As a unit of SCTP message building block, many different types of chunks have been defined, categorized into either control chunks or data chunks [4]. Multiple chunks can be bundled together in an SCTP packet, regardless of whether they are control chunks or data chunks.

### 2.2.2. Dynamic address reconfiguration (DAR)

DAR extension defines three major parameters: *add IP address* (add-IP), *delete IP address* (delete-IP), and *set primary address* (set-primary-IP) [5]. Add-IP is a parameter to add new IP addresses in an active association. Delete-IP is to delete IP addresses from an existing association. Set-primary-IP function is used to inform the other end node to change the destination IP address.

In order to deliver these DAR parameters, two additional chunks, *address configuration change chunk* (ASCONF) and *address configuration acknowledgment* (ASCONF-ACK), are defined [5]. It should be noticed that ASCONF chunks can be bundled with other data chunks in an active association during an mSCTP handoff procedure. This property of SCTP and DAR makes mSCTP handoff delay be neglected as we will discuss in Section 4.

### 2.2.3. mSCTP: make-before-break with multi-homing and DAR

In this section, we have discussed essential components of *mSCTP* including SCTP and DAR extension. Based on multi-homing feature of SCTP and DAR, *make-before-break* IP-based handoff is able to be performed.

During a handoff period, legacy network layer handoff mechanisms, including MIP, have a certain period in which an MN must communicate with agents other than its peer node, CN. Hence, the existing active connection suffers from packet loss, waste of bandwidth by transport layer slow start, and so on. As handoff delay increases, the performance degradation becomes bigger.

On the other hand, mSCTP node can utilize two network adapters in an association to make a new data path while still communicating with its peer node. To perform a handoff, mSCTP exchange at most three pairs of ASCONF/ASCONF-ACK control chunks bundled with data chunks with its peer node. Figure 2 shows mSCTP handoff signaling messages. An MN has an ongoing communication with the CN through the BS node in this figure. For this communication, the MN used the network interface-1. During its movement to-

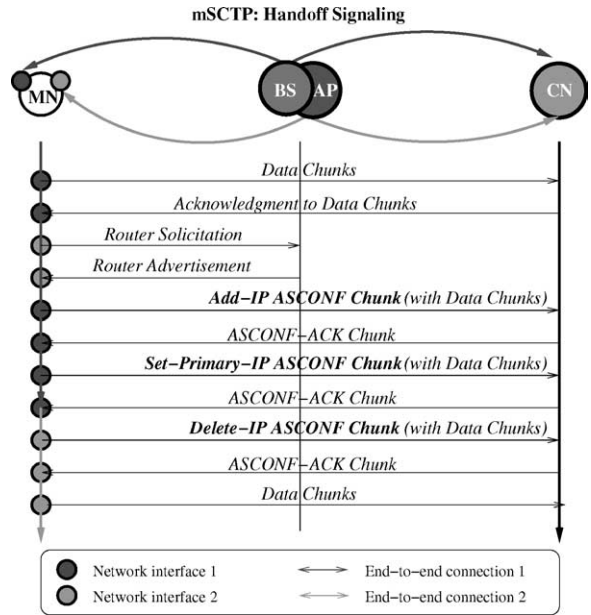


Fig. 2. mSCTP handoff signaling with multi-homing and DAR extension.

ward an AP, the MN exchanges router solicitation and router advertisement with the AP using the network interface-2. The router discovery process is performed independently from the data transmission in the network connection-1.

Once the MN obtains a new IP for the network interface-2, the MN bundle an add-IP ASCONF chunk into other data chunks in the network connection-1. The CN responds with ASCONF-ACK chunk bundled with SACK chunks for received data. When the MN determines to use the interface-2, the MN sends set-primary-IP ASCONF chunk together with data chunks to the CN. The CN replies back with ASCONF-ACK. Upon reception of the ASCONF-ACK chunk, the MN begins to use the interface-2. Delete-IP ASCONF and ASCONF-ACK can be exchanged to erase the old IP address from the association.

## 3. Vertical Handoff Architecture in Heterogeneous Networks

In this section, we discuss one of the promising trends in wireless networking, that is, the *convergent* networks and service. Then we describe to provide mobile services in *wireless convergent networks*, IP-based *vertical handoff* architecture is critical and use IP-based MIP and mSCTP as a case study.

### 3.1. Network and Service Convergence: Trend

Various wireless network services, including 2G, 2.5G/3G, and 802.11 Wi-Fi, are contemporarily wireless systems to meet different needs from customers. This variety comes from the fact that network operators decide type of network based on required coverage of network, target user base, their business profit, and so forth. The deployed networks are divided into three categories: wireless local area networks (LANs), wireless metro-area networks (MANs), and wireless wide area networks (WANs). Wireless LAN includes 802.11b, 802.11g, and 802.11a. Wireless MAN, which is targeting mobile Internet users in a city area, includes 802.16a, 802.16b, and 802.16e [18]. Wireless WAN includes all the current cellular infrastructure access networks such as CDMA2000, GSM, GPRS, and UMTS [19].

Among the different wireless access networks and services, Wireless LAN and wireless WAN have complementary characteristics. Wireless LAN generally provides relatively high data rate with small coverage while wireless WAN networks offer relatively low data rate with large coverage area. For instance, a wireless LAN standard, 802.11b, provides up to 11 Mbps data rate with a few hundred meters of radius while UMTS terrestrial radio network (UTRAN) supports up to 2 Mbps data rate for fixed node and 386 Kbps for pedestrian users with a few kilometers of radius. Therefore, by using both of the network services, wide coverage area is guaranteed with high data rate where wireless LAN services are available.

However, we should notice that certain problems exist due to the heterogeneity of the different technologies. Since it is wireless mobile communication, mobility support is again one of the most fundamental problem to be solved. In this paper, in order to evaluate the performance of handoff effects in integrated networks, UMTS/802.11b integrated networks are investigated. In the next section, we discuss the IP-based

handoff mechanisms for the *vertical handoff* between UMTS and 802.11b networks.

### 3.2. IP-Based Vertical Handoff in UMTS/802.11b

In addition to the existing IP-based wireless networks and services such as Wi-Fi hot spots, commercial cellular networks and services are also evolving toward all-IP networks due to many technological and economical reasons. Thus, IP-based handoff solutions will be an essential part of future wireless IP convergent networks. In the paper, in order to focus on only the handoff performance, we study loosely-coupled vertical handoff architecture [20] where 802.11b wireless LAN does not have any correlation with UMTS network but directly connected to public IP networks.

#### 3.2.1. Mobile IP vertical handoff

MIP, as an Internet Standard, has been deployed widely in various networks and services. For instance, 3GPP2 adopted MIP as a standard IP mobility solution and uses MIP in the packet switching domain of the networks. Moreover, the deployed MIP service has already been used to integrate the CDMA2000 network and the 802.11 wireless LAN network. An example of this integrated network services is the network service product called, *NETSPOT SWING*, provided by a Korean network operator, KTF. The concept of the service is that the multi-homed user device always connects to the CDMA2000 cellular service and MIP Handoff occurs whenever the user enters a Wi-Fi hotspot area.

For UMTS network, MIP has not yet been adopted as a standard IP mobility solution. One of the disadvantages of MIP from an architectural point of view is that it requires additional components, that is, HA and FA have to be installed in certain routers and/or network gateways. Figure 3 shows an example of MIP vertical handoff architecture in

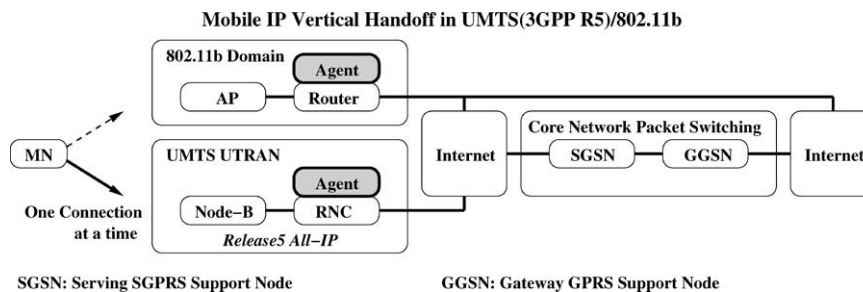


Fig. 3. MIP vertical handoff architecture: UMTS (3GPP R5)/802.11b.

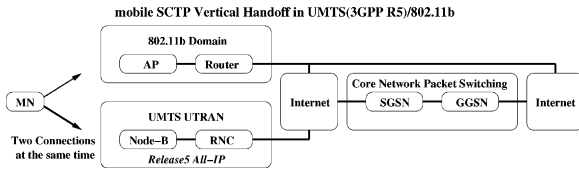


Fig. 4. mSCTP vertical handoff architecture: UMTS (3GPP R5)/802.11b.

UMTS (3GPP R5<sup>§</sup>)/802.11b-integrated network. In 3GPP R5 all-IP network, agent modules can be sit on the radio network controller (RNC) in UMTS. In this architecture, UMTS UTRAN supports IP-based protocol stack. This all-IP access network provides more flexible IP routing service, but MIP handoff overhead cannot be avoided. It should be noted that vertical handoff is an inter-technology handoff. Hence, MIP registration and tunneling incur higher delay than that of in homogeneous environment due to the routing overhead between different technologies. Moreover, in MIP vertical handoff architecture, certain modification has to be made into legacy network components. This task is not very favorable to network operators.

### 3.2.2. mobile SCTP vertical handoff

Unlike MIP vertical handoff, mSCTP vertical handoff does not require any additional modification on existing components. As discussed in Section 2, mSCTP directly interacts with CN to trigger transport layer handoff. This end-to-end handoff procedure is one of the most important advantages that SCTP offers such that no handoff architecture difference exists between homogeneous environments and heterogeneous environments. Figure 4 represents an example of mSCTP vertical handoff architecture in UMTS (3GPP R5)/802.11b-integrated network. mSCTP does not require any modification in this architecture, either.

## 4. Performance Analysis of Mobile IP and mobile SCTP Handoff

In this section, we evaluate the performance of MIP and mSCTP handoff with regard to *handoff delay*, *end-to-end throughput*, and *packet loss*. In Section 5, we conduct a simulation study with corresponding system models in UMTS/802.11b-integrated network.

<sup>§</sup> 3GPP R5 is the release that followed R99 and R4. The main issues in R5 are GSM/EDGE RAN (GERAN) and IP transport within the access network [21].

