5-2006

On secure communication in integrated internet and heterogeneous multi-hop wireless networks.

Bin Xie 1970-
University of Louisville

ON SECURE COMMUNICATION IN INTEGRATED INTERNET AND HETEROGENEOUS MULTI-HOP WIRELESS NETWORKS

By

Bin Xie
B.S., Central South University, 1995
M.S., University of Louisville, 2003

A Dissertation
Submitted to the Faculty of the
Graduate School of the University of Louisville
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

Computer Science and Engineering
University of Louisville
Louisville, Kentucky

May 2006

ON SECURE COMMUNICATION IN INTEGRATED INTERNET AND MULTI-HOP WIRELESS NETWORKS

By

Bin Xie
B.S., Central South University, 1995
M.S., University of Louisville, 2003

A Dissertation Approved on

April 13, 2006

by the following Dissertation Committee:

Advisor – Anup Kumar

S. Srinivasan, Ph.D.

Rammohan K. Ragade, Ph.D.

Xiangqian, Liu, Ph.D.

Ahmed H. Desoky, Ph.D.

Dar-Jen Chang, Ph.D.

DEDICATION

This dissertation is dedicated to my parents

Mr. Changsxiang Xie

and

Mrs. Cuiyun Gong

who have given me invaluable educational opportunities.

ACKNOWLEGEMENTS

This dissertation is the result of the collaborative efforts of numerous individuals who have directly or indirectly supported me in my doctorial dissertation journey. The dissertation would have been impossible without help from my supervisor and mentor, Dr. Anup Kumar. I would like to acknowledge him for his tremendous support in facilitating my graduate journey, and for his wisdom and guidance in inspiring me. I gratefully acknowledge Dr. S. Srinivasan for his encouragement and time-comsuming dedication in my research. I would like to express my sincerest gratitude to members of my committee, Rammohan K. Ragade, Xiangqian Liu, Dar-Jen Chang, and Ahmed H. Desoky, for their time and effort in reviewing my work, and for providing pertinent and valuable advice on my research. It has been a great privilege to spend several years in the Speed School surrounding by their intelligence and humanity.

I would also like to fully acknowledge the supports from Grosscurth Fellowship, NSF Grant 0416900, CECS department, and the assistance from Dr. Dharma P. Agrawal, ORB Center for Distributed and Mobile Computing Department of ECECS, University of Cincinnati.

In addition, I would like to thank Pei Zhang, my soul mate, for her love, understanding, unselfish support, and constant encouragement. Finally, I am forever grateful to my parents, for all of their unconditional support, faith, and love throughout this dissertation and my life. I wouldn't be who I am today without them.

ABSTRACT

ON SECURE COMMUNICATION IN INTEGRATED INTERNET AND HETEROGENEOUS MULTI-HOP WIRELESS NETWORKS

Bin Xie

April 13, 2006

Integration of the Internet with a Cellular Network, WMAN, WLAN, and MANET presents an exceptional promise by having co-existence of conventional WWANs/WMANs/WLANs with wireless ad hoc networks to provide ubiquitous communication. We call such integrated networks providing internet accessibility for mobile users as heterogeneous multi-hop wireless networks where the Internet and wireless infrastructure such as WLAN access points (APs) and base stations (BSs) constitute the backbone for various emerging wireless networks (e.g., multi-hop WLAN and ad hoc networks).

Earlier approaches for the Internet connectivity either provide only unidirectional connectivity for ad hoc hosts or cause high overhead as well as delay for providing full bi-directional connections. In this dissertation, a new protocol is proposed for integrated Internet and ad hoc networks for supporting bi-directional global connectivity for ad hoc hosts.

In order to provide efficient mobility management for mobile users in an integrated network, a mobility management protocol called multi-hop cellular IP (MCIP) has been proposed to provide a micro-mobility management framework for heterogeneous multi-hop network. The micro-mobility is achieved by differentiating the local domain from global domain. At the same time, the MCIP protocol extends Mobile IP protocol for providing macro-mobility support between local domains either for single hop MSs or multi-hop MSs. In the MCIP protocol, new location and mobility management approaches are developed for tracking mobile stations, paging, and handoff management.

This dissertation also provides a security protocol for integrated Internet and MANET to establish distributed trust relationships amongst mobile infrastructures. This protocol protects communication between two mobile stations against the attacks either from the Internet side or from wireless side. Moreover, a secure macro/micro-mobility protocol (SM^{3}P) have been introduced and evaluated for preventing mobility-related attacks either for single-hop MSs or multi-hop MSs. In the proposed SM^{3}P, mobile IP security has been extended for supporting macro-mobility across local domains through the process of multi-hop registration and authentication. In a local domain, a certificate-based authentication achieves the effective routing and micro-mobility protection from a range of potential security threats.

TABLE OF CONTENTS

PAGE

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

FIGURE PAGE

CHAPTER I

INTRODUCTION AND BACKGROUND

Recent advances in communication and computing technologies clearly indicate that future wireless networking environments will be heterogeneous in terms of diversified access technologies [1]. While some wireless technologies are already part of our everyday lives, such as Wireless WANs (e.g., 2G, 2.5G and 3G cellular systems), Wireless MANs (e.g., IEEE 802.16), WLANs (e.g., IEEE 802.11a/b/e/g), and Wireless PANs (e.g., Bluetooth), the emerging IEEE 802.15.3/4 WPANs and IEEE 802.22 technologies are expected to provide even more exciting and efficient services. The available services would include voice, multimedia, messaging, e-mail, information services (e.g., news, stocks, weather, and travel), M-commerce, entertainment, location-based public utility and health-care services, and so on. Availability of a multitude of wireless technologies has motivated research efforts towards the next generation of wireless and mobile communication systems, usually called "Beyond 3G" (B3G) or "4G" networks that are expected to have integrated heterogeneous wireless networks that would enable users equipped with either multi-interface or SDR-based wireless devices to use the appropriate network that meets their service levels and cost requirements, in a transparent and seamless way. In heterogeneous wireless networks, different types of network have distinctive underlying features, for example, WLAN provides higher bandwidth. On the contrary, the cellular system can only support low bandwidth.

However, the cellular network can cover larger area, and can possibly provide worldwide seamless connection. Also, integrating ad hoc networks with infrastructure networks can be beneficial for extending the service coverage of wireless networks via multi hop communication. In this case, wireless access points or base stations (APs/BSs) employ different radio spectrums to provide Internet accessibility for mobile stations (MSs) or wireless networks (e.g., a MANET network) to access the Internet. An ad hoc MS outside the coverage of any BS or AP may obtain the Internet services by way of multi-hop relaying. Ad hoc networks can be used to extend the coverage of not only the cellular networks, but also the high frequency Wireless LANs as well as other wireless networks [2]. This phenomenon is totally different from a pure mobile ad hoc network where a self-configurable network capability is provided by communication among ad hoc MSs without any Internet-based infrastructure (e.g., BS/AP) for mobility and security support. Once attached to the Internet, the MANET communication is not isolated anymore, but is an integral part of the infrastructure-based networks for Internet access. In such an integrated network, a given MS of the ad hoc part needs to relay data packets hop by hop before they can reach a BS or another mobile station having access to a BS, which acts as a communication bridge between wireless network and the Internet.

Our investigation in Chapter 2 shows that an integrated heterogeneous multi-hop network may be attacked from the Internet or wireless side for the purpose of endangering the MANET and other multi-hop communications. Firstly, the potential security vulnerability of the wireless infrastructure may impede the Internet accessibility of such wireless networks, thus destroy the civilian or other applications from these wireless networks (e.g., an ad hoc network). Secondly, mobility-related attacks may occur in an integrated network and result in that a MS cannot find a correct BS/AP for Internet connection. Thirdly, a communication route may be distorted by an adversary with a routing attack. Because of these, infrastructure and information protection for secure communication is a crucial design issue in integrated heterogeneous multi-hop network. The security design and implementation should take the heterogeneous communication path into account, which could span the Internet, cellular networks, WLAN, and MANET. Therefore, many security issues in integrated networks should be reconsidered in this integrated environment rather than only as an individual cellular network, WLAN or MANET. It is obvious [64] that the current authentication protocols of UMTS, IEEE 802.16 and IEEE 802.11 security are not applicable to the integrated networks because these protocols are developed based on the assumption that a MS can connect to a BS directly and exchange authentication packets. Since the internet mobility protocol (i.e., mobile IP [33] [34]) in the conventional cellular and WLAN lacks security support for a multi-hop MS, new security features should be developed for the multi-hop MS to access the network with mutual authentication at every hop. In the integrated networks, possible non-cooperative behavior in MANET may significantly and adversely affect the performance of network in the case of multi-hop packet relaying. Therefore, collaborative protocols are needed in the integrated network to encourage cooperation during multi-hop packet relay.

The design of such integrated Internet and heterogonous multi-hop wireless networks involves two fundamental problems with many sub-issues for each problem, which are addressed in this dissertation.

- **Problem 1:** Secure communication for integrated Internet and MANET

A. Integrated protocol for Internet and ad-hoc network communication

B. Secure interconnection protocol for integrated internet and ad-hoc networks

- **Problem 2:** Secure mobility management in heterogeneous multi-hop wireless networks

  A. Macro/Micro-mobility approach to heterogeneous wireless networks

  B. Secure Macro/Micro-mobility protocol

The following subsections give an overview of significant contribution made by this dissertation and describe the background as well as related work regarding the two problems in heterogonous multi-hop wireless networks.

1.1 Secure Communication for Integrated Internet and MANET

The dissertation in part A addresses the development of secure communication framework for integrated Internet and MANET. In an integrated Internet and MANET [3-10], each ad hoc MS executes a MANET routing protocol to construct a communication path between two network stations (source and destination). The basic process of ad hoc route discovery such as AODV [16] and DSR [17], involves two messages: route request and route reply. In the beginning, the source MS broadcasts a route request if it has no route to a destination. The route request will be forwarded by intermediary node until it reaches the destination. Then the destination station (e.g., a BS or ad hoc MS) responses a route reply message carrying the reversed route. The process of ad hoc route discovery enables the ad hoc MS with the capability to reach a FA. Through the established route to the FA, a separate protocol [3-10] coordinates mobile IP [33] [34] and the MANET routing protocol to obtain the Internet connectivity for ad hoc MSs. Ad hoc routing protocols can be coarsely divided into two groups: proactive and on-demand routing protocols. In the proactive routing protocols, each node keeps one or more tables that contain the routing information about the other nodes. The tables are created and updated through network routing control messages. The routing table in each MS maintains consistent and up-to-date routing information. This kind of protocols includes Destination-Sequence Distance-Vector (DSDV) [11], Wireless Routing Protocol (WRP) [12], Global State Routing (GSR) [13], Hierarchical State Routing (HSR) [14], Fisheye State Routing (FSR) [21], and Zone-based Hierarchical Link State (ZHLS) [15], etc. On the other hand, in an on-demand routing protocol, each node creates a route when communication is needed. When a MS in a MANET requires a route to a destination, the host initiates a route discovery process. The protocols in this category are Ad Hoc On-demand Distance Vector (AODV) [16], Cluster Based Routing Protocol (CBRP) [18], Dynamic Source Routing (DSR) [17], Temporally Order Routing Algorithm (TORA) [19], Associativity Based Routing (ABR) [20], and Signal Stability Routing (SSR) [22]. A detailed survey and performance comparisons of the above ad hoc protocols can be found in papers [23][24].

The approaches discussed in [3-10] address the work of integrating MANETs with wired networks by extending mobile IP. In the approach [5], the MSs run a modified Routing Information Protocol (RIP) Daemon (*routed*) for dealing with mobile IP messages. Also a mobile IP routing daemon (*mipd*) runs on each MS. The *routeds* of the mobile hosts under the coverage of a FA receive and handle FA's agent advertisements and related mobile IP messages. These *routeds* relay the mobile IP messages to the *mipds* of the MSs that are outside the transmission range of the FA. The MIPMANET [7] provides the global Internet connectivity by using mobile IP with FA's care-of-address and reverse tunneling. A FA periodically broadcasts its *advertisement.* The *advertisement* message spreads through the whole ad hoc network. This approach uses AODV protocol for communications in ad hoc network. The performance of this approach has been improved by more recent approaches given in papers [8] [25]. The approach used in [25] extends mobile IP to ad hoc networks using the on-demand AODV protocol; therefore each mobile node registers with a FA only when it requires the global Internet connectivity. The paper [8] further improves the performance of the approach given in [25] by reducing the mobile IP overhead. This approach minimizes the number of FA advertisements by using controlled flooding in the ad hoc network.

In Chapter 3, the dissertation proposes an integrated protocol for internet and ad hoc network communication. Compared to the aforementioned integrated protocols, my protocol has four advantages (i) bi-directional internet connection for ad hoc host that means the Internet (i.e., BS) and each ad hoc host can reach each other all time, (ii) less control overhead in maintaining the bi-directional connection for ad hoc hosts, (iii) high throughput communication for ad hoc host to access the internet, (iv) less delay for packet forwarding.

Since protocols mentioned above are based on a benign environment, these protocols may suffer from security threats in an adversarial environment. Security of ad hoc routing protocols has become an important issue. The Secure Routing Protocol (SRP) [26] uses message authentication codes (HMAC) with pairwise shared keys to ensure the packet forwarding between any two MSs without violating the routing protocol. The computation of HMAC with a shared secret symmetric key between two MSs is very efficient. SRP initiates a trustworthy relationship between sender and destination. Both the Secure Efficient Distance Vector Routing Protocol (SEAD) [27] and the ARIADNE [28] are one-way hash MAC based security protocols. The authentication of routing message in SEAD and ARIADNE depends on TESLA [29] hash-chain-based protocol for key management. The SEAD is an ad hoc secure routing protocol that uses efficient one–way hash functions to guarantee the routing safety on DSDV [11]. With the assumption of using a trusted node to sign public keys for each node and to distribute those keys, SEAD authenticates the sequence number and metric of a route table by using hash chain elements. ARIADNE [28] copes with the security in on-demand routing protocols by point-to-point authentication of routing message. One-way hash chain is less expensive than asymmetric cryptography. However the drawbacks of one-way hash MAC are: (i) one-way hash-based authentication requires clock synchronization at granularities that is hard to reach, (ii) delayed key disclosure results in the delay in the verification of routing message.

ARAN [30] and SAODV [31] are the protocols based on digital signature with asymmetric key cryptography. ARAN assumes a trusted certificate server to maintain fresh certificates for nodes. Based on the public key of the trusted server, the source node begins a route discovery by initiating a packet with the signature of its private key. Mandatory end-to-end authentication, using certificate, timestamp and nonce are policies to guarantee the authentication, message integrity and non-repudiation during the processing of routing recovery. The drawback of the digital signature is that it involves more computation overhead.

A detailed overview of these ad hoc security protocols can be found in paper [32]. These ad hoc security protocols in [26-30] are based on the standalone MANETs. All of these security protocols endure a serious problem that there are no prior trust relationships among MSs because of the absence of central authority in the MANET. It is difficult to consistently identify a MS with a unique identifier because it is easy for a MS to change its identity. Therefore it is difficult to establish the trust relationships among MSs. A malicious ad hoc node can use a forged ad hoc identity and then makes feigned trust relations with other nodes and then attacks the MANET internally. In the integrated Internet and MANET, it is critical that a MS can be identified in a consistent and unique manner. This dissertation in part A provides a security protocol for the integrated Internet and MANETs by tightly combining Mobile IP security and ad hoc security. The dissertation further develops the trust relationship for MANET through the authentication from the Internet authorities.

1.2 Secure Mobility Management in Heterogeneous Multi-hop Wireless Networks

The dissertation in part B addresses a new mobility protocol and its Secure Mobility in integrated Internet and heterogeneous networks. Mobile IP [33] [34] provides continuous mobility functions for MS with two entities on the Internet: home agent (HA), and foreign agent (FA). A HA is the server on the mobile host's home network that maintains the information about the host's current location, as identified as care-of-address (CoA), and security credentials. On the other hand, FA is the server on the visiting network providing the CoA and security administration of the visiting network. Mobile IP handles the mobility of MSs, but is inadequate for mobile wireless networks because of its high update latency, large Internet signaling load, and lack of support of micro-mobility. The standard mobile IP protocol addresses the processes of discovery, registration, and tunneling. FA/HA periodically send out an *advertisement* message indicating its presence to mobile stations (MSs). From the *advertisement,* a MS obtains a CoA (care-of-address) or co-located CoA for registration. The registration of a MS informs its HA to update the mobility binding of the MS. A *registration reply* message issued by FA indicates the status of registration request. Only upon successful registration, datagram can be tunneled between CN (correspondent node) and MS. Recently a variety of enhancements have been proposed to overcome the shortcomings of the base mobile IP [35], e.g., MIP-RO, MIPv6, HMIPv6, IDMP, HAWAII, TeleMIP, Cellular IP, Fast handoffs, EMA, and Proactive Handoff. For example, Cellular IP [36] [37] differentiates the global domain and the local domain, and thus supports micro-mobility. Also, the Telecommunication Enhanced Mobile IP (TeleMIP) [38] is a scalable and hierarchical IP-based architecture that provides lower latency and signaling overhead in comparison with the standard Mobile IP. However, all above mobility protocols can not support heterogeneous multi-hop communication in an integrated Internet and heterogeneous multi-hop wireless networks. A new mobility management protocol is needed to support the migration for both single and multi-hop MSs in an integrated network. Each MS has a multi-hop paging/routing cache maintaining the necessary information for location management and connection management. In Chapter 5, a new mobility management protocol is proposed to support the multi-hop communication. When a single hop or multi-hop MS is in an idle state, the location management of the protocol provides micro-mobility management while is moving in the local domain. On the contrary, when the MS have packets to send or receive the MS moves to active state and a connection management scheme enables the MS to maintain the high performance route for connecting the Internet.

Although the basic mobile IP protocol defines an Authentication Extension (AE) to support authentication in registration (*secret-key based authentication*), there are some deficiencies in this protocol [40] [41]. Since the protocol only requires authentication between HA and MS, network may be attacked unless overall authentications are enforced between the FA and HA and between the FA and MS. To enforce an overall authentication, besides the secret-keys between MS and HA, extra heavy secret-key management is necessary between MS and FA. The Mobile IP protocol doesn't prevent replay attack between FA and either MS or HA. The MoIPS [39] implements a public key management system that supports the mobile IP as well as the route optimized mobile IP. The system is built upon a DNS-based X.509 public key infrastructure in order to supply cryptographic keys to authenticate mobile IP registrations and to establish IPsec tunnels for mobile IP redirected packets.

Public-key based authentication [40] provides a strong, scalable authentication strategy for MS, HA, and FA to authenticate one another based on asymmetric or public key cryptography. The protocol enables mobile IP to work exactly the same way as secret-key-authentication. The *minimal public-key based authentication* [41] protocol uses secret key cryptography in order to minimize the requirement of computing power, as well as the administration cost imposed on MS. In addition, the protocol provides the scalability and non-repudiation that are likely to be demanded by the users.

In an integrated Internet and heterogeneous multi-hop wireless network, however, all of these authentication protocols suffer two serious problems:

- A multi-hop MS, which is outside the direct coverage of a BS and connects to the BS with a multi-hop route, doesn't have the capability of verifying the authenticity of the *advertisements* that are forwarded by an intermediate node.
- The existing mobile IP protocol cannot provide security protection for the registration of a multi-hop MS. Current mobile IP security is deployed on single wireless hop in which each MS can exchange registration and authentication directly with the FA.

Therefore, above authentication protocols cannot be used for an integrated multi-hop network to support authentication of a multi-hop MS.

In a single hop network, the above authentication protocols [39] [40] [41] have no need to maintain a multi-hop cache for each MS. Data packets are sent between each MS and the FA directly. However, in a multi-hop network, a multi-hop routing cache at the BS is required for maintaining the path to each multi-hop MS [2] [10] [42]. Thus, beside the mobile IP registration security, it is very important to create and update a multi-hop routing cache in a way to prevent it from poisoning.

1.3 Dissertation Structure

The rest of this dissertation is organized as follows. After illustrating the network architecture of heterogeneous multi-hop networks, Chapter two provides a detailed investigation of potential vulnerability in heterogeneous multi-hop networks and then develops a security thread model covering a variety of security attacks, including Internet connectivity attacks, MS mobility attacks, multi-hop routing attacks, packet forwarding attacks, and specific protocol attacks. Then, the dissertation is divided into two parts:

- On secure communication for integrated Internet and MANET,
- On secure mobility management in heterogeneous multi-hop wireless networks.