### 4.1. Handoff Delay

*Handoff delay* is  $t_{n+1} - t_n$ , for all  $n$ , where  $t_n$  is the time a packet sequence  $n$  is received on IP(A), an existing IP address of an MN, and  $t_{n+1}$  is the time a packet sequence  $n + 1$  is received on IP(B), which is a newly obtained IP address by the MN's movement into a new IP subnet. For analysis of handoff delay, we denote certain time of delay in particular operations during handoff procedures as  $T$  with subscripts.

#### 4.1.1. Handoff delay in Mobile IP: $T_{MIP}$

During the defined handoff delay period, MIP incurs two activities, agent discovery and registration, to process a handoff (see Subsection 2.1). Hence, MIP *handoff delay* ( $T_{MIP}$ ) consists of two phases: agent discovery ( $T_{ad}$ ) period and registration ( $T_{reg}$ ) period. That is,

$$T_{MIP} = T_{ad} + T_{reg} \quad (1)$$

Agent discovery ( $T_{ad}$ ) and registration ( $T_{reg}$ ), in turn, are composed of the following operations, respectively:

- *Agent discovery* ( $T_{ad}$ ) Agent discovery ( $T_{ad}$ ) period consists of agent solicitation ( $T_{as}$ ), agent advertisement ( $T_{aa}$ ), and CoA processing time ( $T_{CoA}$ ) of an MN. As we have discussed in Subsection 2.1, MIP handoff occurs when an MN moves from one subnet to another. For the movement detection of MN, agent discovery ( $T_{ad}$ ) is executed by an MN and an FA (or possibly HA) cooperation.

Agent solicitation ( $T_{as}$ ) is a modified ICMP message broadcasted by an MN to search for an agent. Agent advertisement ( $T_{aa}$ ) is also a modified ICMP broadcast message sent by an agent [22]. CoA processing time ( $T_{CoA}$ ) represents time taken by MN to dispatch a CoA from FA's agent advertisement message [13]. As a result, the delay incurred by agent discovery is:

$$T_{ad} = T_{as} + T_{aa} + T_{CoA} \quad (2)$$

By accomplishing the above three operations, agent discovery period of MIP ( $T_{ad}$ ) makes MN ready to trigger registration procedure ( $T_{reg}$ ) with its obtained CoA.

- *Registration* ( $T_{reg}$ ) Once an MN obtained a CoA from FA's agent advertisement, it triggers registration ( $T_{reg}$ ) of its CoA with the HA. Registration ( $T_{reg}$ ) period is composed of registration request ( $T_{reg-REQ}$ ) by MN, binding entry update for a new



CoA ( $T_{BU}$ ) by the HA, and registration response ( $T_{reg-RES}$ ) by the HA. That is,

$$T_{reg} = T_{reg-REQ} + T_{BU} + T_{reg-RES} \quad (3)$$

Registration process can be successfully completed only when all the parties do not encounter any error during  $T_{reg-REQ}$ ,  $T_{BU}$ , and  $T_{reg-RES}$  periods. If any error occurs (mostly authentication failure), the HA sends a registration reply message with a corresponding error code and registration request message is retransmitted by the MN after proper handling of the error specified in error code field. In order to prevent *denial-of-service* attack, which an unauthorized malicious node can flood registration traffic to an HA, registration request ( $T_{reg-REQ}$ ) and registration response ( $T_{reg-RES}$ ) messages should be authenticated by an HA and an MN, respectively. Registration request ( $T_{reg-REQ}$ ) and registration response ( $T_{reg-RES}$ ) are delivered in UDP packets.

Finally, agent discovery period and registration period constitutes MIP handoff delay ( $T_{MIP}$ ) as follows:

$$\begin{aligned} T_{MIP} &= T_{ad} + T_{reg} \\ &= \underbrace{T_{as} + T_{aa} + T_{CoA}}_{T_{ad}} \\ &\quad + \underbrace{T_{reg-REQ} + T_{BU} + T_{reg-RES}}_{T_{reg}} \end{aligned} \quad (4)$$

In Equation (4), agent discovery ( $T_{ad}$ ) period is an inevitable procedure due to the fact that the movement of an MN should be detected at IP layer. However, registration ( $T_{reg}$ ) period is part of the handoff delay in MIP because MIP is a network layer, modified routing mechanism. MIP updates this routing information located in HA's binding update table upon a registration request from an MN which obtained a new CoA.

Including binding update ( $T_{BU}$ ) period and an additional authentication overhead, registration delay ( $T_{reg}$ ) can be significantly long when an MN communicates with others in a large scale of mobility environment. It should be noticed that handoff delay ( $T_{MIP}$ ) may interrupt on-going data transmission such that end-to-end throughput decreases and data packet loss possibly occurs.

#### 4.1.2. Handoff delay in mobile SCTP: $T_{mSCTP}$

During the handoff period, mSCTP handoff generates router discovery procedure performed between an MN

and an access router, and DAR procedure between an MN and a CN. Hence, in order to analyze mSCTP handoff delay ( $T_{mSCTP}$ ), we employ router discovery period ( $T_{rd}$ ) and DAR period ( $T_{DAR}$ ), then mSCTP handoff delay  $T_{mSCTP}$  is given by:

$$T_{mSCTP} = T_{rd} + T_{DAR} \quad (5)$$

Router discovery ( $T_{rd}$ ) period and DAR ( $T_{DAR}$ ) procedure, in turn, are composed of the following operations, respectively:

- *Router discovery ( $T_{rd}$ )* Router discovery ( $T_{rd}$ ) period is composed of router solicitation ( $T_{rs}$ ), router advertisement ( $T_{ra}$ ), and processing time of newly obtained IP ( $T_{new-IP}$ ) in an MN's protocol stack. That is,

$$T_{rd} = T_{rs} + T_{ra} + T_{new-IP} \quad (6)$$

Router discovery ( $T_{rd}$ ) procedure of mSCTP is different from agent discovery ( $T_{ad}$ ) of MIP in two folds. First, agent advertisement ( $T_{aa}$ ) of MIP uses a modified ICMP router advertisement ( $T_{ra}$ ) [23]. On the other hand, mSCTP handoff can use a standard ICMP router advertisement message, which is at least 12-byte shorter than an agent advertisement of MIP. This means that the signaling overhead in router discovery ( $T_{rd}$ ) of mSCTP is not greater than that of agent discovery ( $T_{ad}$ ) of MIP. That is,

$$MIP : T_{ad} > mSCTP : T_{rd} \quad (7)$$

Second and more importantly, router discovery ( $T_{rd}$ ) of mSCTP can be performed while transmitting data packets in an SCTP association exploiting multi-homing feature of SCTP. Hence, actual delay caused by mSCTP router discovery ( $T_{rd}$ ) can be neglected. That is,

$$T_{rd} \approx 0 \quad (8)$$

- *Dynamic address reconfiguration ( $T_{DAR}$ )* The other part of mSCTP handoff procedure is a transport layer dynamic IP address configuration procedure employed from DAR extension of SCTP. DAR procedure period ( $T_{DAR}$ ) consists of three ASCONF parameters: add-IP ( $T_{add-IP}$ ) procedure, set-primary-IP ( $T_{set-primary-IP}$ ) procedure, and delete-IP ( $T_{del-IP}$ ) procedure. That is,

$$T_{DAR} = T_{add-IP} + T_{set-primary-IP} + T_{del-IP} \quad (9)$$

Each ASCONF parameter is delivered in an ASCONF chunk and replied in an ASCONF-ACK chunk. Therefore, three pairs of ASCONF/ASCONF-ACKs are exchanged (see Subsections 2.2.2 and 2.2.3):

$$T_{\text{DAR}} = 3 \times (T_{\text{ASCONF}} + T_{\text{ASCONF-ACK}}) \quad (10)$$

As we discussed in Subsection 2.2.2, all the control chunks in DAR procedure can be bundled with other data chunks in mSCTP transmission. Thus, the actual delay caused by DAR ( $T_{\text{DAR}}$ ) procedure also becomes zero. This is a very important fact such that the handoff delay of mSCTP ( $T_{\text{mSCTP}}$ ) becomes zero. Consequently, mSCTP handoff delay is expressed as:

$$\begin{aligned} T_{\text{mSCTP}} &= T_{\text{rd}} + T_{\text{DAR}} \\ &= (T_{\text{rs}} + T_{\text{ra}} + T_{\text{new-IP}}) \\ &\quad + (T_{\text{add-IP}} + T_{\text{set-primary-IP}} + T_{\text{del-IP}}) \\ &= \underbrace{(T_{\text{rs}} + T_{\text{ra}} + T_{\text{new-IP}})}_{T_{\text{rd}} \approx 0 \text{ (by multi-homing)}} \\ &\quad + \underbrace{[3 \times (T_{\text{ASCONF}} + T_{\text{ASCONF-ACK}})]}_{T_{\text{DAR}} \approx 0 \text{ (by chunk-bundling)}} \\ &\approx 0 \end{aligned} \quad (11)$$

As we will discuss in Subsection 4.2, although DAR procedure incurs a constant throughput decrease, it allows mSCTP supports quasi-seamless handoff (i.e.,  $T_{\text{mSCTP}} \approx 0$ ).

#### 4.1.3. Comparison of handoff delay in MIP ( $T_{\text{MIP}}$ ) and mSCTP ( $T_{\text{mSCTP}}$ )

As discussed in Subsections 4.1.1 and 4.1.2, MIP agent discovery ( $T_{\text{ad}}$ ) and mSCTP router discovery ( $T_{\text{rd}}$ ) are as follows:

$$\underbrace{T_{\text{as}} + T_{\text{aa}} + T_{\text{CoA}}}_{\text{MIP:}T_{\text{ad}}} \gg \underbrace{T_{\text{rs}} + T_{\text{ra}} + T_{\text{new-IP}}}_{\text{mSCTP:}T_{\text{rd}} \approx 0 \text{ (by multi-homing)}} \quad (12)$$

That is,

$$\text{MIP} : T_{\text{ad}} \gg \text{mSCTP} : T_{\text{rd}} \quad (13)$$

As shown<sup>Q3</sup> in Equation (12), agent discovery ( $T_{\text{ad}}$ ) of MIP is greater than router discovery ( $T_{\text{rd}}$ ) of mSCTP since multi-homing feature of SCTP makes the actual

delay of router discovery ( $T_{\text{rd}}$ ) of mSCTP be neglected (refer to Equation (5)).

Likewise, the relationship between MIP registration delay ( $T_{\text{reg}}$ ) and mSCTP DAR procedure ( $T_{\text{DAR}}$ ) is as follows:

$$\begin{aligned} &\underbrace{T_{\text{reg-REQ}} + T_{\text{BU}} + T_{\text{reg-RES}}}_{\text{MIP:}T_{\text{reg}}} \\ &\gg \underbrace{T_{\text{add-IP}} + T_{\text{set-primary-IP}} + T_{\text{delete-IP}}}_{\text{mSCTP:}T_{\text{DAR}} \approx 0 \text{ (by chunk-bundling)}} \quad (14) \end{aligned}$$

That is,

$$\text{MIP} : T_{\text{reg}} \gg \text{mSCTP} : T_{\text{DAR}} \quad (15)$$

This is because the ASCONF chunks in the DAR procedure ( $T_{\text{DAR}}$ ) can be bundled with other data chunks in mSCTP transmission (refer to Equation (5)). Finally, we come to the following conclusion for handoff delay of MIP ( $T_{\text{MIP}}$ ) and handoff delay of mSCTP ( $T_{\text{mSCTP}}$ ).

$$\text{MIP} : T_{\text{ad}} + T_{\text{reg}} \gg \text{mSCTP} : T_{\text{rd}} + T_{\text{DAR}} \quad (16)$$

That is,

$$\text{MIP} : T_{\text{MIP}} \gg \text{mSCTP} : T_{\text{mSCTP}} \quad (17)$$

As demonstrated in Section 6, when handoff rate increases, the accumulated total handoff delay of MIP increases and affects the overall transmission throughput while that of mSCTP is fairly constant near zero, which means seamless data transmissions in a large scale mobility environment.

## 4.2. End-to-End Throughput

The second parameter we analyze is end-to-end throughput of MIP and mSCTP. As discussed in Subsection 4.1, MIP incurs certain handoff delay, which directly affects end-to-end throughput decrease. In addition, tunneling overhead is to be considered as a throughput degradation factor as well.

In case of mSCTP, router discovery and DAR procedure are required to allow a CN become aware of the newly obtained IP address of an MN. However, mSCTP handoff delay can be neglected as shown in Equation (5) in Subsection 4.1, and no other significant factors affect end-to-end throughput of mSCTP. In this section, we introduce a new variable, *number of handoff* or *handoff rate*, to analyze end-to-end throughput

of MIP and mSCTP with regard to handoff rate. End-to-end throughput shows how the two different handoff protocols affect the overall transmission efficiency.

*End-to-end throughput* is the total data bits an end node receives during transmission duration.

#### 4.2.1. End-to-end throughput in MIP $\eta_{\text{MIP}}$

Since MIP handoff incurs data loss during its handoff delay periods, end-to-end transmission throughput of MIP ( $\eta_{\text{MIP}}$ ), when  $\delta$  number of handoff occurs during  $T_s$  of transmission period, can be denoted as follows:

$$\begin{aligned} \eta_{\text{MIP}} &= \frac{\mu_{\text{TCP}} - \left( \sum_{i=1}^{\delta} l_{\text{MIP}}(i) + l_{\text{tunnel}} \right)}{T_s} \\ &\approx \frac{\mu_{\text{TCP}} - \left[ (\delta \times \bar{l}_{\text{MIP}}) + l_{\text{tunnel}} \right]}{T_s} \quad (\text{bps}) \quad (18) \end{aligned}$$

As shown in Equation (18), end-to-end throughput of MIP ( $\eta_{\text{MIP}}$ ) is total TCP traffic offered ( $\mu_{\text{TCP}}$ ) subtracted by total data loss due to MIP handoff delay (sum of  $l_{\text{MIP}}(i)$  where  $i = 1$  to  $\delta$ ) and tunneling overhead ( $l_{\text{tunnel}}$ ) divided by total transmission duration ( $T_s$ ). Hence, end-to-end throughput of MIP ( $\eta_{\text{MIP}}$ ) depends upon loss of traffic due to MIP handoff ( $l_{\text{MIP}}$ ) and tunneling overhead ( $l_{\text{tunnel}}$ ). End-to-end throughput of MIP ( $\eta_{\text{MIP}}$ ) decreases proportional to number of handoff during the transmission period ( $\delta$ ). Now, let us examine amount of data loss due to MIP handoff ( $l_{\text{MIP}}$ ) and tunneling overhead ( $l_{\text{tunnel}}$ ):

$$\bar{l}_{\text{MIP}} = \mu_{\text{TCP}} \times \frac{\bar{T}_{\text{MIP}}}{T_s} \quad (19)$$

which represents the average data loss per MIP handoff ( $\bar{l}_{\text{MIP}}$ ) in (bit) in terms of MIP handoff delay ( $T_{\text{MIP}}$ ) discussed in Subsection 4.1.

$$l_{\text{tunnel}} = \mu_{\text{TCP}} \times \frac{T_{\text{fl}}}{T_s} \times \frac{F_{\text{MIP}}}{\text{PMTU}} \quad (20)$$

where  $F_{\text{MIP}} = 160$  (bit) and  $\text{PMTU} = 4608$ (bit), then

$$l_{\text{tunnel}} \approx 0.0347 \times \frac{\mu_{\text{TCP}}}{T_s} \times T_{\text{fl}} \quad (21)$$

Equation (21) shows data loss incurred by MIP tunneling ( $l_{\text{tunnel}}$ ) during the transmission period. The variable  $T_{\text{fl}}$  denotes the total time an MN stays in foreign links other than its home link such that tunneling is required.  $F_{\text{MIP}}$  is the size of tunneling (outer) IP header

attached to tunneling packets. PMTU is a path maximum transmission unit to which all the packets larger than this should be fragmented. Here, we assume that PMTU is set to 576 bytes (4608 bits) since 576-byte is the size of a packet that all IPv4 nodes must be able to receive. As a result, in Equation (4.21), tunneling overhead ( $l_{\text{tunnel}}$ ) can be explained as 0.0347 ( $F_{\text{MIP}}$  bit per fragmented IP datagram) multiplied by the fraction of data traffic when an MN stays in foreign links.

Therefore, we can observe in Equations (19) and (21) that data loss due to MIP handoff ( $l_{\text{MIP}}$ ) and tunneling overhead ( $l_{\text{tunnel}}$ ) increase proportional directly to MIP handoff delay ( $T_{\text{MIP}}$ ) and the fraction of period of time an MN stays in foreign links ( $T_{\text{fl}}$ ).

Now, by plugging the result of (?) and (21) to (?), end-to-end throughput of MIP ( $\eta_{\text{MIP}}$ ) is derived as follows<sup>Q4</sup>:

$$\begin{aligned} \eta_{\text{MIP}} &\approx \frac{\mu_{\text{TCP}} - \left[ (\delta \times \bar{l}_{\text{MIP}}) + l_{\text{tunnel}} \right]}{T_s} \\ &= \frac{\mu_{\text{TCP}} - \left[ (\delta \times \mu_{\text{TCP}} \times \frac{\bar{T}_{\text{MIP}}}{T_s}) + (0.0347 \times \frac{\mu_{\text{TCP}}}{T_s} \times T_{\text{fl}}) \right]}{T_s} \\ &= \frac{\mu_{\text{TCP}} - \frac{\mu_{\text{TCP}}}{T_s} \left[ (\delta \times \bar{T}_{\text{MIP}}) + (0.0347 \times T_{\text{fl}}) \right]}{T_s} \\ &= \frac{\mu_{\text{TCP}}}{T_s} - \frac{\mu_{\text{TCP}}}{T_s^2} \left[ (\delta \times \bar{T}_{\text{MIP}}) + (0.0347 \times T_{\text{fl}}) \right] \quad (22) \end{aligned}$$

Thus, end-to-end throughput of MIP ( $\eta_{\text{MIP}}$ ) also decreases directly proportional to MIP handoff delay ( $T_{\text{MIP}}$ ) and the fraction of period of time an MN stays in foreign links ( $T_{\text{fl}}$ ). In addition, when the number of handoff during the transmission ( $\delta$ ) increases, end-to-end throughput ( $\eta_{\text{MIP}}$ ) of MIP decreases proportional to this handoff rate.

#### 4.2.2. End-to-end throughput in mSCTP:

$\eta_{\text{mSCTP}}$

End-to-end throughput of mSCTP ( $\eta_{\text{mSCTP}}$ ) with  $\delta$  handoff (number of handoff during  $T_s$ ) is denoted as follows:

$$\begin{aligned} \eta_{\text{mSCTP}} &= \frac{\mu_{\text{SCTP}} - \sum_{i=1}^{\delta} (l_{\text{mSCTP}}(i) + l_{\text{DAR}})}{T_s} \\ &\approx \frac{\mu_{\text{SCTP}} - \left[ \delta \times (\bar{l}_{\text{mSCTP}} + l_{\text{DAR}}) \right]}{T_s} \quad (\text{bps}) \quad (23) \end{aligned}$$

where  $l_{\text{mSCTP}}(i)$  denotes data loss incurred by  $i$ th mSCTP handoff delay.  $l_{\text{DAR}}$  represents DAR chunk overhead required for an mSCTP handoff. End-to-end throughput of mSCTP ( $\eta_{\text{mSCTP}}$ ) is the total SCTP traffic offered ( $\mu_{\text{SCTP}}$ ) subtracted by the total data loss due to mSCTP handoff delay (sum of  $l_{\text{mSCTP}}(i)$  where  $i = 1$  to  $\delta$ ) and DAR chunk overhead, divided by the total transmission duration ( $T_s$ ). Hence, end-to-end throughput of mSCTP ( $\eta_{\text{mSCTP}}$ ) depends on the amount of data loss due to mSCTP handoff ( $l_{\text{mSCTP}}$ ) and decreases directly proportional to handoff rate. Now, let us discuss data loss incurred by mSCTP handoff delay ( $l_{\text{mSCTP}}$ ) and DAR chunk overhead ( $l_{\text{DAR}}$ ) (We employ a parameter,  $\bar{l}_{\text{mSCTP}}$ , average data loss per mSCTP handoff):

$$\bar{l}_{\text{mSCTP}} = \mu_{\text{SCTP}} \times \frac{\bar{T}_{\text{mSCTP}}}{T_s} \quad (24)$$

where  $\bar{T}_{\text{mSCTP}} \approx 0$

$$\bar{l}_{\text{mSCTP}} \approx 0 \quad (25)$$

Because the average handoff delay of mSCTP is zero (see Equation (5)),  $\bar{l}_{\text{mSCTP}}$  becomes zero.

$$l_{\text{DAR}} = 3 \times (C_{\text{ASCONF}} + C_{\text{ASCONF-ACK}}) \quad (26)$$

where  $C_{\text{ASCONF}} = 192$  bits, and  $C_{\text{ASCONF-ACK}} = 64$  bits,

$$\begin{aligned} l_{\text{DAR}} &= 3 \times (192 + 64) \\ &= 768 \text{ bits} \end{aligned} \quad (27)$$

Equation (26) shows DAR chunk overhead ( $l_{\text{DAR}}$ ), similar to Equation (10), three DAR ASCONF parameters (add-IP, set-primary-IP, and delete-IP) require three pairs of ASCONF/ASCONF-ACK chunk. Thus, DAR procedure incurs three times  $C_{\text{ASCONF}}$  and  $C_{\text{ASCONF-ACK}}$  bits overhead per mSCTP handoff. Since the size of ASCONF chunk and ASCONF-ACK chunk are 192 bits and 64 bits, respectively (see Subsubsection 2.2.2), ( $l_{\text{DAR}}$ ) becomes 768 bits Equation (26).