Each part has two chapters. In part A, Chapter three provides a design and implementation of an integrated Internet and MANET protocol for providing bi-directional Internet connectivity for ad hoc hosts. Our extensive experimental results show the protocol achieves higher packet delivery rate and less latency in multi-hop communication in comparison with other leading protocols. Chapter four proposes an effective authentication and multi-hop routing protocol between ad hoc MSs and their accessing IP networks in a heterogeneous multi-hop wireless network. Our authentication protocol is different from other mobile IP authentication protocols in two important ways (i) the protocol supports the authentication request from a single hop or multi-hop MS, (ii) the protocol uniquely identifies each MS to enhance the trust relationship between MSs and provides identity and integrity protection for multi-hop communication. Based on the authentication protocol, a secure routing protocol is further developed to provide protection for the process of route discovery for multi-hop communication. The security analysis, comparisons and performance studies show the effectiveness of secure multi-hop communication in integrated Internet and MANET.

Part B has two chapters illustrating the Secure Mobility Management in heterogeneous multi-hop wireless network. In Chapter five, a proposed mobility protocol, called Multi-hop Cellular IP (MCIP), integrates mobile IP and cellular IP in support of heterogeneous multi-hop communication for ad hoc MSs. This protocol supports the migration for both single and multi-hop MSs in heterogeneous multi-hop networks. In the architecture of a MCIP network, a heterogonous multi-hop network is divided into multiple domains. For each MS in a local domain, a HA/FA acts as the administrator with a multi-hop paging/routing cache to maintain necessary information for location management, connection management as well as multi-hop routing reconfiguration. Then, in Chapter six, a secure macro/micro-mobility protocol (SM^{3}P) is introduced for mobility security of the MCIP. In the proposed SM^{3}P, mobile IP security has been extended for supporting macro-mobility across local domains through the process of multi-hop registration and authentication. In a MCIP local domain, a certificate-based authentication achieves the effective routing and micro-mobility protection from a range of potential security threats. Our evaluation and simulation demonstrates the effectiveness of the SM^{3}P.

CHAPTER II

SECURITY SURVEY IN INTEGRATED INTERNET AND HETEROGENEOUS MULTI-HOP WIRELESS NETWORKS

It is widely recognized that many new wireless technologies have been introduced to cater ever growing demands for diversified services [1]. WLANs (e.g., IEEE 802.11a/b/e/g and HiperLAN/2), MANs (e.g., IEEE 802.16), and WANs (e.g., 1G, 2G, 2.5G, 3G, GSM, the proposed IEEE 802.20, etc.) employ different operating radio spaces for satisfying various communication scenarios. A multitude of architectures and protocols [3-10] [43-52] can extend traditional cellular and WLAN to multi-hop communication using the revolutionary paradigm of MANET (mobile ad hoc network). In these approaches, a MS outside the coverage of a BS may access the Internet through ad hoc relaying with the help of other MSs. Moreover, mesh network technology [53], "opportunistic MANETing", intends to integrate wireless PAN (i.e., IEEE 802.15), WLAN, MAN, and WAN with MANET as a commodity multi-hop MANET. Integration of wireless Cellular (MAN/WAN), WLAN, and MANET (ICWM) implies providing ubiquitous Internet connectivity for MSs. A remarkable characteristic of the aforementioned ICWM networks is that the MANET communication is not isolated anymore, but is an integral part of the infrastructure-based networks for Internet access. Research in ICWM networks is increasingly gaining popularity due to availability of multi-mode MSs. The Third Generation Partnership Project (3GPP) has recently taken an initiative to develop a cellular-WLAN interworking architecture to enable 3GPP system operators provide public WLAN as an integral component of the services offered to cellular subscriber [54]. Thus, ICWM networks provide flexible and effective communication where a MS can communicate with the Internet either by a single hop or by a multi-hop path. These networks also support peer-to-peer connection when the source and destination are in close vicinity [1] [53]. These proposed ICWM architectures expectedly offer one or more of the following benefits in terms of ubiquitous Internet connectivity:

- An ICWM network provides heterogeneous wireless network access (e.g., cellular, IEEE 802.11, IEEE 802.16) to multi-mode MSs;
- An ICWM network supports the Internet connectivity to the MSs, which are located outside the radio coverage of the BSs, and thus extends the network coverage to multi-hop MSs;
- In a ICWM network, a single cellular transmission between a MS and a BS can be broken into multi-hop HDR (Higher Data rate) communication links for the purposes of supporting higher communication speed and reduced co-channel interference;
- An ICWM network improves the communication throughput and reduces packet transmission latency, based on availability of different types of radio network in the roaming area (cellular, WLAN or MANET);
- An ICWM network can enhance the capacity of a BS by using multi-hop relaying and diverting the traffic from a congested BS to a neighboring BS and the network provides load balancing among BS, thereby increasing the effective network system capacity; and
- An ICWM network allows MANET communication between two MSs without going through the BSs, and these results in saving radio bandwidth of the BSs thereby increasing the overall capacity of BSs.

Because of ICWM's deployment flexibility, it is very attractive from a commercial perspective [2] [3-10] [43-52]. However, before practical applications of ICWM networks become widely available, such networks need to have adequate built-in security. Various security threats are possible in ICWM networks due to the vulnerability of heterogeneous communication and relaying by the MANET. Existing ICWM architectures are based on the assumption that all MSs trust each other where each MS honestly participates in the process of route discovery, and faithfully forwards all data packets to their respective destinations. In multi-hop communication, however, a malicious MS may attack a network just by injecting or modifying the messages. The malicious MS may also masquerade itself as someone else. A forged wireless infrastructure (such as BS) may imperil the trustworthiness between a MS and the Internet. A multi-hop route may be modified intentionally during the process of route discovery. These attacks certainly degrade the efficiency of packet relay, increase packet delivery latency, lower packet delivery rate, mislead packets, or even turn the Internet connectivity down. As opposed to research on routing protocols and internetworking architectures, research on ICWN network security is still in its inception stages and has gained attention recently [54][55]. It is an important issue to investigate the security properties of ICWM networks so that comprehensive security control protocols can be developed.

The goal of this chapter is to provide a better understanding of the security threats and challenges in this emerging field. Our main contributions in this chapter can be divided into four groups.

- We discuss the unique characteristics of ICWM networks to better understand security attacks and solutions.

- We provide significance of and the nature of security attacks against ICWM networks. To the best of our knowledge, this is the first comprehensive investigation that identifies security weaknesses associated with the ICWM architectures.
- We introduce several possible attacks against ICWM networks and provide novel solutions.
- We provide an insight into the research challenges and latest developments on ICWM security protocols and present various open research issues that need to be explored in a much greater depth.

The reminder of this chapter is structured as follows. Section 2.1 investigates the ICWM architecture and its communication protocols. We particularly emphasize these network characteristics that impact the secured communication. Then, in Section 2.2 the security model provides a comprehensive study of ICWM security threats. Section 2.3 describes open challenges for designing and implementing security protocols. Finally, the concluding remarks are included in Section 2.4.

2.1 The Architecture of Integrated Cellular, WLAN and MANET

Before studying the possible attacks and evaluating any ICWM security solution, it is necessary to have a clear understanding of the network architecture as well as possible breaches in the network security.

2.1.1 An Overview of ICWM Network Architecture

The architecture in Figure 1 depicts a simplified ICWM network that integrates infrastructure and wireless multi-hop networks covered in [3-10] [43-52]. As can been seen from Figure 1, an ICWM network is divided into domains/subnets (e.g., domain 1 and 2) and the basic components of this network are MSs, BSs/APs, Home Agent (HA)/Foreign Agent (FA), and the core IP network. A BS is an access point and serves as

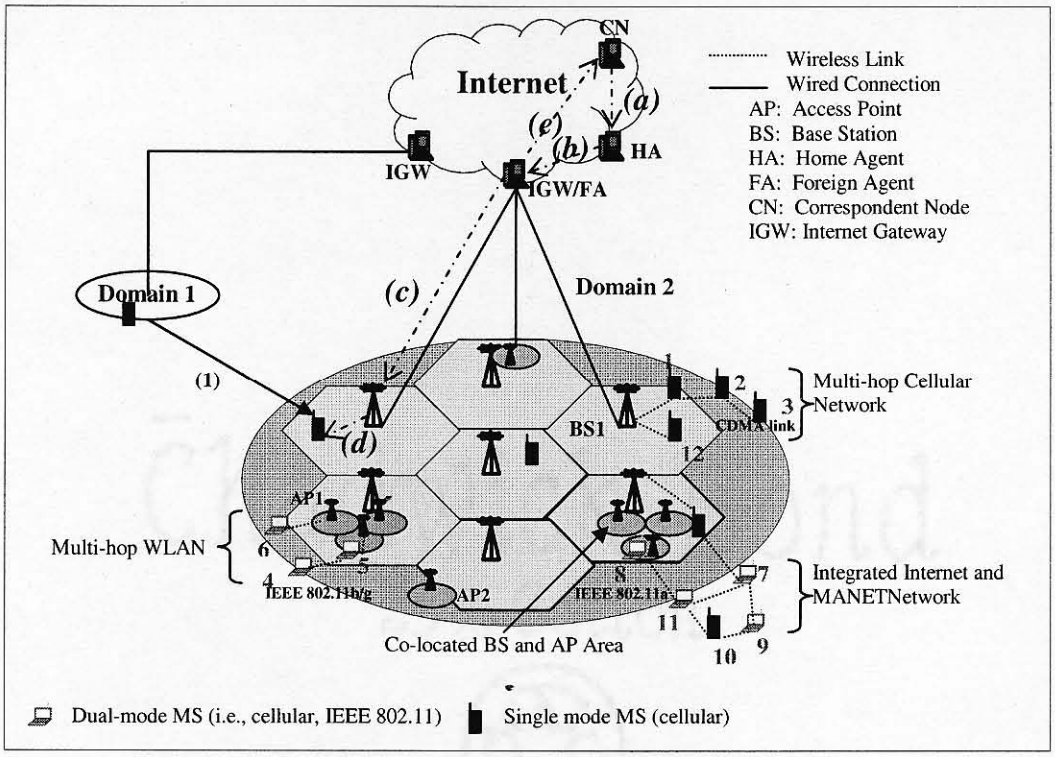


Figure 1. A Simplified ICWM network Architecture.

the communication bridge between wired and wireless networks for MSs. A BS provides one or more wireless radio access interfaces (i.e., cellular, IEEE 802.11, and IEEE 802.16) to MSs. The communication path for a MS spans through wireless link (link *d* as shown in Figure 1) and wired/Internet (a->b->c and c->d as shown in Figure 1). The BSs can be connected in many different ways [36] [53]. A BS can be connected to the Internet by cables or connected to other BSs through wired [36] or wireless connection [53]. A BS and an AP can be co-located in the hot-spot area where the traffic is high, as shown in Figure 1. A multi-mode MS possesses multiple radio interfaces, i.e., cellular interface or WLAN interface. The MS can connect to a BS with a single hop or multi-hop path, using an appropriate radio interface. In the multi-hop communication, a MS operates in MANET communication mode and can move in an arbitrary direction and speed. At a given instance, a particular MS can be either located within or outside the coverage of

BSs. If a MS moves outside the coverage of direct transmission from BSs, multi-hop relaying is needed for MSs to obtain services from BS to communicate with the Internet.

When the MS is visiting a foreign network, it registers with the FA for the purpose of creating a mobile binding at its HA so that upon receiving packets from a CN (correspondent node), the HA can deliver them to the FA, which further forwards the packets to the destination by single hop or multi-hop path. In an ICWM network, multi-hop route discovery protocol is needed for supporting multi-hop communication. According to routing protocols, ICWM architectures can be divided into three categories: multi-hop cellular networks, multi-hop WLANs, and integrated Internet and MANET as illustrated in Figure 1:

- Multi-hop WLAN networks,
- Multi-hop cellular networks, and
- Integrated Internet and MANET networks.

The Multi-hop WLAN networks allow the MSs to obtain services from the Internet through WLAN. A WLAN AP provides a higher communication speed to MSs at the expense of smaller radio coverage as compared to a cellular BS. For example, MS5 in Figure 1 accesses WLAN and obtains services from the Internet with the help of MS4 using the IEEE 802.11g link. The routing protocols for multi-hop WLAN networks include Two-Hop-Relay [50], HWN [51], 1-hop and 2-hops Direct Transmission [52].

The multi-hop cellular networks provide the Internet connectivity to MSs through WAN/MAN BSs such as GSM, IEEE.802.20, and IEEE 802.16. For example, as shown in Figure 1, MS3 connects to the Internet with the path MS3- MS2-MS1-BS1. A cellular BS may be connected to the Internet directly or via cellular infrastructure gateway such as a TCP/IP gateway. The routing protocols for multi-hop cellular networks include A-GSM [43], MCN [44], iCAR [45], MADF [46], UCAN [47], ODMA [48], and SOPRANO [49].

In an integrated Internet and MANET networks, each ad hoc MS runs a MANET routing protocol [32] such as DSDV, AODV, or DSR. The MANET routing protocol is used to construct communication path between two MSs. Meanwhile, a separate protocol [3-10] coordinates mobile IP and the MANET routing protocol to obtain the Internet connectivity.

2.1.2 ICWM Network Traits

Similar to other wireless systems (single hop cellular, WLAN, and MANET), an ICWM network has the basic characteristics such as open wireless medium, mobility and constrained terminal power capability. However, as shown in Figure 1, an ICWM network differs from other distributed mobile systems in some important ways: infrastructure-support, multi-hop, and multi-mode terminals with multiple radio interfaces. Many security preconceptions must be discarded because of these differences. Table 1 illustrates the crucial distinctions between ICWM networks with other networks.

2.2.1 Infrastructure-support vs. Ad Hoc

A MANET is a self-configurable network with the capability of communication among ad hoc MSs without any infrastructure or any centralized administration. The traffic in MANETs is typically between any pair of MSs. ICWM networks share similarities with MANETs: multi-hop networking and multi-hop communication. The dominant distinction between ICWM networks and MANETs is infrastructure-support in ICWM networks. In contract, the Internet and wireless infrastructure provides the

Table 1. Comparisons of Network Architectures and Security Factors

Security factors	Single hop wireless networks		Multi-hop wireless networks			
			MANETs	Integrated Cellular, WLAN and MANET (**ICWM**)		
	Cellular Network	WLAN	MANET	Multi-hop cellular network	Multi-hop WLAN	Integrated Internet and MANET
Internet Access Point	Cellular BS	WLAN AP	No	Cellular BS	WLAN AP	Mobile IP Gateway or BS
Terminal radio interfaces	Single radio interface (GSM or 802.16 etc.)	Single radio interface (HiperLAN/2 , HiSWABa, IEEE802.11a /b/e/g)	Single radio (IEEE 802.11) operating in ad hoc mode	Single or Dual-mode radio interfaces (cellular, IEEE 802.16 a/c/c/g, IEEE 802.11a/b/e/g)	Single or Dual-mode radio interfaces (cellular, IEEE 802.16 a/c/e/g, IEEE 802.11a/b/e/g)	Single or Dual-mode radio interfaces (cellular, IEEE 802.16 a/c/e/g, IEEE 802.11a/b/e/g)
Terminal operating mode	Single hop cellular access	Single hop WLAN access	Ad hoc Multi-hop	Multi-hop in cellular and ad hoc	Multi-hop in WLAN and ad hoc	Multi-hop in cellular/WLAN and ad hoc
Routing Protocols	No	No	DSDV, AODV, DSR etc	A-GSM, MCN, iCAR, MADF, UCAN, ODMA, and SOPRANO	Two-Hop-Relay, HWN, 1-hop and 2-hops Direct Transmission	Integrated mobile IP and ad hoc protocols [3-10]
Infra-structure access security	GSM, UMTS, or IEEE 802.20 Security	HiperLAN/2, HiSWABa, IEEE802.11a /b/e/g security	No	Extending cellular security (e.g., GSM and IEEE 802.20 security) from single hop to multi-hop	Extending IEEE 802.11a/b/e/g or HiperLAN/2 security from single hop to multi-hop	Securing integrated mobile IP and ad hoc routing protocol
Routing security	No	No	Securing DSDV, AODV, DSR etc	Securing all communication alternatives	Securing all communication alternatives	Securing ad hoc routing protocol

centralized administration for MSs in ICWM networks. In an ICWM network, most of the traffic goes through BSs. It is difficult to establish any distributed trust relationships among MSs as a central authority is absent and MSs' have ability to forge identity. A malicious MS can use a forged ad hoc identity and then makes feigned trust relations with other MS to attack the network internally. However, in an ICWM, a MS can access the Internet by way of single or multi-hop connectivity through a BS or an Internet Gateway. The wireless and Internet infrastructure (i.e., Authentication, Authorization, and Accounting - AAA server [74]) can serve as an authentication authority and a security administrative center for ICWM security. The infrastructure-supported security deployment has a fundamental impact on ICWM security implementation in comparison with pure ad hoc security.

2.2.2 Single Hop vs. Multi-hop

A MS in an ICWM network communicates with a BS or another MS by using single hop or multi-hop path with the help of other MSs. On the contrary, a MS in traditional cellular or WLAN connects to Internet through BS using single wireless hop. Multi-hop imposes many new challenges such as open network architecture, shared wireless medium, limited resource constraints, and dynamic network topology that have a significant impact on security. All possible attacks found in MANETs can be easily mounted on ICWM networks due to the common characteristic of multi-hop route. In cellular or WLAN, security solutions only provide protection for one hop connectivity between a MS and BS by securing MAC/link-layer protocols. However, in an ICWN network, the security solutions should extend beyond the single hop to multi-hop routing security at the network layer. Therefore, traditional cellular and WLAN security protocols are not applicable to multi-hop communication between a MS and a BS. Although the security protocols for MANETs provide routing security, unfortunately, they can not be adopted for ICWM networks directly because these protocols are based the assumption of the absence of infrastructure-supported authentication.

2.2.3 Single Radio Interface vs. Multi-mode

Another distinguishing characteristic of ICWM networks is multiple radio interfaces in MSs. A dual-mode MS equipped with two radio interfaces may switch from one radio interface to another (e.g., redirecting a flow from a cellular radio interface to WLAN radio interface when a dual-mode MS moves to the WLAN area) as network accessibility or topology changes. Cellular, WLAN and MANET security protocols are based on a single radio interface and do not have any provision for protection when communication migrates from one radio interface to another. It is necessary to develop integrated schemes for security interworking between multi-mode radios. For example, a 3GPP subscriber MS, which is equipped with cellular–WLAN radio interfaces, may handoff its service from a 3GPP network to a public WLAN network. Here, in order to provide security features to the MS while accessing the public WLAN network, it is necessary for the public WLAN network to reuse the 3GPP subscription and 3GPP-based authentication/authorization as well as 3GPP-based security key agreement using SIM/USIM card [1] [55]. In the case of multi-hop route, the infrastructure security protocols (e.g., 3GPP cellular security, mobile IP security for WLAN) must coordinate with multi-hop routing protocols for multi-hop MSs. At the same time, it is important for the security protocols to minimize the authentication latency induced by the multiple radio interfaces (e.g., networking selection) and multi-hop route.

2.2 Security Threat Model

Several types of attacks are possible on the Internet including Domain Name Service (DNS) “hacking”, routing table “poisoning”, packet “mistreatment” and denial-of-service (DoS) attacks. Numerous approaches have been developed for defending against these types of attacks. An excellent survey of various types of attacks can be found in [56]. In an ICWM network, the security issue is more complex than that of individual cellular, LAN or MANET. Because of multi-hop transmissions, BSs have the task to locate a destination MS, and then forward packets to the destination MS. For this purpose, it is necessary for each BS to record the routes for the destination MSs. In this case, a malicious MS may poison the multi-hop route, resulting in a packet loss. Most of the attacks against ICWM networks fall into one of the following categories:

- Eavesdropping and traffic analysis, unauthorized Internet access, Internet-wireless DDoS, Wireless-Internet DDoS,
- Registration attacks (registration poisoning, bogus registration, registration replay attack), forged BS, BS cache poisoning,
- Avoiding tracking by having unknown mobility, avoiding tracking by changing identity,
- Multi-hop routing attacks (modification, impersonation, routing loop, duplication, selfishness and other denial of services),
- Packet forwarding attacks (modification, selfishness, and power drain), and
- Mobility handoff attacks, specific wireless protocol attacks (WLAN, UMTS, and IEEE 802.16).

This section provides a general discussion on these types of attacks and emphasizes the specific attacks that differ from other wireless networks or remain undocumented in the literature.