Although the three pairs of ASCONF and ASCONF-ACK chunks can be bundled with other data chunks, it incurs 768 bits throughput decrease. As a result, loss of traffic due to an mSCTP handoff ( $l_{\text{mSCTP}}$ ) converges to a constant value, 768 bits, which is not significant over handoff rate.

Now, by applying the results of Equations (25) and (26) to Equation (23), end-to-end throughput of

mSCTP ( $\eta_{\text{mSCTP}}$ ) is derived as follows:

$$\begin{aligned} \eta_{\text{mSCTP}} &\approx \frac{\mu_{\text{SCTP}} - [\delta \times (\bar{l}_{\text{mSCTP}} + l_{\text{DAR}})]}{T_s} \\ &= \frac{\mu_{\text{SCTP}} - [\delta \times (0 + 768)]}{T_s} \\ &= \frac{\mu_{\text{SCTP}} - (\delta \times 768)}{T_s} \end{aligned} \quad (28)$$

We can see that end-to-end throughput of mSCTP ( $\eta_{\text{mSCTP}}$ ) is represented as the total SCTP traffic load without any MN movement ( $\mu_{\text{SCTP}}$ ) subtracted by 768-bit DAR procedure overhead for number of handoffs during the transmission, divided by the transmission duration ( $T_s$ ). That is, mSCTP handoff does not incur much loss of traffic and is able to support seamless handoff with a constant 768-bit throughput decrease per handoff.

#### 4.2.3. Comparison of end-to-end throughput in MIP and mSCTP

To compare end-to-end throughput of MIP ( $\eta_{\text{MIP}}$ ) and that of mSCTP ( $\eta_{\text{mSCTP}}$ ), we refer to the derived throughput of  $\eta_{\text{MIP}}$  in Equation (22) and  $\eta_{\text{mSCTP}}$  in Equation (23):

$$\begin{aligned} &\frac{\mu_{\text{TCP}}}{T_s} - \frac{\mu_{\text{TCP}}}{T_s^2} [(\delta \times \bar{T}_{\text{MIP}}) + (0.0347 \times T_{\text{fl}})] \\ &\ll \frac{\mu_{\text{SCTP}} - (\delta \times 768)}{T_s} \end{aligned} \quad (29)$$

Thus, end-to-end throughput of MIP ( $\eta_{\text{MIP}}$ ) decreases by the product of number of handoff ( $\delta$ ) and MIP handoff delay ( $T_{\text{MIP}}$ ) and the fraction of the period of time MN stays in foreign links ( $T_{\text{fl}}$ ). On the other hand, end-to-end throughput of mSCTP ( $\eta_{\text{mSCTP}}$ ) maintains total offered traffic load excluding 768-bit per handoff (which does not depend on handoff delay) divided by transmission duration ( $T_s$ ), which proves:

$$\text{MIP} : \eta_{\text{MIP}} \ll \text{mSCTP} : \eta_{\text{mSCTP}} \quad (30)$$

#### 4.3. Packet Loss

In this section, we analyze packet loss based on handoff delay and data loss discussed in the above two sections. *Packet loss* is the total number of packets lost during handoff periods.

### 4.3.1. Packet loss in MIP: $L_{MIP}$

Packet loss of MIP handoff ( $L_{MIP}$ ) can be represented as follows:

$$\begin{aligned}
 L_{MIP} &= \sum_{i=1}^{\delta} \frac{l_{MIP}(i)}{S} \\
 &\approx \frac{\delta \times \bar{l}_{MIP}}{S} \text{ (number of packet)} \\
 &= \frac{\delta \times \frac{\mu_{TCP}}{T_s} \times \bar{T}_{MIP}}{S} \\
 &= \delta \times \bar{T}_{MIP} \times \frac{\mu_{TCP}}{S \times T_s} \quad (31)
 \end{aligned}$$

where  $l_{MIP}(i)$  denotes data loss incurred by  $i$ th MIP handoff delay. Packet loss of MIP handoff ( $L_{MIP}$ ) with  $\delta$  handoff during transmission duration ( $T_s$ ) is the sum of data loss incurred by each handoff divided by a packet size ( $S$ ). As a result of derivation of Equation (32), MIP handoff ( $L_{MIP}$ ) is directly proportional to the handoff delay and handoff rate product.

### 4.3.2. Packet loss in mSCTP: $L_{mSCTP}$

Packet loss of mSCTP handoff ( $L_{mSCTP}$ ) can also be denoted as the sum of data loss incurred by each handoff delay divided by a packet size ( $S$ ).

$$\begin{aligned}
 L_{mSCTP} &= \sum_{i=1}^{\delta} \frac{l_{mSCTP}(i)}{S} \\
 &\approx \frac{\delta \times \bar{l}_{mSCTP}}{S} \text{ (number of packet)} \\
 &= \frac{\delta \times \frac{\mu_{SCTP}}{T_s} \times \bar{T}_{mSCTP}}{S} \\
 &= 0 \quad (32)
 \end{aligned}$$

However, in the above Equation (32), packet loss of mSCTP ( $L_{mSCTP}$ ) becomes zero since an average handoff delay of mSCTP is zero as derived in Equation (5). That is, mSCTP packet loss is theoretically zero regardless of handoff rate.

### 4.3.3. Comparison of packet loss in MIP ( $L_{MIP}$ ) and mSCTP ( $L_{mSCTP}$ )

Packet loss, in the paper, shows how reliably handoff protocol supports data transport between an MN and a CN. From Equations (31) and (32), it turned out that

the packet loss of mSCTP handoff is less than that of MIP.

$$MIP : \delta \times \bar{T}_{MIP} \times \frac{\mu_{TCP}}{S \times T_s} \gg mSCTP : 0 \quad (33)$$

Especially, as handoff rate becomes higher, packet loss of MIP and mSCTP has greater difference, which means end-to-end transmission of mSCTP can be more reliable and efficient than that of MIP. Finally, we have the following result with packet loss of MIP and mSCTP.

$$MIP : L_{MIP} \gg mSCTP : L_{mSCTP} \quad (34)$$

## 5. Simulation Methodology

In previous sections, we have discussed the network layer handoff, MIP, and the transport layer handoff, mSCTP. We evaluated the performance of the two different handoff mechanisms with regard to *handoff delay*, *end-to-end throughput*, and *packet loss*. In this section, we describe details of our simulation and results. First, we introduce major modules used in our simulation. Second, our system model, including the network architecture, handoff protocol stacks, and designed scenarios, is followed. Finally, we represent and analyze the simulation results in terms of total handoff delay, end-to-end throughput, and packet loss in UMTS/802.11b-integrated network.

### 5.1. NS-2 and Related Modules

For the simulations of the described network architectures in this chapter, NS-2 network simulator [6] has been used. In addition, a contributed module, NS-2 SCTP [24], has been patched into an original NS-2 version 2.26 [25]. Meanwhile, NS-2 does not provide any standard CDMA implementation at the moment. Hence, we modified settings of existing wireless LAN MAC protocol to simulate UMTS access network.

### 5.2. System Models

#### 5.2.1. Assumptions

We start with the basic assumptions made for our system model.

- *Coverage of networks* Throughout our simulation work, we assume that the coverage of networks is

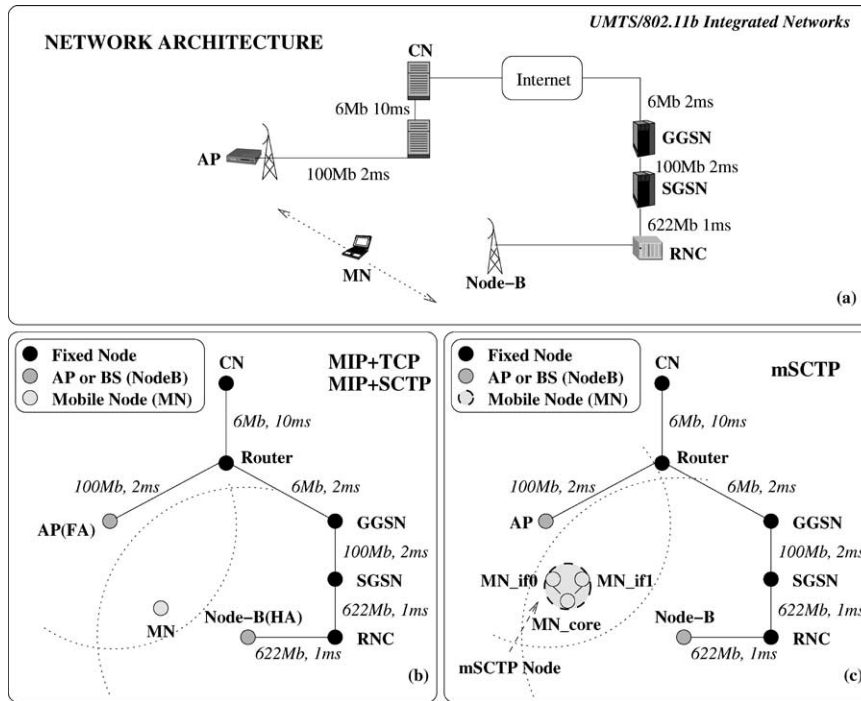


Fig. 5. Network Architecture: (a) UMTS/802.11b integrated network (b) NS-2 Node Topology for Mobile IP + TCP and Mobile IP + SCTP (c) NS-2 Node Topology for mobile SCTP.

overlapped. No matter which handoff protocols are used, physically exclusive networks cannot avoid fundamental interruption of data transmission. Thus, we set this assumption to evaluate handoff protocols in seamlessly covered areas.