2.2.1 Eavesdropping and Traffic Analysis

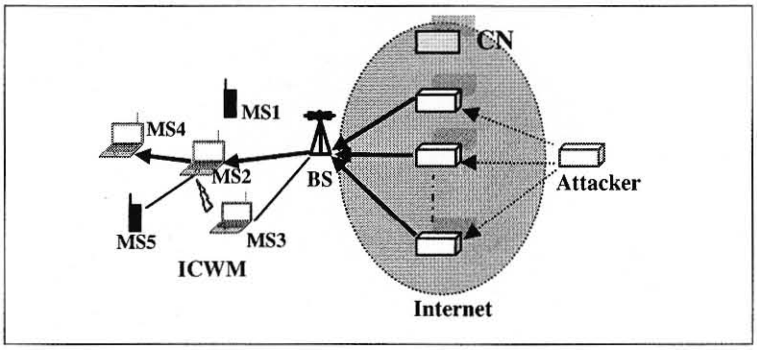


Figure 2. Eavesdropping and Traffic Analysis.

The broadcast nature of the wireless transmission medium renders radio links insecure in ICWN networks. An attacker can easily eavesdrop on all ongoing communication. In addition, an attacker in an ICWN network may reside on the multi-hop path and present itself as an intermediate relaying MS. The data packets can be copied and distributed by the attacker when it forwards packets to next hop (e.g., MS2 is the intermediary relaying MS for MS4, MS5 and MS1). The attacker can read the transmitted data, and also gather information by examining the monitored packets such as address, size, number and time of transmission. The attacker can seize the information of BS such as location and IP address. Also the attacker can know the critical MSs, which provide the Internet connectivity for other MSs (e.g., MS2). The information obtained is useful for many attacks. As an example, consider Figure 2. When the MS3 (an attacker or attacker conspirer) knows that MS2 is the critical MS for providing the Internet connectivity for MS1, MS4 and MS5, the attacker can launch the attack towards MS2 for the purpose of turning down the Internet connection of MS1, MS4 and MS5.

2.2.2 Unauthorized Internet Access and Attacks

A malicious MS may access the ICWM network and enjoy free network usage by way of single hop or multi-hop communication. Wireless Internet service providers (ISPs) may be accessed by an unauthorized MS because of the lack of correct ISP configuration. As for the network devices without having adequate security measures, the security threats may come from within the network itself. A registered MS of the network may access, read, copy and distribute data file that is no business of the MS.

When the Internet is accessed by way of a multi-hop, the malicious MS consumes precious resources like power and bandwidth. Although the free network usage may not be a significant threat to the Internet, an authorized access is the first step for the MS to control the Internet infrastructure and to attack these components. After entering the Internet, the attacker may use some techniques like Medium Access Control address spoofing to gain access to the network infrastructure. For instance, in Figure 3 MS1 implements the man-in-the-middle attack between MS3 and its default network router (Router shown in Figure 3) from which MS1 sees all the traffic between MS and Router. First MS1 connects to the Internet and sends a malicious ARP (address resolution protocol) reply to Router, associating MS1's MAC address with MS3's IP address. With such access, the Router assumes MS1 to be MS3. Next, MS1 sends a malicious ARP reply to MS3 associating MS1's MAC address with the Router's address. In this case, the MS3 believes MS1 is its router (Router). Finally, MS1 can access the session between MS3 and Router. Thus, all data packets from MS3 will be delivered to MS1 first and then MS1 forwards the received data packet to Router. Finally, the router sends the packets to destination (e.g., MS4 as shown in Figure 3). In the opposite direction, all the packets from MS4 will be forwarded to MS1 by Router and MS1 sends the packets to MS3. In this manner, MS1 intercepts the traffic between MS3 and MS4.

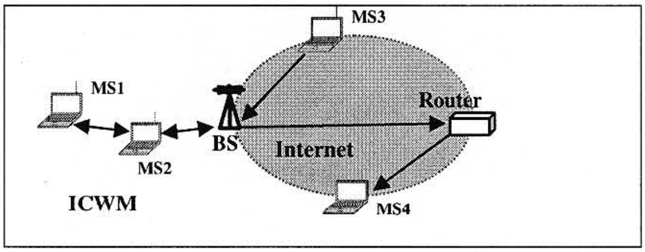


Figure 3. Unauthorized Network Access and Man-in-the-middle Attack.

2.2.3 Internet-Wireless DDoS

We introduce a common attack that could be disastrous against ICWM networks as Internet-Wireless DDoS (distributed denial-of-service) and shown in Figure 2. An ICWM network may be crippled by the attacks from the Internet. Such attacks are referred to as from the internet as the Internet-Wireless-DDoS. Compromising a BS or wireless router can disable or congest the wireless communication in a domain. As a typical example, an attacker first initiates the control over one or several computers on the Internet by using some kind of automatic intrusion software to hack them [56]. After getting control of the computers, the attacker synchronizes them to send traffic in bulk towards one or more multi-hop MSs that are associated with the same BS. The traffic first travels through the Internet toward the BS. The BS is not aware of the illegality of the packets. Then, the BS forwards the traffic towards the destination MSs (e.g., MS4 in Figure 2) according to the address given in each packet header. In the end, the packets travel through the multi-hop wireless network. This process can achieve the following goals.

- Exhausting the wireless resources of the BS: Limited availability of BS radio spectrum is always a bottleneck in the Internet-wireless communication. When the compromised MSs transmit a huge number of packets to a BS, the BS may exhaust its radio resource by forwarding the packets on its air interface. Therefore, the attack can immediately block a large number of MSs that are communicating through the BS.
- Depleting the wireless resource of MSs (power, radio/bandwidth): A MS has limited communication capability as compared to fixed devices in terms of available power and bandwidth. The packets transmitted by compromised

computers travel through multi-hop wireless network definitely cause the power and bandwidth consumption at each intermediate MS. The multi-hop transmissions can simply disable or slow down other communications along the multi-hop path. It can be observed from Figure 2 that the traffic from the compromised MSs consumes the bandwidth of the intermediary relaying MS2. The attack at least slows down the communication between MS5 and CN because MS5 uses the same intermediary relaying MS2 as the attack. Moreover, if the power of MS2 is exhausted, then the Internet connections for MS 5 and MS 4 will also be turned down because MS2 is the only MS to reach BS from MS 5 and MS 4.

2.2.4 Wireless-Internet DDoS

This section introduces a new attack called wireless-Internet DDoS which could be initiated at the wireless side to the Internet and the multi-hop wireless network. It can attack the BS radio spectrum when a number of wireless devices around a BS send packets simultaneously in bulk to the BS by single or multi-hop routes. In this case, at a given instant the BS may be disabled after using up all spectrum resources. At least the bandwidth for normal traffic may be significantly reduced when a number of junk packets take up the majority of communication channels.

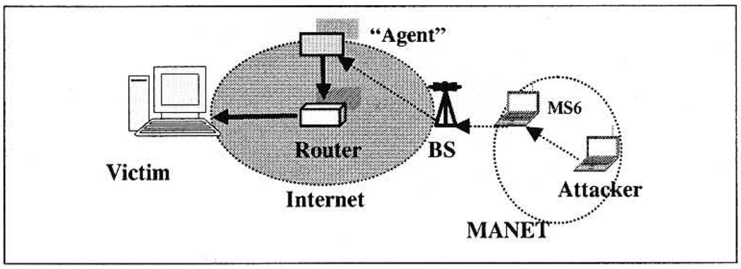


Figure 4. Wireless-Internet DDoS.

The attacker accessing the Internet through a BS by single hop or multi-hop can impair the Internet infrastructure and its communication. As shown in Figure 4, the attacker, which operates in an ad hoc mode and runs the ad hoc routing protocol, connects to BS via the MANET. In this case, attacker attacks wired network using its ad hoc identifier. The attacker can attack the cellular/WLAN/internet infrastructure including router, DSN server, TCP/IP gateway etc. In a similar manner, an attacker can also access the wired network with an invented or spoofed IP address through multi-hop cellular or WLAN network. Moreover, the attacker can collude with other attackers or compromised internet components for implementing more sophisticated attack on the Internet or wireless network. For example, to execute DDoS, the attacker from the wireless network can compromise several computers on the Internet as attacking "agents". Then the attacker directs the "agents" to send a huge number of packets (i.e., UDP, TCP, or ICMP packets) to disable a target (e.g., a server). The compromised "agents" may also launch the "Internet-Wireless DDoS" to attack an ICWM network as illustrated in Section 2.2.3. Since the attacker has accessed the Internet by multi-hop path and a temporary ID (e.g., an ad hoc ID), it is very difficult for victim to trace back the source of the attack (see Section 2.2.8 and 2.2.9).

2.2.5 Registration Attacks

Mobile IP provides continuous Internet accessibility for MS when it visits a foreign network. When switching between networks, the MS has to create a mobility binding at home network through a registration procedure with visiting network. The registration procedure occurs immediately after the MS is switched on or moves to a visiting network. Although the specific registration procedures for different wireless networks (cellular roaming, IPv6) are different, the general registration procedure in ICWM networks is as follows:

1. In case when a MS is single hop away from the BS, the MS obtains the connection to the BS directly. On the other hand, if the MS is outside of all BSs' transmission range, then the MS establishes the connection to the BS with a multi-hop route.
2. After obtaining the connection to the Internet, a MS sends a registration request to the foreign network. The registration request contains the MS's identity, the home network address etc.
3. Upon receipt of the registration request, the foreign network forwards the request to the MS's home network for the purpose of creating a mobility binding for the MS.
4. After the mobility update for the MS, the home network replies the foreign network with a registration reply message carrying the registration result.
5. In the end, the foreign network forwards the registration reply to the MS with a reversed route.

The registration procedure provides the desirable property of mobility, but results in several serious security threats. Three typical types of registration attacks are possible including registration poisoning, bogus registration, and replay attack.

- **Registration Poisoning**: Malicious MS in the ICWM network can poison the registration procedure. In the ICWN registration procedure, a MS registers with the foreign network through a multi-hop route with the help of some intermediate MSs. A malicious MS can entice a multi-hop MS to choose the malicious MS as an intermediate MS by claiming to have a short or fast route to a BS. When a malicious MS is selected as the intermediate MS for a registering MS, the malicious MS can modify or drop the MS's registration request/reply. When the malicious MS modifies or rejects the registration request, the MS cannot correctly register with the foreign network. If the malicious MS modifies the registration result in a registration reply, the MS cannot access the Internet as if its registration request is rejected. For instance, in Figure 2, when MS4 moves to the ICWM network, it initiates a registration by sending out a registration request. When MS2 receives the registration request from MS4, MS2 modifies the address of MS4's home network in the registration request before forwarding the request to BS. Based on the modified request, the foreign network will forward the request to a wrong home network so that the foreign network cannot obtain a successful registration reply. Registration poisoning prevents multi-hop MSs from obtaining services from the wired network.

- **Bogus Registration**: This occurs when a malicious MS does a fake registration by spoofing an IP address and masquerading itself as someone else. The bogus registration causes a wrong mobility binding at the home network so that all packets are tunneled to the illegitimate MS, in place of the correct MS. By bogus registration, the attacker obtains the right to access the Internet so that it can implement further attack on the Internet such as Wireless-Internet DDoS. For instance, in Figure 2, MS3 does a forged registration by masquerading itself as MS5. Thus, all packets coming from Internet for MS5 are forwarded to MS3. In this case, MS5 cannot receive any packet from the Internet.
- **Registration Replay Attack**: In a replay attack, an attacker captures a legitimate registration request and replays the message to a BS. Even through the registration request message may be signed or encrypted and the attacker may not

know what the actual security keys are, the retransmission of the valid request is sufficient to gain access to the network if mechanisms to make the request unique (e.g., nonce or timestamp) are not used. In this case, the attacker gets a valid session with the right of the victim. In Figure 2, MS3 may forward a copy of the registration request originated from MS5. Without protection, the MS3 could perform a valid registration in the name of MS5 and obtain the right to access the Internet. Such attacks are more prominent in cases where tunneling is not used while carrying out the authentication.

2.2.6 Forged BS

In an ICWM network, an attacker can attack the multi-hop wireless network by advertising itself as a genuine BS using some forged messages or duplicate beacons recorded from a correct BS procured by eavesdropping. When a MS hears the fraudulent beacons from the malicious MS, it assumes that it is within the radio coverage of a genuine BS and then initiates a registration procedure. A registration request is issued from the MS to the forged BS. The forged BS replies with a bogus registration reply carrying the acceptance of the registration request. After receiving the registration result, the MS assumes that it has obtained the Internet connection through the forged BS and disconnects its communication from the genuine BS. One by one, the forged BS could entice a number of MSs to disconnect from the genuine BSs and establish connections with the forged MS either by single hop or multi-hop route. However, the MSs cannot obtain any Internet service correctly from the forged BS. This attack is valid in cases where the BS is not authenticated by the MS. For instance, in Figure 2, malicious MS3 advertises a high-speed connectivity to the Internet by sending bogus beacons to its neighbor. After hearing the beacons from the forged BS (MS3) but without realizing the fraud, MS2 and MS5 register with MS3 by a single hop. After registrations, MS2 and MS5 believe that they are connected with a BS with a higher speed, and thus disconnect the connection with genuine BS (BS1). An attacker using forged BS achieves the following:

- The forged BS captures registration information of MSs such as their home IP addresses, home network IP addresses. It can also break a proper Internet connection and can cause unwarranted registration delay.
- The forged BS can act as the gateway to the Internet and can seize the data packet or capture sensitive personal (e.g., password) or network data of MSs.

2.2.7 BS Cache Poisoning

In a single hop wireless network, it is difficult for a malicious MS to modify the radio mappings from which MSs enter a BS because each MS and the BS interact with each other directly. On the contrary, the BS in ICWM suffers from possible BS cache poisoning as multi-hop communication is now allowed. To support multi-hop communication, the routing cache is needed at each BS for the purpose of recording the multi-hop routes between the BS and each multi-hop MS. The multi-hop routing cache of a MS may be poisoned in several ways. For instance, when a multi-hop MS sends a multi-hop route update packet for creating or updating its multi-hop paging cache at a BS, the malicious MS may modify the packet that could result in multi-hop routing cache poisoning. Also, a malicious MS may send a wrong route-update packet on behalf of a genuine one. And the BS updates the routing information for the genuine MS with the wrong information sent by the malicious MS. To locate a multi-hop MS, the BS finds the first hop MSs in routing cache that can reach the destination MS. The data packet from the BS is forwarded, hop-by-hop, to the destination in accordance with the multi-hop route. When a multi-hop route is poisoned, the BS is unable to locate the destination MS by following the multi-hop route provided in the BS cache. For instance in Figure 2, MS3 sends a multi-hop route-update packet to the BS on behalf of MS5, and the BS updates the multi-hop routing cache from MS5-MS2-BS to MS5-BS. In this case, the packets of MS5 from the Internet will be lost due to the incorrect routing information. Note that cache poisoning has to be accompanied by address spoofing.

2.2.8 Avoiding Tracking by Mobility

The ability to avoid tracking is important to an adversary when it uses spoofed addresses during attacks. Many technologies, including link testing, logging, ICMP trace back, and IP trace back [57]; have been developed to enable the victim to trace back the source of the attacker. For example, link testing iteratively checks the upstream link until the source is reached [57]. It is a challenge for a victim to track the malicious MS in the ICWM network because: (i) the multi-hop route from a MS to a BS may change when the mobility forces network topology to alter in an ICWM network (ii) the visiting BS (network attachment) may change due to the mobility of MS. The tracking back technologies based on the fixed network may be invalid in the ICWM networks because of the change in multi-hop path. Meanwhile, the attacker can take advantage of the flaw of mobility protocol to implement some sophisticated technology so as to hide its IP address from being tracked. As for mobile IP mobility protocol, a MS has two IP address: the home address and the care-of-address. The care-of-address is temporarily assigned by foreign network and used as the current address for communication. The Non Disclosure

Method (NDM) [57] prevents against traceability of network connections in mobile environment by hiding the source and destination addresses of an IP packet from every forwarding device except the packet destination.

2.2.9 Avoiding Tracing by Changing Identity

Avoiding identification tracking is another important issue for attackers but remains unexplored in ICWM networks. MANETs may use different addressing solutions [32]; AODV, DSR, and TORA use Node ID; HSR has a hierarchical addressing solution; ZLHS use <zone id + node id> as MS's ad hoc address. When a MS enters a MANET, the MS will be automatically configured with an identity by the ad hoc protocol, using one of the above address solutions. For example, a DSR network will assign a MS with a Node ID (e.g., 001) when the MS enters the DSR network. If the MS leaves the DSR network and enters the DSR network again, the MS will be assigned a new available Node ID (e.g., 002). In the integrated Internet and the MANET, a malicious MS can participate in a MANET and establish its Internet connectivity by using its ad hoc identification. In this case, the malicious MS can easily implement its attacks to the Internet or wireless network over and over again by masquerading itself. When the malicious MS enters the Integrated Internet and MANET, the network configures the MS with an ID. If the malicious MS leaves the network, and enters the network again, the network will automatically configure the malicious MS with a new ID without knowing its last ID. The network cannot track and monitor the history of the malicious MS because each MS does not have a unique and consistent ID while leaving and entering the network. When malicious MS exhibits illegal actions on the Integrated Internet and MANET, the malicious MS can clear its bad record by reentering the network and reinitiating a trust relation with network by using a new ID. As shown in Figure 5, the attacker exhibits its Wireless-Internet attack by using ID A. If the network finds the attack from ID A, the malicious MS reenters the network with a new ID (ID B), and could attack again. The ability of the malicious MS to modify packets to spoof MAC addresses prevents a detection mechanism to quickly identity the malicious MS.

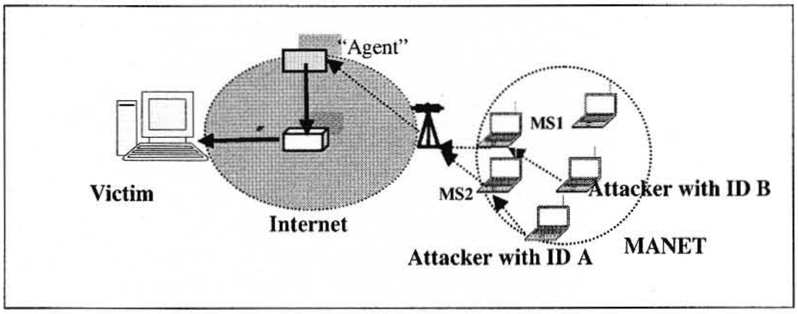


Figure 5. Wireless-Internet DDoS.

2.2.10 Multi-hop Routing Attacks

Almost all ad hoc routing attacks in a MANET can be carried to ICWM networks because of the common characteristic: multi-hop communication. Multi-hop route discovery is responsible for detecting the multi-hop routes between MSs and BSs. An attacker may exhibit its intentions by refusing to participate fully and correctly in a multi-hop route discovery process, without following the principles of integrity, authentication, non-duplication, confidentiality, and cooperation. Therefore, multi-hop routing attacks can be grouped into five categories: anti-integrity, impersonation, duplication, anti-confidentiality, and anti-cooperation [58] [32].

- Anti-integrity is the action of breaking away the integrity of a message. Modification is a typical example of anti-integrity. The malicious MSs modify, inject or delete some fields of a routing packet, and then forward the packet with falsified values in the packet fields. These fields may include the source or destination address, hop count, sequence number, etc. A typical example is the routing loop attack wherein, a malicious MS creates incorrect routing information at the routing table of each node. This is done by modifying the correct route packets destined to the victim MS in such a manner that a loop is created. Routing loop results in the consumption of power and bandwidth at the traveled MSs.

- Impersonations are those actions in which a malicious station spoofs an existing or forged IP address, or uses broadcast address to generate or duplicate one or more messages, and then forwards them to other MSs. In Figure 2, MS3 may masquerade itself by spoofing an invented address or an IP address of another MS. With the spoofed address, MS3 could entice MS5 to use MS3 as the shortest route for reaching BS1. Once MS5 starts sending packets to the BS, MS3 can easily compromise MS5's communication.
- In a duplication attack, a malicious MS sends a legitimate message more than once. These duplicated messages cause multiple receptions and processing overheads on adjacent MSs.
- In the anti-confidentiality attacks, an attacker may reveal sensitive information, such as the private key. Stealing, eavesdropping, guessing, brute-force and cryptanalysis are common ways for an attacker to identify sensitive information.
- Anti-cooperation includes dropping of packets or colluding with other attackers to disrupt a routing process. Selfishness is an example of anti-cooperation that a MS does not participate in the routing protocol or perform packet forwarding for the purpose of conserving its own energy.

2.2.11 Packet Forwarding Attacks

The protection for routing security cannot guarantee that each MS forwards the data packet in accordance with the routing table. It is possible that a MS honestly participates in the route discovery but mishandles data packets during packet forwarding. For instance, a MS participating in a route discovery may become a selfish MS during the multi-hop packet transmissions. In the packet forwarding stage, several types of security threats are possible.

- Modification: A MS may intentionally modify, drop, or inject data packets.
- Selfishness: A MS may drop all or a fraction of packets for the purpose of saving its power.
- Power drain: A malicious MS sends unnecessary packets or broadcast packets to drain the battery of the other MSs.

2.2.12 Mobility Handoff Attack

As stated earlier, when a MS visits a new network, it initiate the process of mobility handoff, indicating its willingness to redirect its data flow to the new network. When the mobility handoff is between two different BSs, but in the same type of network, the migration is horizontal handoff. A vertical handoff is the migration between two heterogeneous components, i.e., from a BS to IEEE 802.11 WLAN. Also, a vertical handoff involves the migration between two different radio interfaces at a multi-mode MS. A typical procedure for a vertical handoff includes three stages:

1. Detecting a single or multi-hop route to a new BS in a new radio interfaces,
2. Interworking either from a cellular to WLAN and vice versa, and
3. Flow redirection.

When a MS detects a new BS with better performance for the route (e.g., less delay, higher speed) than the current BS, the MS initiates a migration procedure to move from the current BS to the new BS. However, as seen from Section 2.2.10, the multi-hop routing could be easily misled. When a cellular subscriber MS migrates from a cellular service to a WLAN network, the WLAN network has the responsibility for securing interworking for the MS so that the data flow over a cellular network can be safely redirected to the WLAN network. In the mobility handoff attack, a malicious MS may entice a MS to initiate a false multi-hop handoff by

- Claiming a higher speed or better multi-hop route metric to a new BS from the malicious MS,
- Impeding interworking by dropping or modifying the registration packets, and
- Advertising forged beacons or replaying the stale beacons of a legal BS to cause the forged BS attack as illustrated in Section 2.2.6.

2.2.13 Specific Wireless Protocol Attacks

IEEE 802.16 is a standard for constructing wireless metropolitan area networks (MANs). The IEEE 802.16 security is implemented as a privacy sub-layer below the MAC protocol. It provides protection in terms of one-hop connection between a MS and a BS. However, the IEEE 802.16e security cannot prevent a MS from the BS forgery or replay attacks due to lack of a BS certificate. Like IEEE 802.16 security, most of the wireless protocols suffer from various security threats. For instance, the GSM security protocol may suffer from the attacks on authentication algorithm like cloning, confidentiality attacks like brute-force attacks, crypt analytical attacks, and attacks using loopholes in the protocol as illustrated in [59]. It is well known that IEEE 802.11 WEP protocol is vulnerable to many attacks. For more information, please refer papers [54] [55] [59].

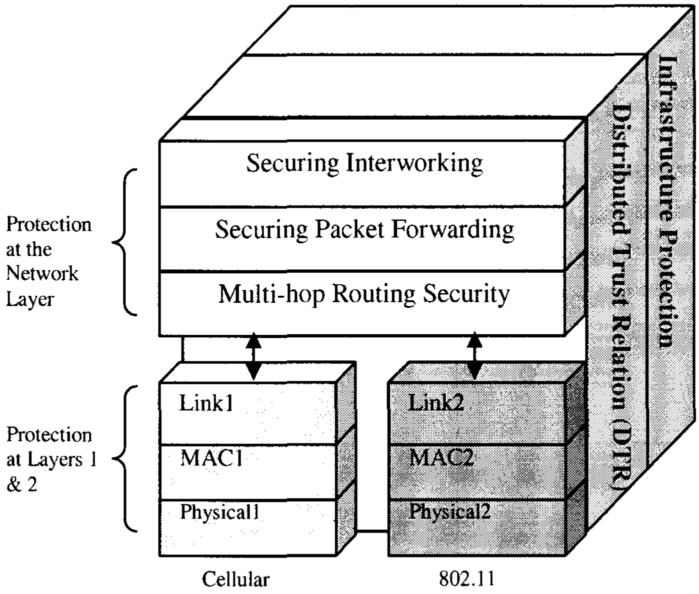

Securing Interworking
Securing Packet Forwarding
Multi-hop Routing Security
Protection at the Network Layer
Protection at Layers 1 & 2
Link1
MAC1
Physical1
Link2
MAC2
Physical2
Cellular
802.11
Distributed Trust Relation (DTR)
Infrastructure Protection

Figure 6. Security Protocol Stack and Security Challenges for a Dual-mode MS in an ICWM Network.

2.3 Open Challenges

Research in ICWM network security is still in its early stage and many issues remain unexplored. The security protocol stack for a dual-mode MS is given in Figure 6. As can be seen, the dual-mode MS has two Physical, Data Link and MAC layers for the cellular and IEEE 802.11 radio interfaces respectively. The network layer selects a radio interface according to the performance requirement [1]. As illustrated in Section 2.1, the fundamental vulnerability of an ICWM network comes from its distributed environment, open wireless medium, and heterogeneous multi-hop wireless communication. In such system, the distributed trust relationship (DTR) is the base for the security framework as shown in Figure 6. A robust security solution should include a DTR scheme that has a cross layered structure and executes over the Internet and wireless domains. Securing Internet (e.g., DSN, internet router) and wireless infrastructure (e.g., BS) prevent attacks such as unauthorized access or physical seizure. ICWM security should provide an end-to-end communication protection to MSs. However, end-to-end communication protection cannot be achieved easily because of the intricacies of the networks and communication protocols. As shown in Figure 6, the ICWM security issues at network layer include multi-hop routing security, secure packet forwarding, and interworking security.

2.3.1 Distributed Trust Relationship Establishment

Trustworthiness is the foundation of security. Almost all attacks in ICWM networks infringe the trust relationship in some form or the other. For the ICWM security, a DTR model across wired and wireless networks is desired. The DTR model considers the nature and basis of the trust relationships. It enables an inter-component authentication (e.g., BS-MS or MS-MS) based on legal binding. An ICWM network should require each single or multi-hop MS to provide authentication information upon entering the network. If a MS does not have a security binding with the network, the MS must register with the BS to get its identification verified before communication. On the other hand, before entering a network, the MS must verify the authenticity of the network (the accessing BS). When a MS visits a foreign network, the foreign network and home network must be mutually verified. In general, the DTR model provides the capability to mutually validate two components and construct trust relationship between them [55] [59] [64] [65]:

DTR 1.A MS (single or multi-hop) ↔ Home network,

DTR 2.A MS (single or multi-hop) ↔ Visiting network,

DTR 3.Home network ↔ Visiting network, and

DTR 4.A MS (single or multi-hop) ↔ Another MS (intermediate or destination).

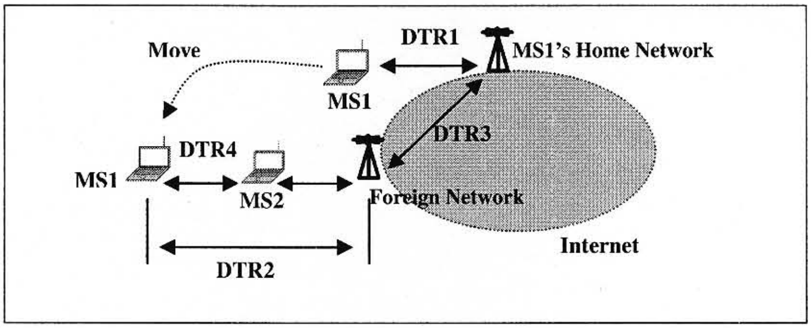


Figure 7. ICWM Network DTRs.

The DTRs in an ICWM network are illustrated in Figure 7. The home network with which the MS has legally contractual agreement (DTR 1) is supposed to be the most trusted domain for ICWM security [1] [55] [59]. DTR1 and DTR2 preclude a MS without credentials to participate in the network. This is achieved by verifying the identification of the MS at its home network. The attacks including unauthorized Internet access, wireless-Internet DDoS, and avoiding tracking by changing identity, violate the principle of correct identification to access the network. DTR1 and DTR2 also provide each MS with the capability to authenticate the accessing network (visiting BS) in order to prevent BS forgery. DTR3 establishes the trust relations between two different domains, and DTR4 constructs the trust relations between any two MSs.

A plethora of security protocols focus on the above DTR establishments based on various single hop wireless networks. These security protocols include GSM security, UMTS security, Mobile IP security, HiperLan/2, IEEE 802.15, IEEE 802.11, IEEE802.16, and IEEE 802.20 etc. The essential scheme with regard to DTR establishments is a procedure from which the single hop MS and the visiting network implement the mutual authentication. This is achieved with the help of home network and the negotiation of encryption keys for communication. In a cellular network, the GSM security (recommendations GSM 02.09 and GSM 03.20) enables the foreign network to authenticate the identity of a MS through the home network, and establishes the encryption keys, which are required for providing the confidentiality services. In a WLAN, the IEEE 802.11x security (e.g., IKEv2 EAP-IKEv2) authenticates a single hop MS and establishes the keys that are used for encrypting/decrypting messages according to the IPsec security specification. IEEE 802.16 security implements X.509 certificate-based authentication for MSs and ciphering key-based data association for confidentiality.

However, it is a new and critical issue to support DTR establishment for multi-hop MSs. This has two essential challenges [59] [64] [65]:

- A multi-hop MS ↔ Home/Visiting network: How can a BS trust a multi-hop MS such as MS1 in the foreign network as shown in Figure 7, and vice versa?
- A multi-hop MS ↔ another MS (intermediate or destination): How can a multi-hop MS trust another MS in an ICWM network (How can MS1 trust MS2, and how can MS2 trust MS1 as shown in Figure 7)?

The current mechanisms in the literature to establish DTR for a multi-hop MS have three steps [64] [59]:

1) The multi-hop MS sends its credentials to the network (BS) so that the BS can authenticate the MS,
2) The multi-hop MS validates the BS so that the MS can trust the BS for accessing the network, and
3) After the mutual authentication, the BS creates a security binding for the multi-hop MS. The security binding provides keys further used for multi-hop routing security and packet forwarding security.

The CAMA (cellular-assisted mobile wireless ad hoc network) security mechanism [64] implements DTR establishment with a mutual authentication scheme through cellular channel. After authentication, a key assignment scheme establishes the security binding for multi-hop WLAN communication on the IEEE 802.11 channel. A security protocol in [59] provides DTR for integrated Internet and MANET. In this approach, the FA/HA serves as the authentication center to check the credentials for each MANET MS. Upon receiving a registration request, the FA and HA carry out a series of authentications. Registration reply indicates the result of the registration. If the registration is successful, the FA binds the identity of MS with its public key in a certificate. Thereby, the MS can initiate a MANET route discovery for communicating with the other MSs by using the certificate.

Open Research Issues:

The process for DTR establishment is a nontrivial process during the multi-hop registration. DTR protocol for ICWM is largely open for study:

- Although a lot of network architectures for multi-hop cellular and multi-hop wireless LAN have been proposed [43-52], research on DRT establishments for multi-hop cellular and multi-hop LAN is very weak. It is a critical issue for a multi-hop MS to establish DTRs in a multi-hop WLAN or a multi-hop cellular network.
- The private/public key-based DTR protocols [59] cause heavy computation delay as compared to secret key-based HMAC. It is necessary to develop secret key-based DTR between a multi-hop MS and the visiting network.

- How to facilitate the process of DTR establishment and provide enough protection in a multi-hop registration.

2.3.2 Infrastructure Protection

In case of the Internet infrastructure, a detailed investigation of the possible attacks and existing solutions can be found in [56]. Wireless infrastructure including cellular BS and WLAN AP serves as the communication bridge connecting the multi-hop wireless network and the Internet. With regard to the ICWM security, each BS typically runs as the centralized authority and administrative center to provide security support for single and multi-hop MSs. A single or multi-hop MS establishes a DTR with the deployed BS and creates a security association. Moreover, a BS also provides support for route management [43-52] to enable the process of multi-hop route discovery for multi-hop MSs. To accomplish these functions, the BS should be sufficiently powerful to defend itself against security threats from the Internet or wireless networks. However, BS is vulnerable to various threats such as eavesdropping and traffic analysis, Internet-Wireless-DDoS, Wireless-Internet DDoS, forged BS, and BS cache poisoning. As a result protocols must be implemented to withstand attacks that can lead to failure of the BS. The BS/AP protection primarily includes access control, and resource and identification protection.

- **Access Control**: A BS may be accessed and reconfigured according to the attacker's convenience. The multi-hop routing or other information stored in a BS may be read or modified by an adversary. The sensitive information such as secret keys may be disclosed. Thereby the BS should have the capability of security access control scheme to prevent unauthorized access and operations.

- **Resource Protection**: Another mandatory requirement is to improve the capability of resource (radio spectrum) management and resilience to the junk packets. Internet-Wireless DDoS, Wireless-Internet DDoS and attacking BS radio spectrum can be used to drain the radio spectrum resource in a BS by a batch of junk packets. The network (BS) should have the capability to identify the authenticity of packets for the purpose of filtering trash packets either from wireless networks or from the Internet.
- **Identification Protection**: It is necessary to provide identification protection for critical BSs. Identification information includes BS's IP address, physical location and physical link connection, etc. By using a directional antenna with a Global Positioning System (GPS), the adversary not only knows that there are BSs in that area but also know its physical position. An adversary can easily know the IP address of a BS by capturing an IP packet from the BS either on the Internet or by wireless eavesdropping.

The personal communication device, such as a multiple air interface MS, should be properly protected. The identification (IP address), secret keys, certificates and other sensitive information should be stored in a proper manner so as to prevent getting compromised and running as a malicious MS. As for a multiple radio interface MS, the MS should properly manage the resource related to different air interfaces such as home IP address and the transition between different air interfaces.

Open Research Issues

A BS or a MS is often required to execute multiple distinct protocol standards so that security operation for various protocol layers and inter-working among different networks (e.g., cellular and wireless LAN) could be easily supported. The design of an efficient BS or a MS platform with security enhancements is still open for further investigation.

- Development of security specific platform for a BS or a MS to accommodate the implementations of different security protocols.
- Improving the security of the local wired networks, where the BSs are located, and enhancing the BS security features to prevent unauthorized operations on the BS configuration and system sensitive data.
- Speeding up the MS's processing capability and optimizing cryptographic algorithms to increase the security operating capability of the MS.

2.3.3 Multi-hop Routing Security

Different routing protocols [3-10] [43-52] have been developed to implement the procedure for MSs to effectively discover and maintain routes in ICWM networks. There are three types of basic routes in ICWM networks: the route from a BS to a MS, the route from a MS to a BS, and the route between two MS without BS. The notorious issue in the multi-hop network could be said to be the multi-hop routing security. Corresponding to the three types of routes, multi-hop routing security should provide security to all the above types of routes.

As illustrated in Section 2.2.1, a major advantage of the ICWM routing security in comparison with the MANET security is the accessibility of the Internet through the BS, thereby having a mechanism for centralized security management. The MANET security has no explicit line of defense to distinguish a MS as trusted or non-trusted due to the absence of infrastructure-based authority. On the contrary, in an ICWM network, the

Internet infrastructure and BS acts as the security authority for multi-hop routing. For instance, a BS can verify the credentials of a single or multi-hop MS when required. By way of registration and authentication, the MSs can be divided into: registered MS and unregistered MSs. It is necessary to execute the required principles to enforce registered MSs to participate in honest route discovery and maintenance and to exclude the unregistered MS from the routing paths. In order to achieve authentication, confidentiality, integrity, cooperation and availability in routing security, some proposed fundamental principles [64] [59] are as follows.

- **Uniqueness**: Identification uniqueness even if a MS enters an ICWM network many times.
- **Access Control**: An ICWM network only provides routing and communication services to the trusted MSs which have established the DTR with the network.
- **Exclusion**: Only the trusted MSs can serve as the intermediary relaying MS in a multi-hop route and not-trusted MSs are excluded from the route discovery.
- **Authentication**: Only the authenticated routing messages can be used to update the BS caching.
- **Integrity**: An appropriate cryptographic scheme to effectively protect routing messages by way of encryption/decryption or signature/verification.
- **Credentials**: The home network has the capability to operate MS's billing and credential when a MS has any malicious action.

In the integrated Internet and MANET [59], the FA acts as the distributed CA (certificate authority) server. Thus it issues a certificate to an ad hoc MS which has a successful DTR process. The certificate is uniquely bound with the MS's home address.

The certificate is also used as a pass for accessing the network as well as the authentication of its routing messages. Therefore, any MS without a legal certificate is excluded from any multi-hop route discovery process. On the other hand, the CAMA security implements these principles for cellular-assisted WLAN networks by a set of keys such as secret key, public/private key, group key, and session key, etc.

The next challenge is to execute the above defined principles across wireless and wired network. The enforcement of the defined principles is a complicated procedure, spanning different layers in the protocol stack and crossing multi-hop wireless network and wired network. In the transport layer, DTR establishment provides multi-hop MSs with security authentication and association based on certification service as specified in the public key infrastructure (PKI) or other security technologies. One of the immediate results is that the process creates a binding that envelope the MS's identification with the keys used for route discovery at the network layer as well as the keys used for encryption at the link layer. Only the MS having a legal binding is marked as a trusted MS on the network layer. Also, the encryption at the link layer is needed for the purpose of confidentiality. During the execution of the above defined principles, an appropriate identification binding procedure is required to prevent a non-trusted MS from participating in routing activities. On the other hand, the process of route discovery across the multi-hop network should be executed without a high control packet overhead and computation of each MS. The security in the routing maintenance stage should be able to accommodate fast network topology changes. In the integrated Internet and MANET, the execution is divided into five stages [59]: key creation, FA discovery, secured registration, certificate issuance, and secured route discovery.

Another important challenge is the evaluation of the aforementioned principles and their executions. The measure need to include quantitative and qualitative evaluation of the security and performance of the defined principles. An important task in security evaluation is to check whether vulnerabilities are still incurred after defining and running the principles on the multi-hop network. The evaluation can be used for the further enhancement of routing security principles and execution until routing security protocol can have adequate robustness.

Open Research Issues:

In an ICWM network, MSs are generally cheap devices having less computation capability. The exchange of routing message with digital signature is based on asymmetric key cryptography (e.g., RSA) and the routing process consequently involves much more computation overhead in signing/decrypting or validating/encrypting operations. If a malicious MS feeds a MS or BS with a large number of bogus routing messages with false signatures, it can easily exhaust the computation resource of a MS or BS. The open challenges for routing security for multi-hop cellular networks are:

- Defining, executing and evaluating desirable security routing principles,
- Developing secret key-based HMAC (message authentication codes) routing security to facilitate the efficiency of route discovery and maintenance, and
- Developing a protocol for DTR establishment and secured route discovery in supporting other different routing protocols (i.e., AGSM, UCAN).

2.3.4 Securing Packet Forwarding

Because of multi-hop communication, an intermediate relaying MS may correctly participate in a route discovery but improperly forward data packets in an ICWM network. There are several challenges [61-71] in secured multi-hop packet forwarding. Few of these are (i) detecting and reporting misbehaviors, (ii) penalizing or isolating malicious MSs, and (iii) charging and rewarding scheme to encourage the packet forwarding.

Several selfishness prevention schemes have been reported since Marti et al. [61] proposed a method to detect misbehaving nodes in August 2000 by dividing the strategies into three groups: *reputation-based scheme*, *credit-payment scheme*, and *game theory scheme*. As the ratio of selfish MS increases, the packet delivery rates get drastically reduced. To detect misbehaviors, an approach in [61] proposes a scheme where all ad hoc MSs maintain a set of matrices to record the past misbehavers of other MSs, and a MS selects well-behaved MSs as the intermediate relaying MS for packet forwarding. In addition, each MS [62] creates a secure association in order that it can authenticate misbehavior reports before isolating a malicious MS. On the other hand, a set of rules and protocols [66] are proposed as incentive for packet forwarding by the exchanges of fictitious currency called Nugget. In the ICWM context, which is infrastructure-based packet forwarding, BS collects the report of the misbehaviors including packet dropping, reduplication of packet, packet filtering, and packet modification etc [65].