- *Multi-homed facility* Multi-homing capability of an mSCTP MN is a basic assumption set by the protocol itself. We have built the mSCTP simulation models based on this assumption.
- *MIP in UMTS* Unlike in 3GPP2, MIP has not yet been adopted as a standard mobility solution in 3GPP. In the released UMTS architecture, we assume that the Node-B has MIP agent capability for simulation purpose.
- *Simplified UTRAN protocol stack* As it is the simulation of IP layer handoff and transport layer handoff targeting all-IP oriented environment, the UMTS UTRAN interface protocol stack has been simplified for the purpose of focusing on the different handoff mechanism itself. Details of protocol stacks will be discussed in Subsection 5.3.

### 5.2.2. Network architecture

The *network architecture* is designed to simulate three handoff protocol stacks in UMTS/802.11b wireless

LAN integrated networks: MIP+TCP, MIP+SCTP, and mSCTP. (We will describe the protocol stacks in Subsection 5.3 with more details.) Figure 5(a) shows the UMTS/802.11b wireless LAN integrated networks architecture. Figure 5(b) and (c) represent the corresponding NS-2 topologies for the network architecture. In Figure 5(b) topology, MIP+TCP and MIP+SCTP have been simulated, and in Figure 5(c), mSCTP has been simulated. In order to simulate mSCTP handoff in UMTS/802.11b-integrated networks in Figure 5(c), three NS-2 nodes (MN\_core, MN\_if0, and MN\_if1) have been employed. During the handoff period, MN\_if0 interface and MN\_if1 interface maintain multiple end-to-end connections in a form of SCTP association. The multi-homing feature enables an mSCTP MN perform virtually seamless handoff by maintaining more than a single stream of communication.

### 5.3. Protocol Stacks

In this section, the three handoff protocol stacks used in the simulation are described. The three protocol stacks include TCP over MIP (MIP+TCP), SCTP over MIP (MIP+SCTP), and mSCTP. MIP+TCP and MIP+SCTP are categorized as network layer handoff solutions, and

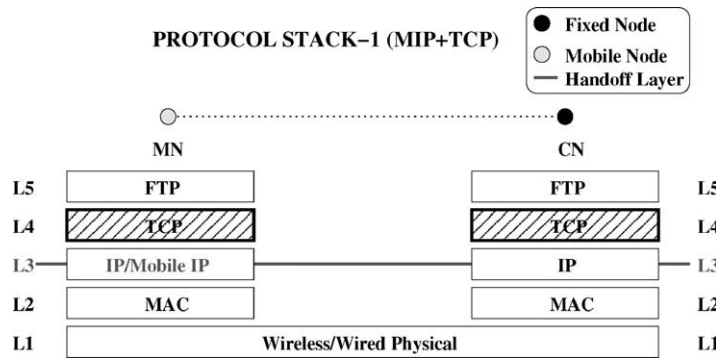


Fig. 6. MIP+TCP protocol stack.

mSCTP is in the category of transport layer handoff solution.

### 5.3.1. TCP over MIP (MIP+TCP)

As the first set of handoff protocol, we use MIP+TCP. This protocol stack is to simulate the traditional MIP handoff approach. Figure 6 shows MIP+TCP protocol stack. In L3, MN and HA/FA are set with NS-2 MIP module to support network layer handoff. Once MN moves into a new foreign link, MIP-enabled AP (HA/FA) node tunnels packets to the MN through router nodes including the FA. Fixed nodes provide regular IP protocol. In L4, two end nodes, MN and CN, establish TCP connection based on NS-2 TCP agent module.

### 5.3.2. SCTP over MIP (MIP+SCTP)

As the second set of handoff protocol, we used MIP+SCTP to evaluate the performance of MIP while giving a fair comparison with mSCTP handoff. By using SCTP as a transport layer protocol over MIP, MIP and mSCTP can be compared under the effect of the

same transmission mechanism. As in traditional MIP, handoff can be supported by MIP+SCTP based on the registration of CoA and packet tunneling in network layer by HA and FA. Only difference from MIP+TCP is that SCTP transport protocol is sat on MIP capability. Figure 7 shows MIP+SCTP protocol stack. As Figure 7 shows, MIP+TCP and MIP+SCTP protocol stacks are identical with regard to the structure except the fact that the two protocol stacks use different transport layer protocols. That is, in MIP+SCTP L4 layer, two end nodes, MN and CN, establish SCTP connection based on NS-2 SCTP module. Both protocol stacks were designed to evaluate network layer (L3) handoff performance based on MIP. However, MIP+SCTP is designed to compare network layer handoff to transport layer handoff with fairness in terms of transmission capability.

### 5.3.3. mobile SCTP (mSCTP)

As a third set of handoff protocol, we use mSCTP which is to simulate the transport layer end-to-end handoff mechanism. With this protocol stack, all the data packets are routed based on regular IP routing mechanism. Figure 8 shows mSCTP protocol stack. In L3, MN and

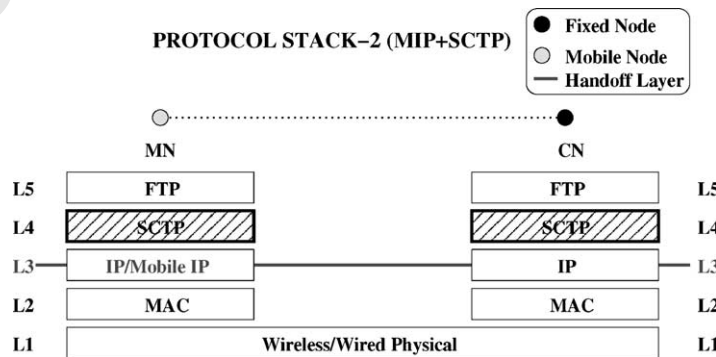


Fig. 7. MIP+SCTP protocol stack.

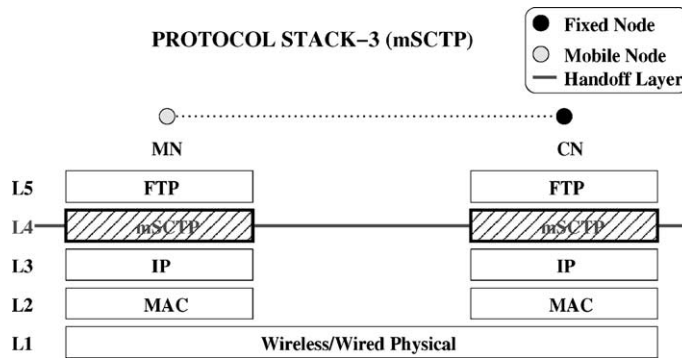


Fig. 8. mSCTP protocol stack in 802.11b wireless LAN topology.

AP nodes are set with regular IP module since no agent is required in this protocol stack. In L4, two end nodes, MN and CN, establish mSCTP association based on NS-2 SCTP module. MN and CN also interact in end-to-end manner based on DAR extension (see Subsubsection 2.2.2) to support transport layer handoff.

#### 5.4. Simulation Scenarios

In this section, we define the simulation scenarios based on the network architecture, the protocol stacks, and additional parameters including the direction of transmission, handoff rate, and MN movement pattern.

##### 5.4.1. Scenarios

The scenarios include two MIP handoff protocols (MIP+TCP and MIP+SCTP) and mSCTP handoff in UMTS/802.11b-integrated network architecture described in Subsubsection 5.2.2. The scenarios are designed to evaluate the performance of MIP handoff and mSCTP handoff in a fair manner. That is, we assumed the transmission behavior of TCP is fairly different from that of SCTP with regard to its congestion control, flow control mechanism, interaction with other layers in the stack, etc., so it should be required to simulate the behavior of SCTP as a pure transport layer protocol over MIP and compare it to mSCTP transmission.

In summary, Table I lists the simulation scenarios designed based on the network architecture and the three protocol stacks.

##### 5.4.2. Handoff Rate

In order to evaluate the performance of the three handoff protocol stacks, we introduce a variable *handoff rate* in the simulation. *Handoff rate* ( $\delta$ ) is number of handoffs during our simulation unit time 600 s (10 min).

We increase handoff rate ( $\delta$ ) and investigated how the three handoff protocol stacks react with respect to total handoff delay, end-to-end throughput, and packet loss. The reason we employed handoff rate ( $\delta$ ) is to measure the variation of performance parameters according to different scale of mobility. Hence our simulation results will show which handoff protocol is more seamless and reliable irrespective of handoff rate ( $\delta$ ). For nomadic users, the performance of handoff protocol does not give much negative effect, but we aimed to evaluate the performance of the handoff protocols for ubiquitous users.

During 600 seconds of transmission duration, 0 to 10 handoff occurrences have been simulated. That is, the lowest rate of 0 handoff to the highest rate of 10 handoffs per 10-min unit time have been simulated to evaluate the performance of the three handoff protocol stacks.

##### 5.4.3. MN movement

Figure 9 shows the MN movement pattern. An AP, A, and a Node-B, B, are placed in UMTS/802.11b-integrated networks topology. The radius of a 802.11b network is set to 250 m, and that of UMTS is set to 500 m. The MN, initially, stays in the coordination of (250, 250) in UMTS region. According to the handoff rate to simulate, the MN moves between UMTS

Table I. Simulation scenarios.

Network		Protocol		Direction	
1	UMTS/802.11b	1	MIP+TCP	1	Uplink
				2	Downlink
		2	MIP+SCTP	1	Uplink
				2	Downlink
		3	mSCTP	1	Uplink
				2	Downlink



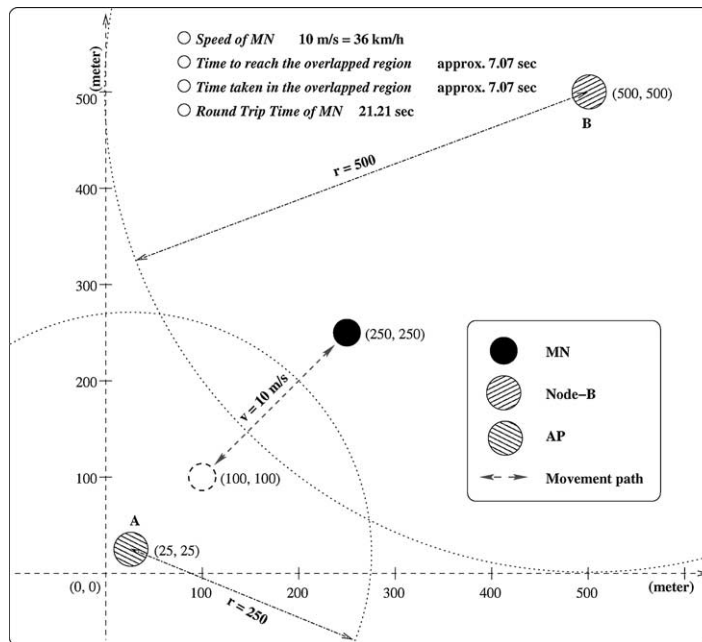


Fig. 9. MN movement pattern for UMTS/802.11b-integrated networks.

region and 802.11b region. The MN moves at a constant velocity of 10 m/s (or 36 km/h). The time for MN to reach the overlapped region is approximately 7.07 s and the time taken in the overlapped region is approximately 7.07 s. In this heterogeneous environment, we also targeted a semi-vehicular speed node in a city area to simulate highly movement-oriented environment.