Based on the reports from multi-hop network, the network (BS) decides which accounts should be charged or credited. To implement misbehavior reporting, two schemes are needed: authentication and packet acknowledgement [65]. An authentication scheme provides the capability for (i) authentication of the source and destination, (ii) authentication of the forwarding MS. The authentication of any message depends on the DTR establishment from which the necessary authentication keys are created and uniquely bound with the MS. An acknowledgement scheme is the process for destination to acknowledge the reception of every packet either in per packet basis or batch basis. The BS can maintain the records of all activities of packet forwarding for each participating MS. The destination reports to the BS every time when it receives a data packet from the source [65]. In order to save resource, the destination acknowledges all received packets in a single batch at the end of packet transmission. Such credit-payment approach requires global participation and a variety of heterogeneous MSs may exist in a practical MANET. Furthermore, some MSs may not earn enough credits not as they may just be badly positioned. Raghavan et al. [68] address this by proposing two layered forwarding service: *priced priority forwarding* and *free best-effort forwarding*.

A charging and remuneration scheme manipulates the accounts of all participating MSs. The manipulation of account depends on several factors: the number and size of packets, the number of hops of wireless route, and the transmission time. In the process of charging and remuneration, it is necessary to protect the network from cheating, refusal to pay, incorrect reward claims, free-riding, and invasive adversary. The trade-based scheme in [58] enforces the cooperation in mobile ad hoc WANs by two approaches: packet purse and packet trade. The natural idea of the trade-based scheme is that a MS used a service should be charged and a MS providing a service should be remunerated. The charging and remuneration scheme in [65] achieves the collaboration of packet forwarding in multi-hop cellular networks. Upon receiving the upstream packets which are from a source MS in the multi-hop network, the BS credits their collaboration after authenticating each relaying MS. In the opposite direction of upstream packets, the destination MS in the multi-hop network acknowledges the receipt of packet from the BS, and then BS makes the credits for each relaying MS. Sprite (a simple, cheat-proof, credit-based system) [69] with a central authority server called CCS (Credit Clearance Service) also utilizes credit to give incentive to the MSs that forward packets to the source. However, the amount of credit charged to the source is not always equal to that given to other MSs. Yau et al. [70] identifies many problems which make CONFIDANT and CORE impractical to be used in actual MANETs and propose a simple reputation system to enhance the robustness of a MANET by MS utilizing only its own experience about all neighboring MSs. Huang et al. [71] address drawbacks and impracticality of both the reputation and the credit-payment incentive methods.

Open Research Issues:

The infrastructure-based scheme for detecting various misbehaviors is an important issue in ICWM networks but has been neglected in current security designs. A lot of issues about securing packet forwarding in ICWM network remain unexplored:

- How to enforce service availability and cooperation in a ICWM network with a secure mechanism to stimulate MS to participating in packet forwarding, to refrain from overloading the network, and to thwart the "selfish" MS.
- How to implement fair charging and rewarding for the cooperation between MSs in packet forwarding protocol and a reasonable fine for misbehavers.

2.3.4 Interworking Security

A dual-mode MS equipped with two radio interfaces discovers and selects an access network (GSM, GPRS, UMTS, and WLAN) from the available networks. The interworking of 3GPP/WLAN enables the provisioning of public WLAN access service to 3GPP system subscribers [1] [55]. Interworking security is a basic prerequisite for providing a cellular subscriber MS with WLAN IP Internet connectivity.

Figure 8 illustrates simplified cellular/WLAN interworking security architecture [54] [59]. When a cellular MS visits a foreign WLAN network, the 3GPP AAA interconnects WLAN security system and cellular (e.g., GSM, UMTS) security system. To reuse SIM/USIM-based authentication in WLAN interworking, two new EAP schemes, EAP SIM and EAP AKA, have been specified for 3GPP-WLAN interworking. The EAP SIM encapsulates GSM SIM authentication and key agreement algorithms within EAP. And the EAP AKA specifies UMTS mutual authentication and key agreement (AKA) on EAP.

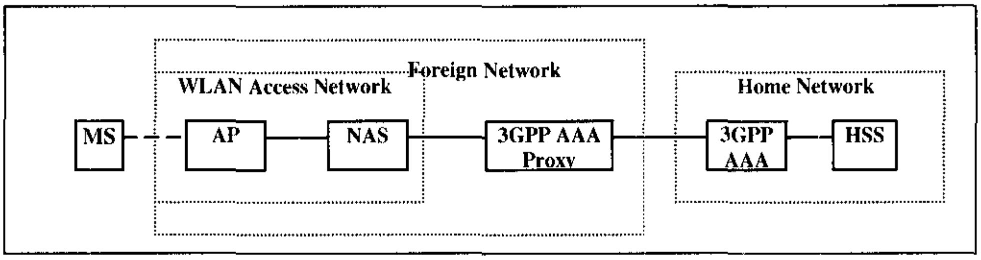


Figure 8. Security Architecture for Cellular/WLAN Interworking.

Open Research Issues:

The AAA framework is chosen as the basis of the WLAN and 3GPP-WLAN. The possible AAA protocols include RADIUS or Diameter. Many challenges remain in securing the cellular/WLAN interworking:

- Extending AAA protocols to allow cellular (e.g., GSM, GPRS and UMTS) MS roaming from cellular system toward the WLAN system.
- Instead of GSM family system, how to support WLAN interworking such as IEEE 802.16 in which X.509 certificate is the basis of access security for MSs.
- Since the proposed 3GPP-WLAN can not support multi-hop access, protocols are needed for the interworking security system to support multi-hop communication and routing.

2.4 Conclusion

Many possible security threats against ICWM networks have been presented in this chapter, along with other open issues. The security attacks such as wireless-Internet DDoS may seriously degrade the performance of Internet services for multi-hop MSs. However, the current security solution for ICWM networks is only the first step toward tackling various security attacks. In this chapter, the ICWM security challenges of such schemes are discussed for providing ubiquitous and secure Internet services for mobile users in support of heterogeneous multi-hop communication. For the practicality of ICWM networks, robust and efficient security schemes have to design across heterogeneous multi-hop wireless networks and the Internet.

PART A

ON SECURE COMMUNICATION FOR INTEGRATED INTERNET AND MANET

CHAPTER III

AN INTEGREATED PROTOCOL FOR INTERNET AND AD-HOC NETWORK COMMUNICATION

The integration of infrastructure networks and ad hoc mobile networks can be used to eliminate dead zones in wireless LANs and cellular networks, and can also be used to extend the coverage of wireless networks. Earlier approaches for the integrated networks either provided only unidirectional global connectivity for ad hoc hosts or caused high overhead and packet delay for full bi-directional connections. In this chapter, a protocol of integrating mobile IP and enhanced DSDV (EDSDV) is proposed to provide full bi-directional Internet connectivity for ad hoc hosts. A detailed performance comparison is conducted between the proposed approach with a leading integration approach based on different network sizes, mobility of ad hoc hosts and other network parameters. The experimental results show that the proposed approach has smaller overhead for full bi-directional connectivity and shorter average packet delay compared to the leading integration approach.

The spectrum of wireless networks can be classified into infrastructure (e.g., wireless LAN and wireless MAN) networks and infrastructureless (ad hoc) network. In infrastructure networks, a base station or an access point serves as a communication bridge for mobile hosts. Mobile users obtain their services under the coverage of a base station or an access point. A base station or an access point has both a wireless and a wired interface and acts as a gateway between two types of networks [25]. Mobile IP has been successfully extended to cellular networks and wireless LANs for providing global internet connectivity [3] [4]. In contrast to infrastructure networks, ad hoc nodes are mobile terminals connected in a dynamic and arbitrary manner without any infrastructure device. The existing wireless LANs and cellular networks only provide a single hop wireless extension to access points and base stations. Integrating ad hoc networks with infrastructure networks can be beneficial for extending the coverage of wireless networks via multi hop communication of ad hoc hosts. Many communication scenarios may benefit from multi-hops networks with the desirable properties of larger coverage and freedom of mobility. Integrated ad hoc and infrastructure networks can be used to eliminate dead zones in wireless LAN. This chapter proposes an efficient approach that provides bi-directional Internet connectivity to ad hoc hosts via mobile IP protocol implemented on infrastructure networks.

Two important issues for providing the global Internet connectivity for ad hoc networks are bi-directional communication and the total overhead of Mobile IP as well as ad hoc routing protocol. Previous approaches [3] [4] [5] [6] [7] [8] [25] for the integrated Internet and ad hoc networks were developed by extending the mobile IP to ad hoc networks. But these approaches cause high mobile IP and ad hoc network routing overhead for full bi-directional connectivity. The approaches discussed in [5] [7] [25] proactively flood agent advertisements through the entire ad hoc network. In approach [5], the ad hoc hosts run a modified Routing Information Protocol (RIP) Daemon (*routed*) to deal with mobile IP messages. Also a mobile IP routing daemon (*mipd*) runs on each ad hoc host. The *routeds* of mobile hosts within the coverage of a foreign agent (FA) receive FA's agent advertisements and related mobile IP messages. These *routeds* can relay the mobile IP messages to the *mipds* of the ad hoc hosts that are outside the transmission range of the FA. The MIPMANET [7] provides Internet connectivity for ad hoc hosts by using mobile IP with foreign agent's care-of-address and reverse tunneling. FA periodically broadcasts its agent advertisements. Agent advertisements spread through the whole ad hoc network. This approach uses AODV [16] protocol for communication in ad hoc network. The performance of this approach has been improved by more recent approaches given in [8] [25]. The approach used in [25] extends mobile IP to ad hoc networks using on-demand AODV protocol; therefore each mobile node registers with FA only when it requires internet connectivity. However, on-demand approaches only allow unidirectional Internet connectivity from the ad hoc hosts to wired network hosts. Implementing bi-directional connectivity with these routing protocols incurs a significant overhead in maintaining IP registration with the FA. The scheme in [8] further improves the performance of the integrated approach given in [25] by lowering the mobile IP overhead. This approach reduces the number of FA advertisement by using controlled flooding among the ad hoc nodes. However, the mobile IP overhead and related AODV overhead are still very high [8]. The approaches in the [9] [72] involve agent discovery which results in more overhead for full bi-directional connectivity and introduces significant connection delay.

The full bi-directional Internet connectivity for mobile user is very important especially in the integrated Internet and ad hoc networks. The implementation of bi-directional connectivity has not been clearly addressed in the existing approaches. Moreover, the existing approaches have higher packet latency because they are mostly using either AODV or some extension of AODV. These approaches have high route discovery latency compared to the proposed approach. The integration of the mobile IP with table-driven (proactive) protocols, lends itself very well for providing bi-directional internet connectivity: either a host in infrastructure (wired) network or a host in the ad hoc network can initiate a global connection. In this chapter an enhanced DSDV (EDSDV) protocol is proposed that improves the performance of standard DSDV protocol. In the proposed integrated approach FA acts as one of the ad hoc nodes participating in the enhanced DSDV routing protocol. Also the FA serves as the mobile IP proxy for ad hoc hosts.

The rest of this chapter is organized as follows. Section 3.1 describes the basic model of combining ad hoc and infrastructure (wired) networks. Section 3.2 discusses why DSDV protocol [11] has poor throughput at higher mobility of ad hoc hosts. It then discusses an EDSDV protocol for overcoming the problem of lower throughput. In addition, this section proposes a protocol for full bi-directional connectivity by using FA as the mobile IP proxy for ad hoc hosts. Section 3.3 discusses the experimental configuration for integrating the infrastructure (wired) and ad hoc network. The experimental results obtained from the simulations under different scenarios are also discussed in this section. Some conclusions are included in Section 3.4.

3.1 Internet Connectivity for Ad-hoc Networks

The goal of the proposed approach is to integrate ad hoc and infrastructure (wired) network by using mobile hosts, which are located under the coverage of a foreign agent (FA), as the communicating bridges between two heterogeneous networks. Mobile IP and ad hoc routing protocol coordinate with each other to build the connectivity across the heterogeneous networks as shown in Figure 9. There are only two alternatives for a mobile host to obtain the global Internet connectivity. In the first case, ad hoc hosts, which are under the transmission range of a FA, can communicate directly with the FA. In the second case, ad hoc hosts, which are outside the coverage of a FA, communicate with the FA using multi-hop ad hoc links. The key challenge in providing connectivity is to minimize the overhead of mobile IP and ad hoc routing protocol between infrastructure and ad hoc networks. Figure 9 illustrates an example of cooperation between the wired network and the ad hoc network. The inside area of dash circle represents the coverage of a FA; as shown in Figure 9, mobile host 1 and 3 are located in the coverage of the FA. The left side of Figure 9 represents a wired network that consists of Home Agent (HA), Correspondent Node (CN), FA, and the Internet. The right side is the ad hoc network using EDSDV. The ad hoc host 1 and FA can reach each other directly. Assume that all mobile hosts work in a promiscuous receive mode, therefore, each mobile host would deliver every packet received from its neighbors without filtering any information. Meanwhile ad hoc host 4 and 5 are neighbors allowing ad hoc host 5 to communicate with the FA using ad hoc routing protocol via ad hoc host 4 and 1. In the end, a path can be established between CN and ad hoc host 5 via the Internet, FA, ad hoc hosts 1 and 4. Each ad hoc host registers with HA and creates a mobility binding at HA by mapping its home address to the care-of-address (COA) provided by FA. Upon receiving data packets sent by CN and tunneled via Internet, FA delivers the data packets to the ad hoc host 5 through the previously established path. On the other hand, after receiving data packets sent by the ad hoc host 5, FA delivers them to CN using IP routing in the infrastructure network.

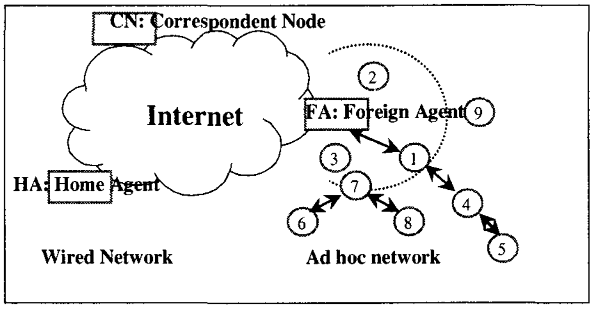


Figure 9. An Integrated Internet and Ad hoc Network.

3.2 Design of Integrated Framework

3.2.1 Motivation for Enhanced DSDV

The DSDV [11] routing protocol is based on classical Bellman–Ford Routing Algorithm for finding shortest paths between ad hoc nodes with some improvements. In DSDV protocol, each mobile host maintains a routing table that stores the number of hops and the sequence number assigned by the destination mobile host for all the destinations. The routing table updates in DSDV could be time-driven or event-driven. In time-driven routing table update, mobile hosts periodically transmit their routing tables to their immediate neighbors. The interval between two updates is referred as the DSDV periodic route update interval. On the other hand, in the event-driven routing table update, the mobile host notifies its routing information if a significant change has occurred in its routing table since its last update. There are two ways of performing routing update: “full dump”, in which an ad hoc host transmits the complete routing table to its neighbors, and “incremental update”, in which an ad hoc host sends only those entries from the routing table that have changed since the last update. A sequence number is used to distinguish a stale route from a new route to avoid the formation of loops in the routing process.

At higher rates of mobility in ad hoc networks, the performance of DSDV protocol is worse than many other ad hoc routing protocols. The packet delivery fraction can drop up to 70 percent with 50 nodes and a maximum mobility of 20 m/s (average speed 10). The low packet delivery rate results from the fact that an ad hoc host uses stale routes to a destination to send packets. Higher mobility rate results in higher number of broken links in an ad hoc network. These broken links result in the creation of stale routes in the routing tables. In our extensive simulation studies, we have found that packet dropping occurs due to traffic forwarding via a stale route generated by a broken link between two ad hoc hosts. Figure 10 illustrates how many packets are dropped during 600 seconds at a source node due to the broken links. The size of DSDV network is 50 nodes. A source node sends packets to a destination at the rate of 10 packets per second. As seen from Figure 10, the average number of packets dropped at the source node increases when the mobility of ad hoc increases. For example, at 50 m/sec mobility, 765 packets were dropped out of 6000 packets due to broken links at the source node.

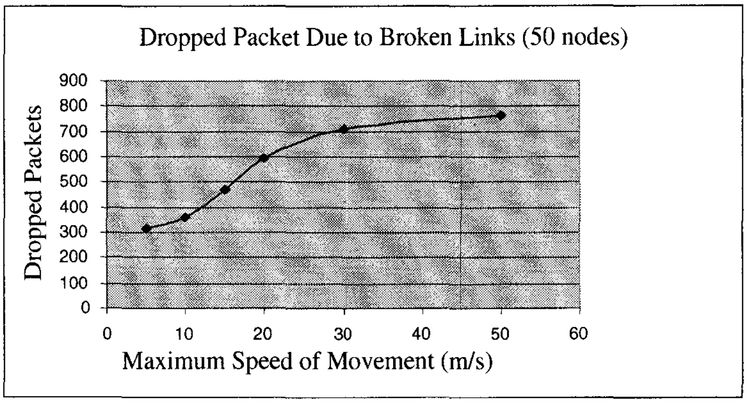


Figure 10. Number of Packets Lost due to Broken Links.

In DSDV, a stale route at an ad hoc node doesn't mean that there is no valid route to a destination. Packets can be forwarded to some other neighbors that may have routes leading to the destination. When an immediate link from an ad hoc host *A* to destination *D* is broken, the proposed EDSDV protocol creates a temporary link via a neighbor that has a valid route to the destination *D* in its routing table. In EDSDV protocol, a temporary link is created by one-hop ROUTE REQUEST and ROUTE ACK messages. The ad hoc host *A* having the broken next hop link to the destination broadcasts a one-hop ROUTE REQUEST message to all its neighbors. The neighbor returns a ROUTE ACK message if it has a route to the destination and the ad hoc host *A* is not the next hop on the route from the neighbor to destination *D*.

In the proposed EDSDV protocol, each entry in routing table has an additional element for recording the route update time. This route update time is sent in ROUTE ACK message and is used for choosing a temporary route. After ad hoc host *A* broadcasts ROUTE REQUEST, there may be more than one neighbor that responds with ROUTE ACK. The ad hoc host *A* chooses a route that has least number of hops to the destination and the latest update time. The EDSDV protocol maintains all the desirable features of standard DSDV but reduces the packet loss due to broken links. The EDSDV protocol uses the following approach to overcome the stale route problem of standard DSDV routing protocol:

- If the link from host *A* to next hop link in route to a destination is broken, then in-coming packets are buffered; the maximum packets buffered at an ad hoc host for each destination is *N*.
- When host *A* identifies that the next hop link in route to a destination is broken, the host *A* suspends forwarding packets to this next hop. In order to find a temporary next hop leading to the same destination, host *A* immediately broadcasts a one-hop ROUTE REQUEST to its neighbors. ROUTE REQUEST includes the host ID, and the destination. If a neighbor has a route leading to the destination in its routing table and the route does not take use of *A* as the next hop, then the neighbor responds by sending a ROUTE ACK message. ROUTE ACK includes its host ID, the destination, the hop count metric for the destination, and the last updated time for this path to destination.
- The host *A* chooses the best neighbor as the temporary next hop, and then resumes forwarding the buffered and in-coming packets via the temporary route. The logic of choosing the best neighbor for the ad hoc host is to select a neighbor with the least number of hops to the destination. If there are several next hop nodes leading to the destination with same hop count, then host A chooses the neighbor having the latest routing update time.
- Later, regular DSDV protocol updates the route to the destination in the routing table of host *A*. An updated route from the host *A* to the destination replaces the stale route. At this time host *A* switches from the temporary route

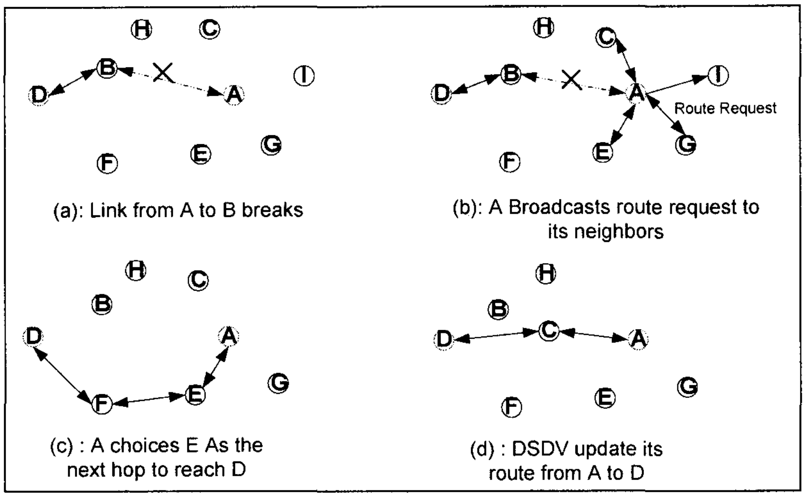
to the updated route for sending the in-coming packets. Also during creation of a temporary route, if the routing table entry from host *A* to the destination is updated by the regular DSDV protocol, then host *A* stops creating the temporary route and sends packets via the updated route.