#### 5.4.4. Traffic

For all the simulation scenarios, the same FTP traffic is used. FTP data packets are transmitted for 600 s (from 30 s to 630 s period). We gave initial 30-s period to obtain stable link status before transmission.

## 6. Results in UMTS/802.11B-Integrated Networks

In this section, we evaluate the performance of MIP and mSCTP handoff by comparative analysis based on the simulation results in UMTS/802.11b integrated networks.

### 6.1. Uplink Behaviors

Figure 10 shows an example of the uplink (MN to CN) transmission behavior (with five handoffs in the unit time, 10-min) of the three protocol stacks in UMTS/802.11b-integrated networks. The x-axis in

Figure 10 represents the simulation duration, and the y-axis represents the number of total bits received at CN. Each line of graph shows the transmission bit rate of each handoff protocol. Five handoff delay periods are recognized in both MIP+TCP and MIP+SCTP protocol stacks. Although MIP+SCTP CN received more data than that of MIP+TCP, both protocol stacks show inherent handoff delay of MIP due to its registration period. On the other hand, mSCTP handoff shows the highest transmission rate without any delayed period.

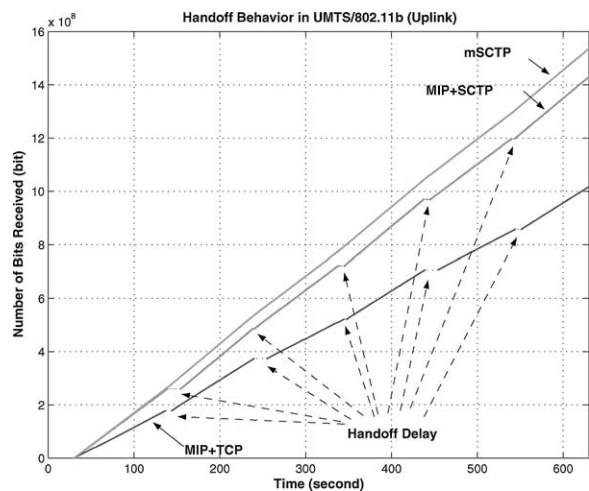


Fig. 10. Handoff behavior in UMTS/802.11b-integrated networks (Uplink): SIM-N2-P $\beta$ -D1.

This is because mSCTP uses multi-homing and DAR procedure, which provides ASCONF chunk bundling in ordinary data packets (see Subsubsection 2.2.2). In Figure 10, it should be noted that the data rate varies during the transmission period. According to the position of MN, whether it is in UMTS Node-B coverage or in 802.11b AP coverage, MN utilizes the underlying network data rates.

### 6.1.1. Total handoff delay

We measured the total accumulated handoff delay during the simulation duration. Handoff delay in MIP includes registration period while mSCTP requires DAR procedure (see Subsection 2.2.2). As mSCTP DAR procedure sends and receives at most three pairs of ASCONF and ASCONF-ACK control chunks (256-bit per each pair) bundling with the other data chunks, the handoff delay can be neglected although it incurs a 768-bit throughput decrease per handoff. Figure 11 shows the uplink (MN to CN) handoff delay of MIP+TCP, MIP+SCTP, and mSCTP over handoff rate.

Total handoff delays versus handoff rate shows how the handoff delay of each handoff protocol reacts when scale of mobility varies. In Figure 11, the total handoff delays of MIP+TCP and MIP+SCTP linearly increase as expected from our analysis in Subsection 4.1.1 as shown in Equations (4) and (5). In contrast, mSCTP handoff does not incur any delay irrespective of the handoff rate. This is due to the fundamental difference between MIP handoff registration procedure and mSCTP DAR procedure. That is, mSCTP can trigger

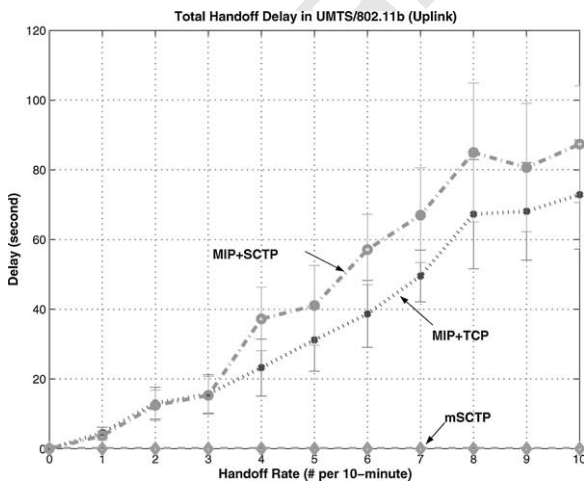


Fig. 11. Total handoff delay in UMTS/802.11b-integrated networks (uplink): SIM-N2-P $\beta$ -D1 (95% confidence interval).

DAR procedure by bundling ASCONF and ASCONF-ACK chunks with the other data chunks.

The result shows that the handoff delay of two MIP protocol stacks becomes more significant as handoff rate increases. As we discussed in Subsections 4.2.1 and 4.3.1, handoff delay and handoff rate product directly affects the end-to-end throughput and packet loss. Thus, MIP cannot be a proper handoff approach in large scale mobility environments. On the other hand, mSCTP does not affect any significant throughput decrease nor packet loss by keeping handoff delay zero regardless of handoff rate.

### 6.1.2. End-to-end throughput

In the analysis in Subsection 4.2, we concluded end-to-end throughput of MIP decreases directly proportional to the product of handoff delay and handoff rate while that of mSCTP decreases at most 768-bit per handoff in Equations (22) and (23). In the simulation, each scenario has been run with ten different handoff rates (0 to 10 handoff occurrences in 10-min unit time). The uplink end-to-end throughput has been calculated as the total number of bits received at CN divided by the transmission duration. Figure 12 shows the uplink (MN to CN) end-to-end throughput in bit-per-second (bps) over the handoff rate. For all the three handoff protocol stacks, the end-to-end throughput with 1-handoff is increased from the case of zero-handoff. This is due to the fact that the MN utilize the higher bandwidth of 802.11b from 1-handoff scenarios (see *MN movement* in Subsection 5.4). This result shows the benefit of the network convergence paradigm we discussed in

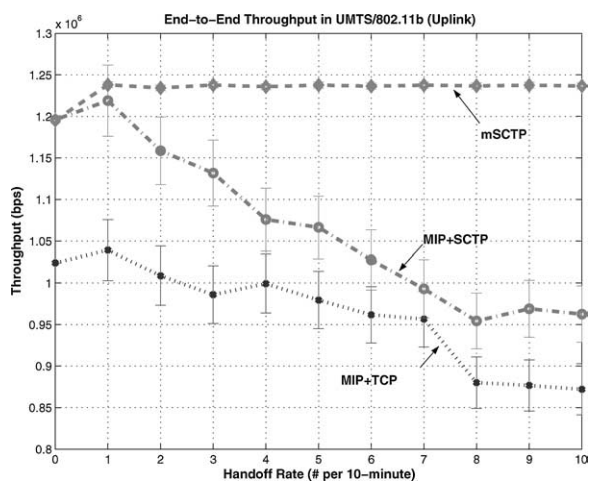


Fig. 12. End-to-end throughput in UMTS/802.11b-integrated networks (uplink): SIM-N2-P $\beta$ -D1 (95% confidence interval).

Subsection 3.1. However, the end-to-end throughput of MIP+TCP and MIP+SCTP still decreases as handoff rate increases while that of mSCTP maintains consistent values regardless of the handoff rate. First, the end-to-end throughput of MIP+TCP, without any handoff, is about 1.024 Mbps. It is the maximum data rate that MIP+TCP can obtain in UMTS coverage. The data rate of MIP+TCP goes up to 1.039 Mbps in 1-handoff scenario, but goes down to 872 Kbps when 10 handoffs occur in the unit time, 10 min. It shows that the handoff delay and handoff rate product affects the degradation of end-to-end throughput in spite of the benefit of the network integration.

Second, MIP+SCTP produces about 1.196 Mbps end-to-end throughput when no handoff occurs. The end-to-end throughput of MIP+SCTP reaches 1.219 Mbps in 1-handoff scenario due to the bandwidth benefit of 802.11b network. However, it goes down to approximately 954 Kbps with 8 handoffs. It shows MIP+SCTP also cannot take advantage of the benefit of network integration when handoff rate increases.

The reason why MIP+SCTP produces better basic transmission rate than that of MIP+TCP is that SCTP uses SACK algorithm as well as multi-chunk message-based stream with enhanced congestion control mechanism. Unlike TCP, SCTP does not mandatorily require in-order data arrival in packet sequence level but in each chunk level. So, all the in-order data chunks can be handed over to the upper layer protocol in the stack irrespective of the fact whether there was any out-of-order data chunk.

Finally, the end-to-end throughput of mSCTP is 1.195 Mbps without any handoff in the UMTS coverage. With 1-handoff, it goes up to 1.238 Mbps with the benefit of 802.11b bandwidth. Unlike MIP handoffs, the end-to-end throughput of mSCTP is maintained consistently irrespective of the handoff rate (except zero-handoff scenario). This means that mSCTP can exploit the higher data rate of 802.11b, without any significant loss of data.

As shown in Figure 12, mSCTP produces very stable and higher end-to-end throughput comparing to MIP+TCP and MIP+SCTP, regardless of the handoff rate. That is, mSCTP, the transport layer handoff can be much more efficient, in terms of throughput, than MIP in large scale mobility environments.