Figure 11 illustrates how host *A* creates a temporary route to destination *D* after the immediate link from *A* to *B* is broken. When the link from host *A* to destination *D* is broken because of its own or host B's movement, host *A* suspends sending packets (Figure 11 (a)). Then host *A* immediately broadcasts a ROUTE REQUEST to its neighbors(*C, E, G* and *I*). As an example, Table 2 represents the current routes leading to destination *D* at each neighbor. Because the next hop in routing entry of the neighbor *I* is *A*, the ad hoc host *I* does not respond to *A*'s request. Ad hoc hosts *C, E,* and *G* respond with ROUTE ACK messages along with hop count metrics and the route update time to ad hoc host *A* (Figure 11 (b)).

Table 2: The snapshot of current routes

Neighbor	Metric (hops)	Next Hop	Route	Route updated time
C	2	H	C->H->D	1005
E	2	F	E->F->D	1010
G	3	E	G->E->F->D	905
I	3	A	I->A->B->D	1002

The ad hoc host *C* and *E* have the same value for hop count metrics, but the latest routing update time for *E* is greater than that of C, which indicates that the path through E is updated more recently. So host *A* chooses *E* as the next hop to the destination *D*. Host *A* resumes sending packets to the destination *D* (Figure 11 (c)). Therefore, the packets in host *A* are forwarded to the destination *D* via *E* and *F*. After a while, the route entry in host *A* to the destination *D* is updated by the regular DSDV process, then host *A* switches its route from the temporary route to the updated route. In the Figure 11 (d), *C* moves close to *D*, and the DSDV routing process updates the route at *A*. Therefore, ad hoc host *A* uses the updated DSDV route <A->C->D> for sending packets to *D*.



(a): Link from A to B breaks

(b): A Broadcasts route request to its neighbors

(c) : A choices E As the next hop to reach D

(d) : DSDV update its route from A to D

Figure 11. Creating a Temporary Route in Node *A*.

3.2.2 Integrated Mobile IP and Ad hoc Network

Mobile IP protocol [33] [34] provides a continuous connectivity for mobile hosts, in which discovery, registration, and tunneling processes are involved. In mobile IP mechanism, a mobile host uses a fixed home address. The home address is bound to a care-of address (COA) provided by a FA while the mobile host visits a FA. This mapping is available at home agent for forwarding packets to FA which can deliver packets to a roaming ad hoc host. However, the role of FA is extended in the proposed approach:

- FA takes part in the routing protocol just as other ad hoc hosts do.
- FA acts as the mobile IP proxy for the ad hoc hosts. The detailed information on registration process for ad hoc hosts is provided in the following sections.

In the proposed integration approach, FA participates in ad hoc routing protocol as ad hoc hosts, thus FA doesn't broadcast agent advertisements for the purpose of integrating ad hoc hosts. This FA broadcast has been a significant source of overhead in earlier approaches [3] [4] [5] [6] [7] [8] [25]. In the proposed EDSDV protocol, ad hoc hosts and

FA automatically know each other's presence via routing update EDSDV protocol. Also each host doesn't send solicitations to request care-of-address from FA. When a mobile host *H* joins the ad hoc network, the host *H* broadcasts DSDV advertisements to its neighbors for the first time with a sequence number of 0. Each neighbor of the host *H* inserts a route entry in its routing table for the host *H* and broadcasts immediately with an increased sequence number to its neighbors. The process continues until the advertisement of the host *H* has reached all the destinations. Eventually the FA inserts a route entry for the host *H* also. At the same time, the host *H* gets a routing table from each of its neighbors and creates its own routing table. The route to FA is also included in the host *H*'s route table via routing update messages from its neighbors. Then the host *H* sends FA the registration information, for example, *H*'s home address. Based on the registration information, FA acts as the mobile IP proxy for the host *H*. FA sends the registration request to the host *H*'s home agent. After successful registration at the home agent, a registration reply message is returned from HA to FA. FA then informs ad hoc host *H* about its registration status. In the proposed approach FA keeps the registration information for all ad hoc hosts and uses it again during re-registrations. FA uses its foreign agent address as COA to register with the HA for each ad hoc host. The mobile IP registration lifetime for each mobile host is 2 to 3 times of the DSDV periodic route update interval.

In the proposed approach, FA serves as the mobile IP proxy for each ad hoc host. Each ad hoc host registers with FA only once when it joins the network. Mobile IP maintenance in the proposed approach includes two parts:

- FA re-registration or deregistration for each ad hoc host,

- Ad hoc host location management.

FA Re-registration or deregistration: Each ad hoc host doesn't participate in the mobile IP registration after exchanging registration information at the moment of joining the ad hoc network. Since FA keeps all the registration information for each node, FA updates the registration for each ad hoc host before its expiration only if an ad hoc host has a valid path from FA. FA has the consistent and up-to-date routes for each ad hoc host via EDSDV so the FA surely knows whether or not an ad hoc host is part of the mobile network. In case the hop count metric for an ad hoc host in FA's routing table becomes infinite, it implies that FA has lost its route to that ad hoc host. It may mean that the ad hoc host has roamed away from the network. If it loses the route to an ad hoc host for 2 times of the periodic EDSDV route update interval, FA assumes that the ad hoc host has moved away from the ad hoc network. The FA sends the ad hoc host's home agent a deregistration message to notify the home agent that the ad hoc host has disconnected with the FA. If the host joins the ad hoc network again, the host needs to resend its registration information to FA.

Ad hoc host location management: The location and route to each destination ad hoc host can be identified from an ad hoc host's routing table. If the route from an ad hoc host to FA is reachable (the hops from the ad hoc host to FA is finite), it means that the global connectivity is possible to the ad hoc host. If a mobile host roams away from the ad hoc network, it will be reflected in routing table of the mobile host because the hop count metric of FA will be infinity.

3.2.3 Bi-directional Internet Connectivity

In order to provide bi-directional communication, the framework allows the following three types of interaction between ad hoc hosts and a CN in Internet:

- Intra-MANET communication
- Inter-MANET communication from ad hoc hosts to CN
- Inter-MANET communication from CN to ad hoc hosts

For the first type of interaction, the ad hoc hosts keep the up-to-date routing information about other hosts in ad hoc network through EDSDV. To communicate with another host, an ad hoc host first checks its routing table. If the destination is inside the ad hoc network, a routing entry leading to the destination will be found in the routing table and packets will be forwarded to next hop in the route table according to EDSDV protocol.

The second type of communication occurs between an ad hoc host and a correspondent node on the Internet, initiated by an ad hoc host. In Figure 9, communications between an ad hoc host 5 and a CN, initiated by ad hoc host 5, is an example of this type of communication. In order to communicate with a CN on the Internet an ad hoc host checks its routing table. If no routing information to the CN is found, then the ad hoc host checks the routing to a FA. If the routing entry to a FA is found, then the packets will be forwarded to that FA. Otherwise the packet is discarded. When a FA receives packets from an ad hoc host for the CN, it forwards those packets to the destination via Internet following the Internet IP routing protocol.

The third type of communication occurs between a CN and an ad hoc host, initiated by the CN. For example in Figure 9, communications between the CN and host 5 is initiated by CN. If the CN wishes to send packets to an ad hoc host, the packets will be delivered to HA that the ad hoc host visits. If the ad hoc host registers with HA domain, HA checks its routing table, finds a route to the ad hoc host, and then forwards the packets to the ad hoc host. If the ad hoc host is roaming away from its HA, HA maintains the ad hoc host's current location via FA registration. Using COA of FA, HA forwards the packets to FA visited by the destination ad hoc host. The FA checks it's routing table, and delivers the packets to the requested destination ad hoc host via EDSDV routing protocol.

Table 3: Simulation Parameter for EDSDV Implementation

DSDV Periodical route update interval	15.0 seconds
DSDV Periodic update missed before link declared broken	3.0 seconds
Initial triggered update weighted setting time	6.0 seconds
Weight settling time weighting factor	7/8
Routing advertisement aggregation time	1 second
Number of times that a ROUTE REQUEST may be resent	2
Time before a new ROUTE REQUEST is sent	1.0 second

3.3 Simulation and Experimental Results

3.3.1 Experimental Configuration

The effectiveness of the proposed integration approach is demonstrated by carrying out extensive experiments, in which FA acts as mobile IP proxy and combines the EDSDV protocol with Mobile IP routing. The experiments are conducted using NS-2 [73] simulator. Different sizes of ad hoc networks (20 and 50 nodes) were tested. In the integrated network, FA is configured as an ad hoc node as well as mobile IP proxy for ad hoc hosts. Table 3 lists the constants used in implementing EDSDV protocol in the simulation. In Table 3, the first five parameters are same as the parameters in paper [24]. When a next hop link from an ad hoc host *A* to destination *D* is broken, the ad hoc host *A* sends ROUTE REQUEST message to its neighbors. The decision for temporary next hop is based on the responses received from the two rounds of ROUTE REQUEST messages. If there is no response after 1.0 second another ROUTE REQUEST is sent to the neighbor.

The maximum mobility speed of ad hoc hosts is set between 1 to 50 m/s during the lifetime of simulation runs according to a uniform random distribution. For example, if the maximum speed of nodes is set at 20 m/s, the nodes move randomly ranging from 0 to 20 m/s; the average speed is 10 m/s. The direction of movement is also changed according to the random waypoint model [6]. The pause time is consistently 10 seconds between each movement. Constant bit rate packets are sent for 600 seconds during the simulation. The ad hoc network size is varied in order to keep consistency in node density. Experimental scenarios are set with the following dimensions:

- 670m x 670m simulation area with 20 mobile nodes
- 1000m x 1000m simulation area with 50 mobile nodes

Instead of just unidirectional global connectivity for ad hoc hosts [3] [4] [5] [6] [7] [8] [25], the proposed approach achieves full bi-directional connectivity. The experiments here show the full bi-directional communication where CN initiates three calls to three randomly selected ad hoc hosts, and two randomly selected ad hoc hosts initiate two separate calls to CN also. The two randomly selected ad hoc hosts send Constant Bit Rate (CBR) packets to CN. In addition, CN sends CBR packets to three ad hoc hosts selected to receive packets from CN. The CBR packet size is set as 512 bytes and does not include the packet head. Figure 12 illustrates the experimental configuration. A router, connected to HA, CN and FA, represents the delay on the Internet. The FA is located in the center of experimental domain. The dash circle represents the coverage of FA. For a mobile host located in the coverage of FA, a bidirectional connection can be established directly between the FA and the mobile host. FA and wireless ad hoc hosts use the EDSDV protocol. For each sampled data point, NS-2 was executed five times under different random mobility scenarios and each data point shown in the graphs is an average of five results.

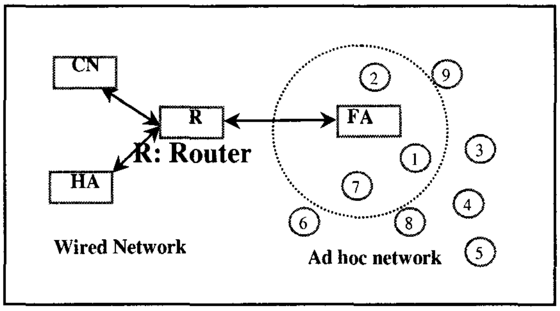


Figure 12. Experimental Configuration.

3.3.2 Experimental Results and Comparisons

The performance of proposed integrated ad hoc network is analyzed based on the following parameters:

- Packet Delivery Fraction (throughput)
- Mobile IP overhead and the related ad hoc protocol overhead,
- Packet Latency

The experimental results obtained for the proposed integrated approach are compared with the leading hybrid approach [8]. In order to lower the mobile IP and related AODV overhead in ad hoc networks, the hybrid approach [8] sets the TTL-field as *N* in the IP header of the advertisements to limit the flooding of advertisements. Advertisements are only spread within the *N*-hop neighborhoods directly. Instead of flooding advertisements through the whole ad hoc networks, any ad hoc host in the hybrid approach can eavesdrop and cache fresh agent advertisements for the mobile IP purposes. In the hybrid approach [8], the ideal TTL value of 2 is used and a best FA beacon interval of 10 seconds is chosen. In the proposed approach in this chapter, it is not necessary to choose a TTL value and the beacon interval because there are no FA advertisements in the ad hoc network. The comparison of performance is based on the 50 node network with different movement patterns as given in [8]. All the results are compared against the results obtained using ideal TTL and the best beacon interval suggested in the hybrid approach [8].

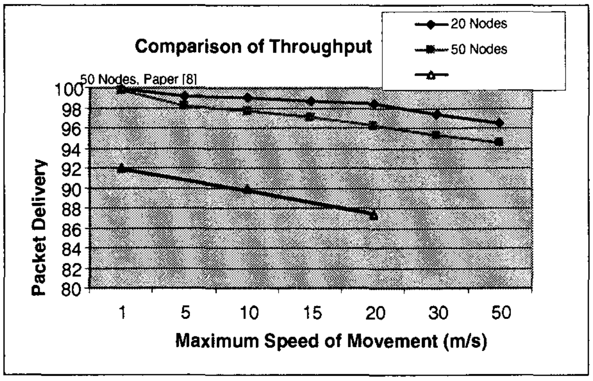
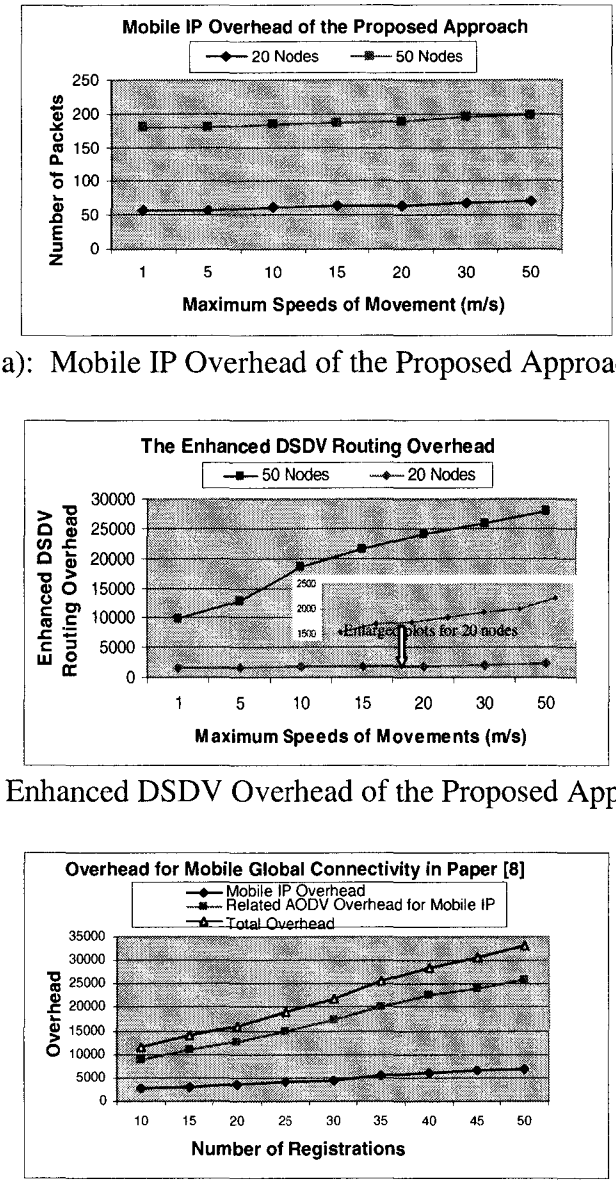


Figure 13. Comparison of Throughput.

Figure 13 illustrates the results of packet delivery fraction in the proposed approach and in the hybrid approach [8]. The impact of mobility on the packet delivery fraction was tested by varying the maximum movement speed of the ad hoc hosts. The mobility of ad hoc hosts is varied from 0 m/s to 50 m/s in the proposed approach. As seen in Figure 13, the packet delivery fraction goes down as the mobility increases. Figure 13 also presents the throughput of the hybrid approach with 50 nodes in which the FA beacon interval is 10 seconds. The throughput of the proposed integrated approach is better than that of the hybrid approach.



(a): Mobile IP Overhead of the Proposed Approach

(b): Enhanced DSDV Overhead of the Proposed Approach

(c): Overhead of the Hybrid Approach (50 nodes, 20 m/s)

Figure 14. Comparisons of Overhead.

Figure 14 compares the overhead of the proposed approach with the hybrid scheme given in [8]. It is evident from Figure 14 (a) that the overhead of mobile IP in the proposed approach is very low. This is because ad hoc hosts send their registration information to FA only at the moment of joining the ad hoc network. Figure 14 (b) shows the overhead of EDSDV for maintaining the routing tables at ad hoc hosts. The overhead increases when the mobility of ad hoc host increases.

To maintain the global connectivity for ad hoc hosts, the hybrid approach [8] results in mobile IP overhead and related AODV overhead. The related AODV overhead is used for establishing the route between an AODV node and FA to exchange mobile IP messages. Figure 14 (c) presents the mobile IP and related AODV overhead in the hybrid approach [8] with different numbers of registrations. The related AODV overhead in Figure 14 (c) is only the AODV routing overhead generated by exchanging mobile IP messages between ad hoc hosts and FA for maintaining Internet mobility of ad hoc host. The overhead of the AODV routing caused by transmitting data packets is not included in Figure 14 (c). Figure 14 (c) also shows the total overhead to maintain global connectivity for different registered ad hoc hosts (from 10 to 50 nodes). The size of network is 50 nodes, and the TTL is 2. The maximum speed of movements is 20 m/s. The number of registered ad hoc hosts varies from 10 to 50 nodes. Therefore bi-directional connectivity is provided for 10 to 50 nodes in accordance with the number of registered nodes. For instance, if the number of registered ad hoc hosts is 30 out of 50 nodes, it only provides 30 nodes with bi-directional Internet connectivity. To provide data packet transmissions, the hybrid approach needs more AODV routing overhead to maintain the path from a source host to a destination host. On the contrary, Figure 14 (a, b) show the overhead incurred due to Mobile IP and the EDSDV protocol.

It is clear from Figure 14 that the total overhead of the proposed integrated approach is lower than that of the hybrid approach developed in [8]. If the hybrid approach keeps the full bi-directional connectivity for all ad hoc hosts (50 nodes), the total overhead will be much higher than that of the proposed approach. The following conclusions can be derived by analyzing the overhead of 50 nodes:

- The total overhead for maintaining full connectivity for ad hoc hosts in the hybrid approach [8] is higher than that of the proposed approach.
- In addition to the overhead for maintaining full connectivity for ad hoc hosts, extra AODV routing overhead is needed for communication between two ad hoc hosts in the hybrid approach [8]. The proposed EDSDV protocol causes no extra overhead for communication between two ad hoc hosts.
- If AODV hosts don't register with FA, CN cannot reach these AODV hosts. Thus, the AODV protocol will require additional overhead for full bi-directional connectivity.

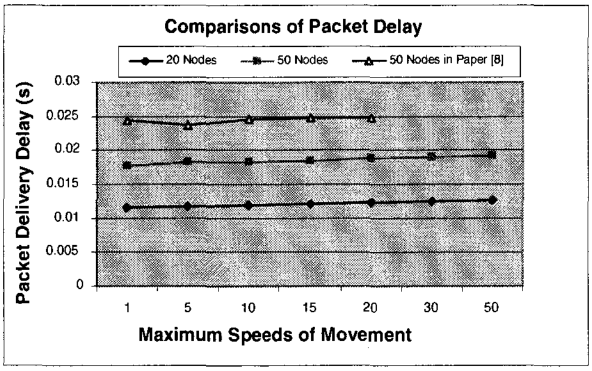


Figure 15. Comparisons of Packet Delivery Delay.

Figure 15 compares the average packet delay from ad hoc network to CN in the two approaches. Figure 15 shows the packet delay in the hybrid approach where TTL is 2. It can also be concluded from this study that the average packet delivery delay for the proposed approach is better than the hybrid protocol [8]. The reasons for better packet delay are:

- The EDSDV protocol uses better routes than AODV so that packet can quickly reach their destination.

- The EDSDV does not need time for Route Discovery before communications since the routes already exist between the source and destination; but in the case of AODV, the route is created on demand.

3.4 Conclusion

The proposed EDSDV protocol overcomes the stale link problem that degrades the performance of standard DSDV protocol at higher rate of mobility of ad hoc host. This chapter also presents a scheme for providing full bi-directional Internet connectivity for ad hoc networks. The global Internet connectivity is especially important when it is used to extend the coverage areas of wireless LAN infrastructure, and to improve the services in poor coverage of wireless infrastructure in the cellular network. In this paper, instead of flooding the FA advertisement through the whole ad hoc networks for registration protocol, FA acts as the mobile IP proxy for ad hoc hosts. The simulations show that the scheme of integrating the Internet with the EDSDV networks achieves higher throughput compared to a leading approach [8]. Also the scheme has extremely low mobile IP overhead for ad hoc hosts while keeping full bi-directional connectivity. The packet delay in the proposed scheme is also better than the hybrid approach [8].

CHAPTER IV

SECURE INTERCONNECTION PROTOCOL FOR INTEGRATED INTERNET AND AD-HOC NETWORKS

The Integration of ad hoc networks with the Internet provides global Internet connectivity for ad hoc hosts through the coordination of mobile IP and ad hoc protocols. In a pure ad hoc network, it is difficult to establish trust relationship between two ad hoc hosts due to lack of infrastructure or centralized administration. In this chapter, an infrastructure-supported distributed authentication protocol is proposed to enhance trust relationships amongst ad hoc hosts. In addition an effective secure routing protocol is discussed to protect the multi-hop routing for internet and ad hoc communication. In the integrated ad hoc networks with Internet accessibility, the ad hoc routing security deployed with the help of infrastructure, has a fundamental impact on ad hoc hosts in term of internet access, integrity, and authentication. The analysis and experimental results show the achievements of the proposed security protocol.

4.1 Introduction

A pure mobile ad hoc network (MANET) is a self-configurable network with the capacity of providing communication amongst ad hoc mobile stations (MSs) without any centralized administration. In contract, a MS in an integrated internet and MANET can access the Internet by way of single hop or multi-hop connectivity through a base station (BS) or an Internet gateway [3] [4] [5] [6] [7] [8] [25]. In an integrated Internet and MANET, each MS runs a MANET routing protocol, such as DSDV, AODV, or DSR. The MANET protocol has the capacity to construct communication path between two MSs. Meanwhile, a separate protocol [3-8] [25] coordinates mobile IP [33] [34] and the MENET routing protocol to obtain the internet connectivity.

In such an integrated environment, MSs can obtain various services and applications from the Internet or the MANET. With the multi-hop internet connectivity, the integrated networks can be used to extend the coverage of wireless WANs (e.g., GSM, 3G, 4G etc.), Wireless LANs (e.g., IEEE 802.11a/b/e/g and HiperLAN/2), and wireless MANs (IEEE 802.16). A MS in an integrated network, which is outside the radio coverage of all BSs, may connect the Internet with a multi-hop path. In the integrated network, the MS can obtain the services from the participating MANET, e.g., sharing files, conference, multimedia, and games etc. Also, as an everyday experience, through the MS a user may enjoy the services from the Internet, including email, voice, messaging, information services (e.g., new stocks, weather, travel).

Similar to a MANET, an integrated network has the basic security-related characteristics such as open wireless medium, multi-hop route discovery, mobility, and constrained power capacity. However, an integrated Internet and MANET differs from a pure MANET in an important way: mobile IP with infrastructure-support. For instance, in the integrated network, Mobile IP [33] [34] could perform continuous mobility functions for mobile hosts with two entities: home agent (HA), and foreign agent (FA). A HA is the server on the mobile host's home network that maintains the information about the host's current location, as identified as care-of-address (CoA), and security credentials. On the other hand, FA is the server on the visiting network providing the CoA and security administration of the visiting network. Thus, an integrated Internet and MANET has some crucial distinctions with a pure MANET with regard to multi-hop routing security:

- In a pure ad hoc network, it is difficult to establish distributed trust relationships between two MSs as a centralized authority is absent and a MS has the capacity to forge identity. A malicious host may attack an ad hoc network many times with different identities. On the contrary, in an integrated network, the wireless and internet infrastructure (e.g., AAA server) may serves as an authentication authority and a security administrative center for MANET.
- In an integrated network, in addition to ad hoc routing security, it is necessary to enforce the security for mobile IP either for single hop or multi-hop. However, the existing mobile IP [33] [34] cannot support multi-hop communication as it is needed in an integrated MANET. Before the initiation of a communication, the Internet has to authenticate the accessing MS by a mobile IP security. Meanwhile, the MS should have the capacity to authenticate the visiting Internet for preventing a forged BS or fraudulent Internet gateway.

Due to the above differences, in the integrated networks, there are several fundamental questions that have to be addressed with regard to the Internet security and multi-hop routing security. Some of the critical questions are:

- How to build a secure FA or Internet gateway discovery protocol for a multi-hop MS?
- How to enforce a protected authentication process for the mutual validation

between a MS and it visiting network, and establish trust relationships amongst multi-hop MSs?

- How to uniquely identify a MS and prevent it from changing its identity in the purpose of attacking the network?
- How to secure an ad hoc route discovery?

However, the existing approaches [3-8] [25] for the integrated networks are based on the assumption that all MSs trust one another such that all messages can be forwarded to various destinations without any modification, drop, or injection. In a trustable environment, each participating node cooperates honestly during the process of route discovery. In practice, the Internet connectivity and routing protocols are susceptible to a wide variety of attacks in adversarial environment. The policies of security for MANETs and mobile IP have been studied separately in the existing literature and have never been considered in an integrated environment. To our knowledge no dedicated work has been carried out to address the problem of securing communication in the integrated MANET and the Internet. The following specific issues are addressed in the chapter: (i) Providing secure FA or Internet gateway discovery from MSs to the Internet, (ii) Providing secure multi-hop authentication to implement the mutual authentication between MS and the accessing Internet, (iii) Providing secure route discovery between MSs.

The rest of this chapter is organized as follows: Section 4.2 describes the background of the integrated Internet and MANET and the security challenges differing from pure MANETs. Then, the potential security threats, including the Internet connectivity and ad hoc route discovery, are discussed in Section 4.3. Section 4.4 gives the details of the proposed security protocol: global Internet security of mobile IP and ad hoc routing security. Section 4.5 analyzes the security achievements on the Internet connectivity as well as the ad hoc network. The experimental results are also shown in Section 4.5. The related works about mobile IP security and ad hoc security are addressed in Section 4.6. Finally, the chapter concludes with Section 4.7.

4.2 Background of the Integrated Network and Security Design Challenges

4.2.1 Integrated Internet and MANET

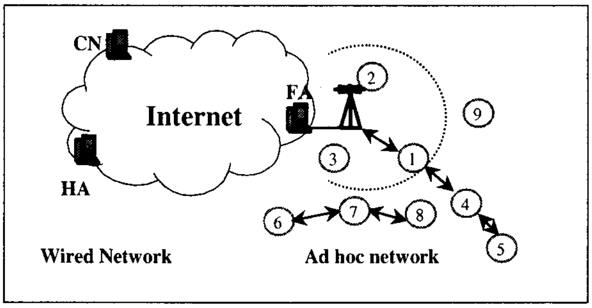


Figure 16. A MANET with global Internet connectivity.

Mobile IP and ad hoc routing protocol coordinate with each other to build the connectivity across the heterogeneous networks as shown in Figure 16. The inside area of dash circle represents the coverage of a FA; as shown in Figure 16, MSs 1, 2, and 3 are located in the coverage of the FA. The left side of Figure 16 represents a wired network that consists of a HA, a Correspondent Node (CN), a FA, and the Internet. The right side is an ad hoc network using ad hoc routing protocol, e.g., AODV, DSR. The MS 1 and the FA can reach each other directly. The basic process of an ad hoc route discovery involves two messages: routing request and routing reply. In the beginning, the source node broadcasts a routing request if it has no fresh route to the destination. The routing request is forwarded by intermediary nodes. In the end, the destination responds with a routing reply which has the route from the source to the destination. The ad hoc routing protocol enables the multi-hop MS with the capability to reach a FA. For example, in Figure 16, MS 5 may find the FA using ad hoc routing protocol via MSs 4 and 1. Moreover, the MS can obtain the Internet connectivity using mobile IP with which the MS creates a mobility binding at the HA through the visiting FA by initiating a mobile IP registration. When the FA receives a registration request from the MS, it forwarded to its HA through the Internet. After updating the mobility for the MS, the HA responds the registration with a registration reply to the MS. There are two kinds of possible communication in the integrated network. Intra-MANET communication involves interaction between hosts within the ad hoc network, e.g., in Figure 16, communication between the mobile host 6 and 8. Inter-MANET is the communication that involves infrastructure nodes (like CN) and ad hoc network nodes (like 1-9). Some properties of the integrated Internet and MANET are:

- In the integrated network, the FA can provide ad hoc nodes the capability to connect to a wired network.
- A MS that do not have direct wireless connection to the FA can communicate with the wired network by establishing a multi-hop ad hoc path.
- In the integrated network, MSs can communicate with other MSs using ad hoc routing protocol. In addition, they can also use a Mobile IP protocol to communicate with the Internet.
- Security must be implemented from the MS to the CN and vice versa. Furthermore, secure routing must also be provided for the communications between any two MSs.

4.2.2 Security Design Challenges

The routing security deals with the protection of routing messages exchanged between the nodes. Before an Intra-MANET communication, an ad hoc route discovery is needed to construct the route between two MSs. To support inter-MANET communication, a wired (Internet) route to CN is required by following the mobile IP [33] [34] that provides the continuous mobility and location management for each MS. There are two critical issues for providing a secure multi-hop route discovery: Internet connectivity and ad hoc routing protocol.

In order to obtain the internet connectivity, each MS has to register with the visiting FA and create a mobility binding at its HA [33] [34]. Before registration, a MS has to discover a FA and establish a multi-hop route to the FA. If the MS is misled by a forged FA or a malicious intermediary node, it cannot get a correct Internet connection with services. During the process of multi-hop registration, the registration messages may be modified by a malicious intermediary node. This may not allow nodes to register with its HA. The existing ad hoc routing security approaches cannot provide a secure FA discovery because all these security protocols are based on the assumption that a MANET has no any centralized infrastructure (FA). Also, the existing mobile IP protocol cannot provide the security protection for a multi-hop registration [39][40][41]. It is the fact that current mobile IP security [39][40][41] is deployed on single wireless hop in which each MS can exchange registration and authentication directly with the FA. Thus, for the Internet connectivity, two schemes are required: a secure FA discovery and a secure mobile IP registration for MS. The security of a FA discovery allows a MS to find a correct FA, and the secure mobile IP registration provides the integrity protection to registration messages across the multi-hop wireless network and the Internet.

In a pure MANET, it has no infrastructure for key management, and it is hard to verify the identity of a MS [61-71]. An integrated internet and ad hoc network is not isolated anymore like a pure MANET. The integration with infrastructure has significant impact on ad hoc routing security. The availability of the Internet enables the possible validation of MS's identify (i.e., home IP address) and credentials (i.e., billing and account). This information is generally stored at its HA and thus the routing security protocols [25-31] for pure ad hoc networks cannot be adopted for the integrated networks.

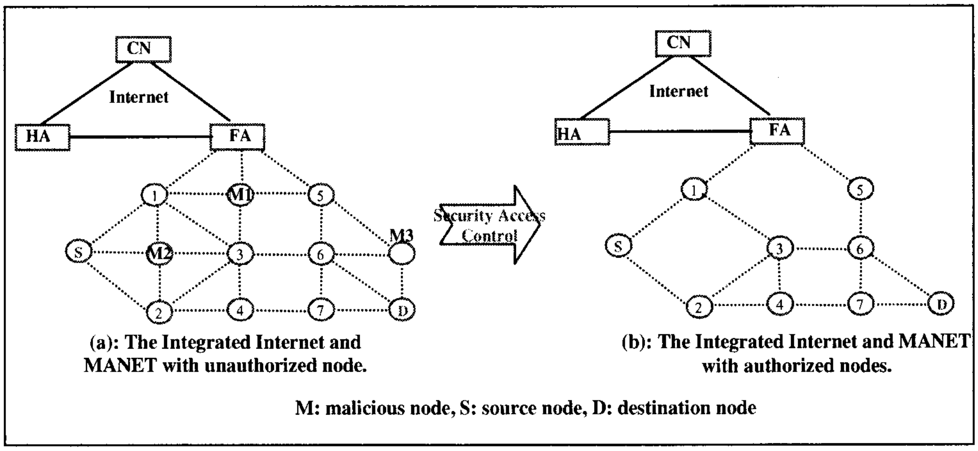
4.3 Security Requirement for Integrated Internet and MANET

Figure 17 illustrates the connectivity of an integrated Internet and MANET that can be implemented according to the routing protocols in [3-8] [25]. Figure 17 (a) shows the integrated network with malicious nodes while Figure 17 (b) illustrates the corresponding network without malicious nodes. The solid lines represent wired connection while the dash lines stand for wireless links. In Figure 17 (a), MSs 1, 5 and malicious node (M1) can reach the FA directly (MSs 1, 5, and M1 are under the coverage of the FA). Therefore MSs 1, 5, and M1 can act as the potential gateways for integrating Internet and MANET. This section summarizes the possible attacks against the Integrated Internet and MANET, and identifies the various security requirements.

4.3.1 Attacks on the Internet Connectivity

In general, malicious nodes, which modify, drop, forge or generate mobile IP messages (e.g., Advertisement, Registration Request, or Registration Reply) to corrupt mobile IP support for MANET, can cause attacks on Internet connectivity. There are mainly three types of attacks on the Internet connectivity at the network layer: bogus registration, replay attack, and forged FA. The attacks includes

- Bogus Registration
- Replay Attack
- Forged FA



(a): The Integrated Internet and MANET with unauthorized node.

(b): The Integrated Internet and MANET with authorized nodes.

M: malicious node, S: source node, D: destination node

Figure 17. An example of Integrated Internet and MANET.

4.3.2 Attacks on Ad hoc Routing Protocol

Ad hoc routing protocols are vulnerable to different types of attacks that have been extensively studied and addressed in Chapter 2. After a multi-hop route is constructed, various attacks may happen in the stage of packet forwarding as illustrated in Chapter 2.

The attacks include:

- Violating integrity (anti-integrity)
- Impersonations
- Duplication
- Anti-confidentiality
- Denial of Service (anti-cooperation)
- Modification

4.4 Proposal for Securing Global Connectivity

Table 4: Notations used for the proposed protocol

M, N	Concatenation of two messages M and N in the order specified
MS_{HM}	MS home address
HA_{id}, FA_{id}	HA and FA IP address as its identity
N_X	Nonce issued by X, e.g. HA, MS (a pseudo-random number).
$<M>K$	MAC value of message *M* under key K
CA	Certification authority
K_X, K^{-1}_X	Public and private key of X
$[M]\ K^{-1}_X$	Digital signature of message *M* generated using private key of X
$Cert_X$	Certificate of X
$T_{issue,}\ T_{expire}$	Issuing and expiration time of a MS's certificate
t	Timestamp, current estimated time
MS_{CoA}	MS's Care-of-Address
$S_{MS\text{-}HA}$	Shared secret key between MS and HA
MS_{HMx}	Permanent Home address of Mobile Station *X*
MS_X	Mobile Station *X*

Table 5: Messages for the proposed protocol

Request	*A bit pattern indicating a registration request*
Reply	A bit pattern indicating a registration reply
Result	A value indicating the result of registration
R_Request	A packet indicating a route request
R_Reply	A packet indicating a route reply
R_Error	A packet indicating a routing error
Advertisement	A bit pattern indicating an advertisement
Solicitation	A bit pattern indicating an advertisement solicitation

This chapter proposes a secure connectivity framework for integrated Internet and MANET. The securing protocol combines the mobile IP security with ad hoc routing security. Before developing the security protocol, three assumptions and related clarifications are made here. Firstly, it assumes that if two entities (e.g., MS, HA, FA) have security association (public/private key or secret key), and their identities have been authenticated by authorities, e.g., Diameter [75] and AAA authorities [74], thereby the two entities are mutually trustable. Secondly, it assumes that each MS belongs to a certain administrative domain (HA) and has a security association with its authority [39] [40] [41] [74]. Otherwise the MS is an unauthorized node. The authorized node and its HA are mutually trusted after validation. The HA is in charge of a MS for maintaining its credit, account, service policy, even marking the MS as a malicious node. Thirdly, it assume that a FA and a HA have a security association [39] [40] [41] [74] [75]. The FA and HA can be trusted by each other after mutual validations. The trust chain for integrated MANET, mobile IP, and AAA (Authentication, Authorization, and Accounting) is provided in Section 4.5.2 (Figure 7). Table 4 and Table 5 list the notations and messages used for the development of the proposed protocol.

The security protocol for integrated Internet and MANET includes two parts: the global Internet security of mobile IP and the security of integrated MANET. In order to communicate with the nodes, a MS performs the following security operations:

1. *Key establishment*: the MS generates a pair of private and public keys.
2. *FA discovery and FA advertisement*: the MS finds a route to a FA, sends its public key to the FA, and obtains FA's advertisement.
3. *MS registration with FA and HA*: The MS follows an authentication protocol to register with the FA and HA.
4. *Identity Binding at FA*: MS's home address, its ad hoc identifier and its public key are bound by the FA in a certificate for a MS.
5. Certificate issuing: the FA issues this certificate for a MS, and the certificate acts as the MS's authenticated pass in the integrated network.

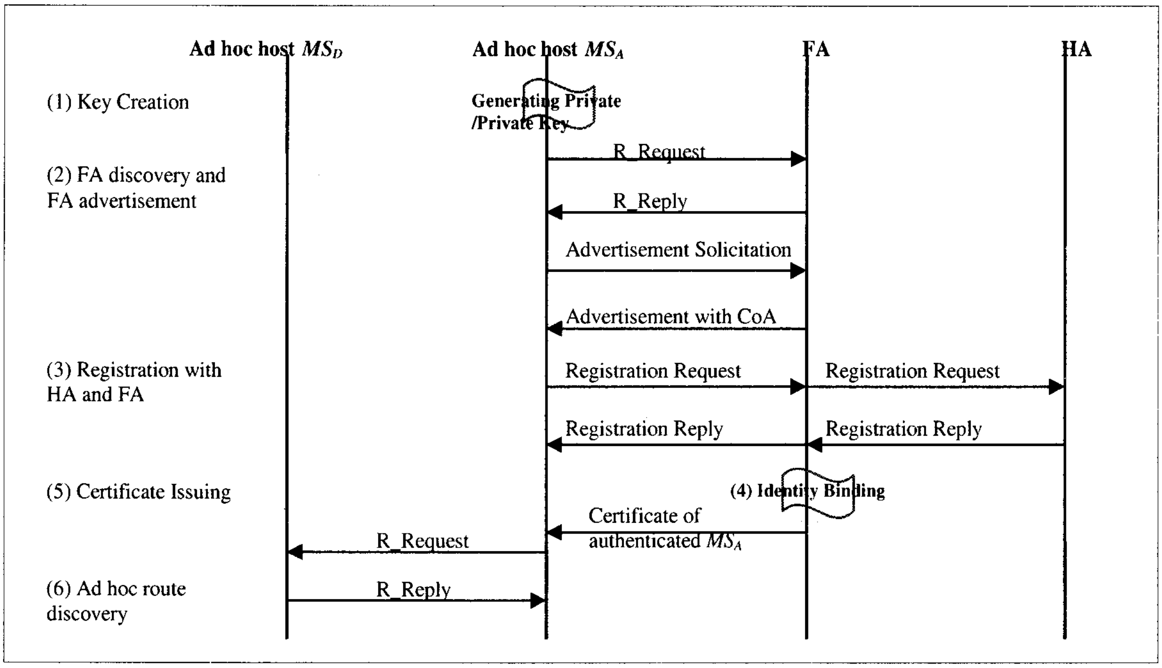


Figure 18. An Example of Routing Security.

The sequence diagram in Figure 18 illustrates the basic process of security implementation in the integrated network. The example assumes that it is the first time for a node (e.g., MS_A) to communicate with a destination (e.g., MS_D). Other MSs have already been successfully registered with the FA. In order to obtain the services from the Internet and MANET, node MS_A must register with the FA first. At first, the MS generates its private and public keys. The private key is kept secret by the MS. Then MS_A initiates a process of *FA discovery* to find a route to the FA for registration. *FA discovery* starts at node MS_A by issuing a routing request message (*R_Request*) with its signature. In the route request message, the MS provides its identity and public key to the FA. *R_Request* is forwarded to the FA hop by hop. The FA selects a route and replies MS_A with *R_Reply*. Then node MS_A sends *solicitation* to the FA to request an *advertisement* with a CoA. After receiving an advertisement from the FA, node MS_A registers with its HA via the FA by issuing a registration request. The FA and HA finish a series of authentications for checking the registration request. Registration *reply* indicates the result of the registration. If the registration is successful, the FA binds the identity of node MS_A with its public key in a certificate. Then FA issues node MS_A a certificate. Hereby, node MS_A can initiate an ad hoc route discovery for communicating with the other mobile hosts, e.g., MS_D in the integrated MANET or a CN in the Internet.