### 6.1.3. Packet loss

To measure packet loss, we counted the total number of data packets lost during total handoff delay periods. The packet loss over the handoff rate shows how the

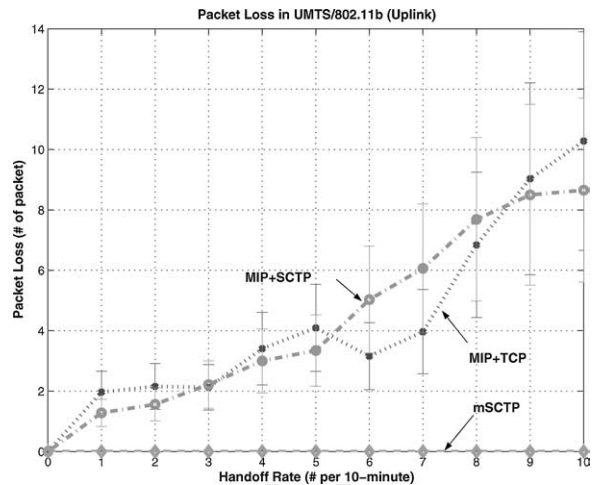


Fig. 13. Packet loss in UMTS/802.11b-integrated networks (uplink): SIM-N2-P $\beta$ -D1 (95% confidence interval).

handoff protocols react under the variation of the intensity of MN mobility. Especially, data packet losses can trigger congestion control of transport layer protocol, and thus, affect the quality of service. Figure 13 shows the uplink (MN to CN) packet loss in number of packets lost over the handoff rate.

Figure 13 shows that the number of packet loss of MIP+SCTP and MIP+TCP increases proportional to handoff rate while that of mSCTP does not increase and remains to be zero as in Equations (31) and (32). First, the number of packet loss in MIP+TCP increases quite linearly showing a little fluctuation with 10.281 highest in 10-handoff scenario. Second, the number of packet loss of MIP+SCTP also has a little fluctuation increasing up to 8.656 at 10-handoff scenario. Finally, mSCTP maintains zero packet loss irrespective of handoff rate. This is the result from the fact that the multi-homing of SCTP and DAR procedure makes no handoff delay (see Subsection 4.3.2). As we have seen in Figure 11, the handoff delay of MIP+TCP and MIP+SCTP shows identical relationship with the packet loss. That is, the number of packet loss in MIP+TCP and MIP+SCTP is directly affected by the total handoff delay periods.

## 6.2. Downlink Behaviors

Generally, it is known that downlink traffic overwhelms uplink traffic in terms of the proportion of the application level services. Figure 14 is an example transmission behavior (with five handoffs in the unit time) of the three handoff protocol stacks for downlink (CN to MN) traffic in UMTS/802.11b-integrated networks.

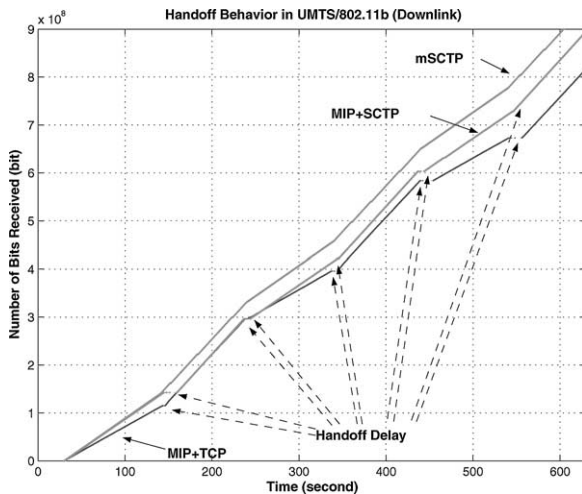


Fig. 14. Handoff behavior in UMTS/802.11b-integrated networks (downlink).

Like in the uplink transmission, the x-axis in Figure 14 is the simulation time and the y-axis is the number of total bits received at MN. The result shows the variation of data rate during the transmission. All the three handoff protocol stacks take advantage of the network integration in Figure 14. The bit rate of mSCTP is higher than that of MIP+TCP and MIP+SCTP. With MIP+TCP and MIP+SCTP protocols, we can notice five handoff delay periods during the transmission.

In the downlink transmission (CN to MN), mSCTP MN triggers DAR procedure with three pairs (add-IP, set-primary-IP, and delete-IP) of ASCONF and ASCONF-ACK control chunks cooperating with CN. After receiving an add-IP ASCONF request, CN adds the received IP address as a new IP address in the existing association and responds with ASCONF-ACK. In turn, MN sends a set-primary-IP chunk to let CN switch the transmission to the new IP address. Upon the reception of the set-primary-IP ASCONF chunk, CN switches the destination address to the requested IP address and respond back to MN with ASCONF-ACK. Delete-IP can also be performed if it is preferable. We implemented this scenario using *set-primary-destination* method provided by NS-2 SCTP module. Now, we analyze the performance of the three handoff protocols in terms of total handoff delay, end-to-end throughput, and packet loss.

### 6.2.1. Total handoff delay

Figure 15 shows the downlink (CN to MN) handoff delay (in second) over the handoff rate.

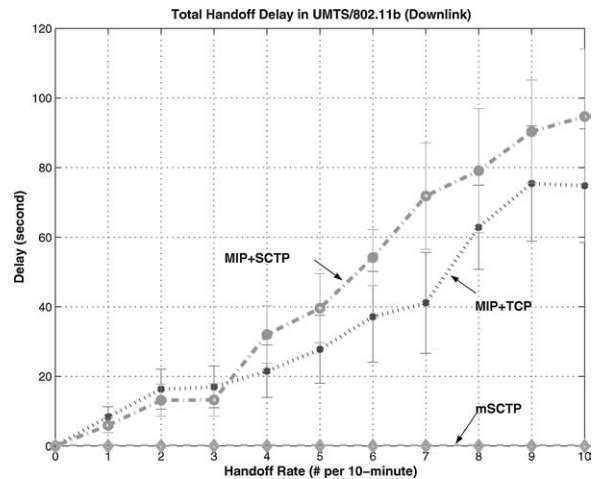


Fig. 15. Total handoff delay in UMTS/802.11b-integrated networks (downlink) (95% confidence interval).

In downlink case, the total handoff delay of MIP+TCP and MIP+SCTP increase proportional to the handoff rate while that of mSCTP keeps zero irrespective of the handoff rate. With the same reason with uplink case, MIP handoff brings about the unavoidable registration delay. On the contrary, mSCTP performs DAR procedure using at most three pairs of ASCONF and ASCONF-ACK control chunks (256-bit per each pair) bundling with the other data chunks in the existing transmission.

In addition, the total handoff delay of MIP+TCP and MIP+SCTP shows quite an identical trend although the difference is not perfectly identical. This is because, as we discussed in the uplink case, the registration procedure of MIP should be processed regardless of the transport layer protocols used.

### 6.2.2. End-to-end throughput

Figure 16 shows the downlink (CN to MN) end-to-end throughput in bit-per-second (bps) over the handoff rate.

As in the uplink case, the end-to-end throughput of all the three handoff protocol stacks shows the highest value from all the different handoff rate scenarios. This is because MN initially stays in UMTS region in the zero-handoff scenario. From the 1-handoff scenario to 10-handoff scenario, MN takes the benefit of 802.11b network bandwidth whenever it raises handoff into 802.11b region. However, in spite of the benefit of network integration, the throughput of MIP+TCP and MIP+SCTP decreases as the handoff rate increases

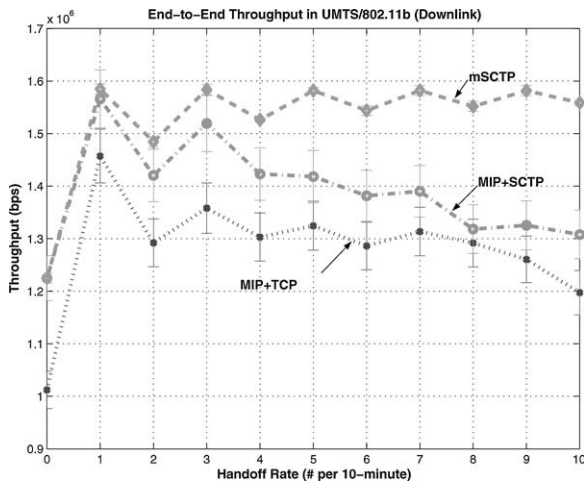


Fig. 16. End-to-end throughput in UMTS/802.11b-integrated networks (downlink) (95% confidence interval).

while that of mSCTP maintains quite consistent values irrespective of the handoff rate.

First, the basic throughput without any handoff in MIP+TCP is about 1.012 Mbps. In 1-handoff scenario, the throughput of MIP+TCP goes up to 1.457 by utilizing the higher bandwidth of 802.11b network. However, it goes down to 1.197 Mbps when 10 handoffs occur within the 10-min unit time. Second, MIP+SCTP produces about 1.225 Mbps throughput when no handoff occurs. As it is in MIP+TCP, the end-to-end throughput of MIP+SCTP goes up to 1.566 Mbps in 1-handoff scenario, but, again, goes down to approximately 1.308 Mbps with 10 handoffs. Although MIP+SCTP produces little bit higher throughput than MIP+TCP, both of two MIP handoff protocol stacks incur unavoidable throughput decrease.

Last, the end-to-end throughput of mSCTP is 1.224 Mbps without any handoff. The throughput of mSCTP reaches 1.585 Mbps in 1-handoff scenario by exploiting 802.11b bandwidth. Then, mSCTP handoff maintains about the same data rates as in 1-handoff scenario with little fluctuations. This means that mSCTP can utilize the benefit of network integration without any significant data loss. mSCTP has only 768-bit DAR procedure overhead per handoff. The accumulated DAR procedure overhead decreases the end-to-end throughput of mSCTP, but the amount of decrease is insignificant comparing to that of MIP. Here, we should notice that the DAR procedure does not incur any delay at all while generating the insignificant throughput decrease.

For all the three handoff protocols, the throughput of 10-handoff scenario is even higher than that

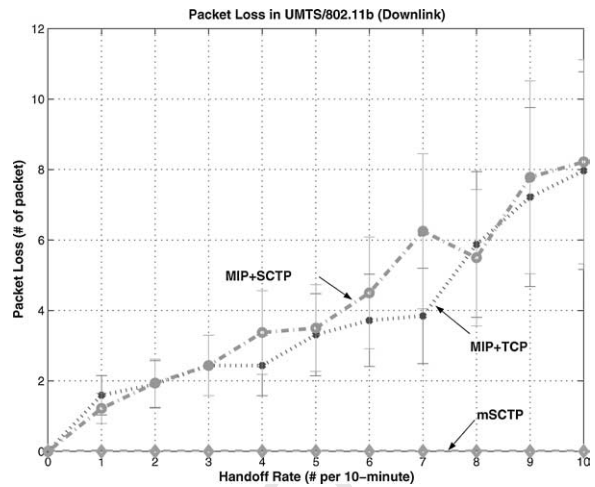


Fig. 17. Packet loss in UMTS/802.11b-integrated networks (downlink) (95% confidence interval).

of zero-handoff scenario in which MN stays only in UMTS region. This shows the benefit of network integration. However, the throughput of MIP+TCP and MIP+SCTP still decreases significantly comparing to that of mSCTP.

As shown in Figure 16, mSCTP produces quite stable and higher end-to-end throughput over MIP+TCP and MIP+SCTP in the downlink case as well. The effect of mSCTP becomes very apparent when the handoff rate increases. mSCTP shows far better end-to-end transmission throughput in highly handoff intensive environment by exploiting its multi-homed seamless handoff mechanism.

### 6.2.3. Packet loss

Figure 17 shows the downlink (CN to MN) packet loss in number of packets lost over the handoff rate.

Figure 17 shows that the number of packet loss of MIP+SCTP and MIP+TCP increases proportional to handoff rate while that of mSCTP does not increase and maintain zero. First, the number of packet loss in MIP+TCP increases quite linearly showing a little fluctuation with the highest value of 7.968 in 10-handoff scenario. Second, the number of packet loss of MIP+SCTP also has a little fluctuation increasing up to 8.219 at 10-handoff scenario. Finally, mSCTP maintains zero packet loss irrespective of handoff rate. This is the result from the fact that the multi-homing of SCTP and DAR procedure makes no handoff delay (see Subsection 4.3.2). As we have seen in Figure 15, the handoff delay of MIP+TCP and MIP+SCTP shows identical relationship with the packet loss. That is, the

number of packet loss in MIP+TCP and MIP+SCTP is directly affected by the total handoff delay periods.

In this section, we presented our simulation system model and evaluated the performance of MIP handoff and mSCTP handoff based on the result of the simulation. mSCTP handoff, having no significant delay and packet loss, produces the highest and consistent end-to-end throughput regardless of handoff rate. It is worthy to mention that the analysis in Section 4 is mainly for uplink scenarios because we consider the delay, throughput, and packet loss for mobile users, but not corresponding users. However, our analytical results for uplink can also be generalized to downlink scenarios with minor modification in counting signaling.

### 6.3. Further Discussion

#### 6.3.1. Remark

In Subsections 6.1 and 6.2, we have discussed the transmission behaviors of three handoff protocol stacks in UMTS/802.11b-integrated networks. We already addressed the advantages of mSCTP over MIP protocol stacks. Apart from the performance difference between MIP and mSCTP, we should also bring our attention to performance improvement coming from SCTP protocol. From Figures 11, 12, 15, and 16, it should be noticed that MIP+SCTP produces higher throughput than MIP+TCP while total handoff delay of MIP+SCTP is larger than that of MIP+TCP in certain handoff rate ranges. It is contradictory to our expectation. For instance, in the 4-handoff rate case in Figure 15, the difference of total handoff delay between MIP+TCP and MIP+SCTP is about 32.8%. MIP+TCP shows even better performance in terms of handoff delay in this case. However, the throughput difference between MIP+TCP and MIP+SCTP in the 4-handoff rate case in Figure 16 shows MIP+SCTP has 8.4% throughput improvement over MIP+TCP. This implies that SCTP protocol is a more proper transport layer protocol than TCP in mobile wireless environments. Thus, mSCTP, inherently adopting SCTP as a transport protocol, can perform more efficient data transmissions than any MIP approach at the moment.

#### 6.3.2. Impact of handoff rate in the simulation

Throughout the simulation study, handoff rate has been adopted as an important variable to evaluate the different handoff protocols with regard to different scale of mobility.

Although the accumulation of handoff delays is supposed to be identical to the total handoff delay generated by corresponding handoff rate, the simulation results showed that the relationship between the two values is not perfectly linear. This observation can be attributed to the fact that the actual performance degradation of handoff protocols in different scale of mobility comes from various reasons including node movement, possible randomness from underlying network components, and so on. Thus, our performance evaluation approach with handoff rate can be considered to be at least more approximate to experimental results than that of linear accumulation approach.

## 7. Conclusions

In this paper, we conducted a performance evaluation of MIP and mSCTP with regard to handoff delay, end-to-end throughput, and packet loss, followed by the corresponding simulation study in heterogeneous network environment. MIP handoff incurs unavoidable handoff delay due to agent discovery and registration periods while mSCTP handoff delay can be neglected since it exploits multi-homing and DAR procedure. It should be noted that total handoff delay of MIP increases as handoff rate increases while that of mSCTP is being kept zero. End-to-end throughput of MIP decreases directly proportional to handoff rate and handoff delay product while mSCTP throughput is maintained constant without any significant data loss irrespective of handoff rate. We also showed that MIP handoff incurs certain packet loss according to its handoff delay. mSCTP, on the other hand, does not incur any packet loss during its handoff period. To conclude, mSCTP provides quasi-seamless IP-based handoff regardless of scale of mobility in heterogeneous network environments. As a promising IP-based handoff solution in future wireless IP convergent networks, mSCTP needs more attention in both academic and industry research society.

## References

1. Ramjee R, Varadhan K, Salgarelli L, Thuel SR, Wang S-Y, Porta TL. Hawaii: a domain-based approach for supporting mobility in wide-area wireless networks. *IEEE/ACM Transactions on Networking* 2002; **10**(3): 396–410.
2. Campbell AT, Gomez J, Kim S, Valko AG, Wan C-Y, Turanyi ZR. Design, implementation, and evaluation of cellular ip. *IEEE Personal Communications* 2000; **7**(4): 42–49.
3. Misra A, Das S, Dutta A, McAuley A, Das SK. Idmp-based fast handoffs and paging in ip-based 4g mobile networks. *IEEE Communications Magazine* 2002; **40**(3): 138–145.

4. Stewart R, Xie Q, Morneault K, Sharp C, Schwarzbauer H, Taylor T, Rytina I, Kalla M, Zhang L, Paxson V. Stream control transmission protocol: <http://www.ietf.org/rfc/rfc2960.txt?number=2960>, 2000.
5. Stewart R, Ramalho M, Xie Q, Tuexen M, Conrad P. Stream control transmission protocol (sctp) dynamic address reconfiguration: <http://www.ietf.org/internet-drafts/draft-ietf-tsvwg-addip-sctp-11.txt>, 2005.
6. The network simulator—ns-2: <http://www.isi.edu/nsnam/ns>.
7. Soliman H. *Mobile ipv6: Mobility in a Wireless Internet*. Addison Wesley, 2004<sup>Q5</sup>.
8. Ma L, Yu F, Leung VCM, Randhawa T. A new method to support UMTS/WLAN vertical handover using SCTP. *IEEE Wireless Communications* 2004; **11**(4): 44–51.
9. Fu S, Atiquzzaman M, Ma L, Ivancic W, Lee Y-J, Jones JS, Lu S. TraSH: A transport layer seamless handover for mobile networks. Technical report technical report, Univ. of Oklahoma, November 2003.
10. Perkins C, Johnson DB. Route optimization in mobile ip: <http://k-lug.org/griswold/drafts-rfcs/draft-ietfmobileip-optim-11.txt>, 2001.
11. Johnson DB, Perkins C, Arkko J. *Mobility Support in IPv6*. Internet Engineering Task Force, June 2004. available here: <http://www.ietf.org>.
12. Gustafsson E, Jonsson A, Perkins CE. Mobile ipv4 regional registration: <http://www.ietf.org/proceedings/04aug/i-d/draft-ietf-mobileip-reg-tunnel-09.txt>, 2004.
13. Soliman H, Catelluccia C, Malki KE, Bellier L. Hierarchical mobile ipv6 mobility management (hmip6): <http://www.ietf.org/proceedings/04aug/i-d/draft-ietf-mipshop-hmip6-02.txt>, 2004.
14. Akyildiz IF, Xie J, Mohanty S. A survey of mobility management in next-generation all-ip-based wireless systems. *IEEE Wireless Communications* August 2004; **11**(4): 16–28.
15. Stewart RR, Xie Q. *Stream Control Transmission Protocol (SCTP): A Reference Guide*. Addison Wesley, 2002<sup>Q5</sup>.
16. Fu S, Atiquzzaman M. Sctp: State of the art in research, products, and technical challenges. *IEEE Communications Magazine* 2004; **42**(4): 64–76.
17. Stewart R, Metz C. Sctp: New transport protocol for tcp/ip. *IEEE Internet Computing* 2001; **5**(6): 64–69.
18. Smith C, Meyer J. *3g wireless with wimax and wi-fi*. McGraw-Hill, 2005<sup>Q5</sup>.
19. Bannister J, Mather P, Coope S. *Convergence Technologies for 3g Networks: IP, UMTS, EGPRS and ATM*. Wiley, 2004<sup>Q5</sup>.
20. Buddhikot M, Chandranmenon G, Han S, Lee YW, Miller S, Salgarelli L. Integration of 802.11 and Third-Generation Wireless Data Networks. In *Proceedings of IEEE Infocom 2003*, volume 1, March 2003, 503–512.
21. Kaaranen H, Ahtiainen A, Laitinen L, Naghian S, Niemi V. *Umts Networks: Architecture, Mobility and Services*. Wiley, 2001<sup>Q5</sup>.
22. Perkins C. Ip mobility support: <http://www.ietf.org/rfc/rfc2002.txt?number=2002>, 1996.
23. Deering S. Icmp router discovery messages: <http://www.ietf.org/rfc/rfc1256.txt?number=1256>, 1991.
24. Ns-2 sctp module: <http://pel.cis.udel.edu>, 2005.
25. Ns-2 network simulator release all-in-one 2.26: <http://www.isi.edu/nsnam/dist/ns-allinone-2.26.tar.gz>.

## Authors' Biographies



**Wenye Wang** (M'98/ACM'99) 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. degree from 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. Dr. Wang is a recipient of NSF CAREER Award in 2006. She has been a member of the Association for Computing Machinery since 2002.



**Jung Kee Song** received his Bachelors degree in Computer Science from Kwangwoon University, Seoul, Korea, in February 1999. He had worked for DreamData, Inc., Seoul, Korea, as a software engineer from November 1998 to May 2002. He has been a Masters student in the Department of Computer Science at North Carolina State University since August 2003. He will work for Samsung Electronics Co., Ltd., Telecommunication Network Business, as an engineer, starting Fall 2005.

## Author Query Form (WCM/415)

**Special Instructions: Author please write response to queries directly on Galley proofs and then fax back. Alternatively please list response in an e-mail.**

Q1: Author: Please check the suitability of the suggested short title.

Q2: Author: Please provide keywords.

Q3: Author: Please check the sentence for correctness.

Q4: Author: Please check symbol (?) for correctness.

Q5: Author: Please provide the publisher location.

UNCORRECTED PROOFS