4.4.1 Global Internet Security of Mobile IP

In the proposed approach, each MS has a local table to record recently received packet information, e.g., packet source address (MS_{HM}), nonce (N_{MS}), and time (t). Each node shares a secret key with its HA for calculation and validation of MAC (Message Authentication Code). Certification Authority, HA, and FA have a pair of public and private keys separately for their mutual authentication.

4.4.1.1 Key Establishment

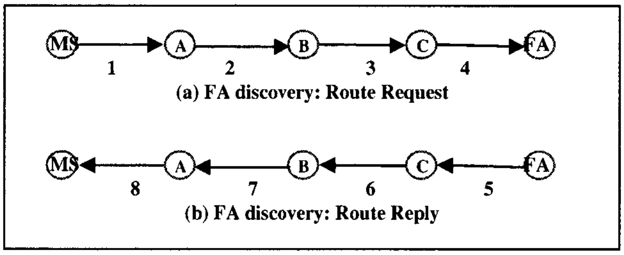
As a first step in the protocol, a MS uses a key generation function to calculate a pair of private and public keys. An example of key generation is provided at appendix. The MS keeps its private key secret. Its public key as well as its identifier (the home address of the MS) is sent to the FA via a FA discovery. The MS starts a process of a FA discovery according to the following subsection.

4.4.1.2 FA Discovery and FA Advertisement

A FA periodically advertises to ad hoc MSs with *advertisement:* M_1, $[M_1]$ K^{1}_{FA} $Cert_{FA}$; where $\boldsymbol{M_1}$ is *advertisement, Sequence, t,* FA_{id} and MS_{CoA}. The sequence number is incremented every time a new advertisement is issued by the FA. While receiving an advertisement from the FA, the MS decrypts the advertisement by using FA's public key, and compares the FA's address, timestamp (t), and sequence number with those of previously received advertisement in its local table. MS discards the duplicate advertisements. If it is a fresh advertisement, MS records the FA's certificate, IP address, timestamp, and sequence number to avoid duplication. The record is also used for tracking the history of FA's advertisement. A MS may rebroadcast this *advertisement* on its interface for the purpose of sending advertisement to the other MSs.

In the beginning, each MS does not have an authenticated public key of the FA. The process of *FA discovery* enables a MS to search an available FA and obtains a certificate of the FA. The public key inside the certificate will be authenticated in the process of followed registration. If a MS has never received a FA advertisement, but it wishes to have the knowledge of the route to a FA (i.e., Internet gateway), the MS issues a FA discovery with a destination address of *FA_Address (224.0.0.11)*; this address is the mobile agent multicasts group address. Otherwise, if the MS has registered with the FA, the MS can start its route discovery according to the steps in Section 4.4.2.1 (*ad hoc route discovery*). The process for a MS to discover a route to a FA includes two stages as shown in Figure 19: FA route request and FA route reply.

The process of the FA route request from a MS has four steps as shown in Figure 19 (a). MS initiates *R_Request* with *FA_Address* that is signed with its private key. In the *R_Request* message, the MS claims its identifier (its home IP address) and its public key (MS_{HM}, K_{MS}). When any neighbor *A* of MS receives *R_Request*, node *A* checks against the local table (MS_{HM}, N_{MS}, and t) to verify whether it has already seen the request and whether the packet has a valid timestamp. If *A* has seen this packet before or an invalid timestamp is found on the packet, *A* discards the duplicated or invalid request. Knowing that the destination is FA, the neighbor *A* cannot verify MS's identifier because the ad hoc MS still has not its certificate issued by the FA. Therefore neighbor *A* leaves the job of verification of MS's signature to FA. The neighbor *A* rebroadcasts the FA discovery request after appending its address and signing the packet. All intermediate nodes, e.g., *A*, *B*, and *C*, must be registered nodes (Figure 19 (a)). Aside from above steps, each intermediate node, except for MS's neighbors (e.g., node *B* and *C*, but no *A*), must validate the signature of its preceding node with the public key of preceding node, which is issued by FA and enveloped inside the certificate. Each intermediate node other than MS's neighbors (e.g., node *B* and *C*) but not *A*, removes the signature of preceding node before its rebroadcast. In the end, FA receives the route discovery packet. The detailed message exchanges based on Figure 19 (a) are given below.



(a) FA discovery: Route Request

(b) FA discovery: Route Reply

Figure 19. Steps of FA Discovery.

1. *MS* broadcasts route request:
   $R\_Request$: $[R\_Request, MS_{HM}, K_{MS}, FA\_Address, N_{MS}, t]\ K^{-1}_{MS}$
2. *A* receives *R_Request* from *MS:*
   $R\_Request$: $[[R\_Request, MS_{HM}, K_{MS}, FA\_Address, N_{MS}, t]\ K^{-1}_{MS}, MS_{HMa}]\ K^{-1}_{HMa}, Cert_{HMa}$
3. *B* receives *R_Request* from *A:*
   $R\_Request$: $[[R\_Request, MS_{HM}, K_{MS}, FA\_Address, N_{MS}, t]\ K^{-1}_{MS}, MS_{HMa}, MS_{HMb}]\ K^{-1}_{HMb}, Cert_{HMb}$
4. *C* receives *R_Request* from *B:*
   $R\_Request$: $[[R\_Request, MS_{HM}, K_{MS}, FA\_Address, N_{MS}, t]\ K^{-1}_{MS}, MS_{HMa}, MS_{HMb}, MS_{HMc}]\ K^{-1}_{HMc}, Cert_{HMc}$

When a FA receives the route discovery packet from node *C* as shown in Figure 19 (b), the FA validates *C*'s signature. Also the FA verified MS's signature using their public keys claimed in the *R_Request* message. The FA records the claimed identifier and the public key of the MS. The identifier and public key will be future verified through the followed process of the MS's registration (Section 4.4.1.3). If the FA receives several valid discovery packets from the same MS, it chooses the best route (e.g., shortest path). Then, the FA initiates an *R_Reply* packet by using FA's address with a new nonce. FA returns the *R_Reply* to MS with the reverse path from MS to FA. Before *R_Reply* arrives to MS's neighbor, each intermediate node validates the signature of the preceding node and prevents duplication by comparing nonce and timestamp with its local table. Then each intermediate node rebroadcasts the *R_Reply* after removing the signature of preceding node and signing with its own private key. When *R_Reply* reaches the MS's neighbor A, A validates the signature of preceding node, and then rebroadcasts the packet to MS without signing on it. The detailed message exchange based on Figure 19 (b) is shown below.

5. *FA* receives *R_Request* from *C*:
   R_Reply: [*R_Reply*, MS_{HM}, MS_{HMa}, MS_{HMb}, MS_{HMc}, FA_{id}, N_{FA}, t] $K^{-1}{}_{FA}$, $Cert_{FA}$
6. *C* receives the R_Reply form *FA*:
   R_Reply: [[*R_Reply*, MS_{HM}, MS_{HMa}, MS_{HMb}, MS_{HMc} FA_{id}, N_{FA}, t] $K^{-1}{}_{FA}$, $Cert_{FA}$] $K^{-1}{}_{HMc}$, $Cert_{HMc}$
7. *B* receives the R_Reply form *C*:
   R_Reply: [[*R_Reply*, MS_{HM}, MS_{HMa}, MS_{HMb}, MS_{HMc} FA_{id}, N_{FA}, t] $K^{-1}{}_{FA}$, $Cert_{FA}$] $K^{-1}{}_{HMb}$, $Cert_{HMb}$
8. *A* receives the R_Reply form *B*:
   R_Reply: [*R_Reply*, MS_{HM}, MS_{HMa}, MS_{HMb}, MS_{HMc} FA_{id}, N_{FA}, t] $K^{-1}{}_{FA}$, $Cert_{FA}$

When the MS receives *R_Reply* from the neighbor A, it validates the signature of the FA. Node A doesn't sign *R_Reply*. It doesn't matter because if any malicious node provides falsified route information, the MS can detect the falsification by checking the signature of FA. The MS extracts the route between itself and the FA from the *R_Reply*. In the above example, < MS_{HM}, MS_{HMa}, MS_{HMb}, MS_{HMc} FA_{id} > is the route from the MS to the FA. Now MS can use the route to send a *solicitation* message to FA. Upon receiving *solicitation* message, the FA returns an *advertisement* to the MS with a CoA. Both the *solicitation* and *advertisement* are protected with signatures. Then, the MS selects the advertised CoA to register with FA by issuing a registration request message. The processing of registration has the following steps.

4.4.1.3 MS Registration with FA and HA

For the purpose of the security of ad hoc network, each node must register with FA and obtain a certificate from FA before participating in ad hoc routing protocol. There are two functionalities during registration for a MS: (i) mutually authentication between the MS and the visiting network (i.e., FA), (ii) mobility binding for the MS according mobile IP protocol. The scheme assumes that a public key infrastructure (PKI) is available by HA and FA [74] [75]. Each MS and its HA share a security association to create MAC (Message Authentication Code) for registration request and reply [41]. A MS performs cryptographic operation using its secret-key ($S_{MS\text{-}HA}$) to protect the integrity of registration request. The MS starts its registration by issuing a registration request carrying the original advertisement as well as its HA address (HA_{id}) etc. The registration request message is signed by the MS with the secret key ($S_{MS\text{-}HA}$). Then the registration request message is forwarded to the FA hop by hop according to the established route constructed via a FA discovery. Then the registration request is sent to the HA through the FA after appending FA's nonce. The HA verifies the registration request by checking the MAC with the associated secret key. The HA further checks advertisement and the FA by validating the certificate and the signature of FA. After the validation the registration request and the FA, it returns the registration result to the FA. Then, the FA validates the HA by checking the certificate of the HA and its signature. After the validation, it returns the registration result to the MS. Once MS receives a successful reply from HA, it is guaranteed that FA's certificate is valid. The registration protocol operates as follows:

(R1) MS -> FA: M_2, $< M_2 > S_{MS\text{-}HA}$. Where M_2 = Registration Request, FA_{id}, HA_{id}, MS_{HM}, MS_{CoA}, N_{MS}, N_{HA}, {Message in *advertisement*}

(R2) FA->HA: {Message in R1}, N_{FA}

(R3) HA: (upon receipt of R2):
Validate < M_2> $S_{MS\text{-}HA}$ using $S_{MS\text{-}HA}$
Check whether FA_{id} in *advertisement* = FA_{id} in M_2
Validate $Cert_{FA}$ based on existing PKI at HA
Validate $[M_1]\ K^{-1}_{FA}$ using authenticated K_{FA}
Continue with the steps in [33][34] (Perkins, Mobile IP support, e.g., updating mobility binding for the MS)

(R4): HA->FA: M_3, $[M_3]\ K^{-1}_{HA}$, $Cert_{HA}$,
Where M_3= M_4, N_{FA};
M_4 = Reply, Result, FA_{id}, HA_{id}, MS_{HM}, $MS_{CoA,}$ N'_{HA}, N_{MS}, < M_4> $S_{MS\text{-}HA}$

(R5): FA: (upon receipt of R4)
Validate N_{FA}
Validate $Cert_{HA}$ based on existing PKI at HA
Validate $[M_3]\ K^{-1}_{HA}$ using authenticated K_{HA}
Log this message as a proof of serving MS
Identity binding and certificate creation
Continue with the steps in [8] (Perkins, Mobile IP support, e.g., creating mobility binding for the MS)

(R6): HA->MS: M_4

(R7) MS:
Validate < M_4> $S_{MS\text{-}HA}$ using $S_{MS\text{-}HA}$
Continue with the steps in [33][34] (Perkins, Mobile IP support, e.g., obtaining registration result)

The process of the secure registrations for all MS creates a trusted integrated MANET. As shown in Figure 17 (a) the integrated Internet and MANET may have malicious nodes. The malicious nodes can attack the integrated Internet and MANET by participating in the activities of routing protocols. The security registration excludes these malicious nodes by disallowing them from participating in the routing protocols shown in Figure 18 (b). The malicious nodes cannot obtain services from the Internet and MANET. The detailed information as regards how to achieve it by using certificates and signatures is discussed next.

4.4.1.4 Identity Binding

Depending on ad hoc network protocols, ad hoc networks may use different addressing solutions; AODV, DSR, and TORA use Node ID; HSR has a hierarchical addressing solution; ZLHS use <zone id + node id> as MS ad hoc address. In a standalone ad hoc network, any MS can easily masquerade itself through changing its ad hoc identity so that it is difficult for ad hoc network to trace its previous actions. However, in an integrated network, the proposed security protocol requires each MS uses its home address as its ad hoc identifier. Moreover, the home address is bound with its public key. For each registered MS, the FA issues a certificate to associate the binding. This binding makes it possible to trace the MS's history so that the bad credit MS can be excluded from the ad hoc network routing. A MS claims its home address and its public key in the process of its FA discovery, and the FA verifies the information through HA during the MS's registration. Therefore if a match is found between a claimed home address and the home address in registration request, the FA associates MS and MS's public key to be a certificate as a passport to the MS.

4.4.1.5 Certificate Issuing

Each successfully registered node maintains a copy of certificates of itself issued by FA as (R8). The certificate ($Cert_{MS} = <<MS_{HM}, FA_{id}, K_{MS1}, T_{issue}, T_{expire}>> K^{1}{}_{FA}$) of an ad hoc MS includes its permanent IP address (MS_{HM}), public key (K_{MS}), issuing time (T_{issue}), and expiring time (T_{expire}). Because of the uniqueness of the permanent IP address of the MS, the certificate is unique for each MS. Each certificate has the signature of the FA, providing the integrity protection from being forged or modified. Since each registered MS has an authenticated public key of the FA, they can validate the legality of a certificate using the public key of FA (K_{FA}). The certificate can be used as the passport for multi-hop routing security. For instance, in the Figure 19, if the registered MS *B* receives a routing request packet from the neighboring MS *A* with the certificate of MS *A*, MS *B* first validates the legality of the certificate of MS *A* ($Cert_{MSa} = <<MS_{HMa}, FA_{id}, K_{MSa}, T_{issue}, T_{expire}>> K^{1}{}_{FA}$) by verifying the signature of the FA and checking the issuing and expiration times of the certificate. If the certificate is validated, the MS *B* reads the public key of MS *A* (K_{MSa}) from the certificate, and uses it to verify the authenticity of route request packet. In the next cycle, if MS *B* receives another packet from MS *A* with the same certificate, the MS *B* can verify the packet directly using the public key of MS *A* without validating the certificate again. Before the expiration of the certificate, if the MS is still on the ad hoc network, it requests the FA with a new certificate and the FA reissues certificate to the MS with an extended period as (R8).

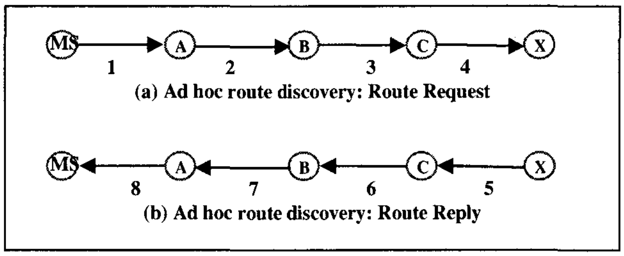
(R8) FA-> MS, *Certificate of MS*: $Cert_{MS} = <MS_{HM}, FA_{id}, K_{MS1}, T_{issue}, T_{expire}> K^{-1}_{FA}$

4.4.2 Security of Integrated MANET

Only authenticated MS has a certificate from the registered FA. It guarantees only the authenticated nodes have the capability to participate in ad hoc route discovery. Unauthorized nodes cannot obtain services from either the Internet or MANET because the packets issued by unauthorized nodes are ignored by other nodes. The routing security for multi-hop communication includes ad hoc route discovery, routing cache, and routing maintenance.

4.4.2.1 Ad hoc Route Discovery

In order to discover a route to a destination node in the ad hoc network, a MS follows the steps outlined in Figure 20. If it is the first time for a MS to discover the destination node, MS creates a pseudo-random number as nonce. Each intermediate node on route to the destination node validates whether the neighbor, from which the packet is received, is an authorized one by validating its certificate and the signature. The validation uses the association between the IP address of the neighbor and its certificate. If in the local table the T_{issue} and T_{expire} of the neighbor are invalid or expired, the received packet must be discarded. Each intermediate node records its reverse route, from which the packet is received, and then signs with its private key before sending a packet out. Ad hoc route discovery has two stages: route request and route reply. Figure 20 illustrates the route discovery with AODV ad hoc routing protocol. In Figure 20, *A*, *B*, and *C* are intermediate nodes, and *X* is the destination node. Firstly, MS broadcasts RREQ (Route Request) marked as *R_Request* with its signature. The *R_Request* includes home addresses of the source and destination node, a nonce, and issue time. The signature of the MS to non-mutable fields (MS_{HM}, MS_{HMx}, N_{MS}, and t) protects the integrity of non-mutable items of *R_Request.* The signature of intermediate node is to protect mutable field, such as *hop_cnt*, from being modified by other nodes. After receiving the route discovery at the destination node, the destination node checks the request and replies with *R_Repy.* For convenience, in the following example, some route fields, such as route lifetime and broadcast id, are not shown but those fields should be included in practice.



(a) Ad hoc route discovery: Route Request

(b) Ad hoc route discovery: Route Reply

Figure 20. Steps of Ad-hoc Discovery.

1. *MS* broadcasts:
   R_Request: $[R\_Request, MS_{HM}, MS_{HMx}, N_{MS}, \mathrm{t}]\ K^{-1}{}_{MS}, Cert_{MS}$
2. *A* receives *R_Request* from *MS*
   R_Request: $[[R\_Request, MS_{HM}, MS_{HMx}, N_{MS}, \mathrm{t}]\ K^{-1}{}_{MS}, hop\_cnt]\ K^{-1}{}_{MSa}, Cert_{MS}, Cert_{MSa}$
3. *B* receives *R_Request* from *A*
   R_Request: $[[R\_Request, MS_{HM}, MS_{HMx}, N_{MS}, \mathrm{t}]\ K^{-1}{}_{MS}, hop\_cnt]\ K^{-1}{}_{MSb}, Cert_{MS}, Cert_{MSb}$
4. *C* receives *R_Request* from *B*
   R_Request $[[R\_Request, MS_{HM}, MS_{HMx}, N_{MS}, \mathrm{t}]\ K^{-1}{}_{MS}, hop\_cnt]\ K^{-1}{}_{MSc}, Cert_{MS}, Cert_{MSc}$
5. Destination node *X* receives *R_Request* from *C* :
   R_Reply: $[R\_Reply, MS_{HM}, MS_{HMx}, N_x, \mathrm{t}, hop\_cnt]\ K^{-1}{}_{MSx}, Cert_{MSx}$
6. *C* receives *R_Reply* from destination node *X*
   R_Reply: $[[R\_Reply, MS_{HM}, MS_{HMx}, N_x, \mathrm{t}, hop\_cnt]\ K^{-1}{}_{MSx}]\ K^{-1}{}_{MSc}, Cert_{MSx}, Cert_{MSc}$
7. *B* receives *R_Reply* from destination node *C*
   R_Reply: $[[R\_Reply, MS_{HM}, MS_{HMx}, N_x, \mathrm{t}, hop\_cnt]\ K^{-1}{}_{MSx}]\ K^{-1}{}_{MSb}, Cert_{MSx}, Cert_{MSb}$
8. *A* receives *R_Reply* from *B*
   R_Reply: $[[R\_Reply, MS_{HM}, MS_{HMx}, N_x, \mathrm{t}, hop\_cnt]\ K^{-1}{}_{MSx}]\ K^{-1}{}_{MSa}, Cert_{MSx}, Cert_{MSa}$

4.4.2.2 Routing Cache

Routing cache is used to provide a faster routing response via intermediate node before the routing request arrives at the destination node. A MS starts an ad hoc route discovery by issuing a routing request (*R_Request*) to search the destination node. If an intermediate node, which receives the *R_Request*, has a valid route to destination node in its route table, the intermediate node issues a route reply (*R_Reply*) and forwards it to the source node. In the *R_Reply* issued by the intermediate node, the original route reply, which is issued and signed by destination node, should be included. Therefore the source node can construct a route from the source node to the destination node after checking the signature of the intermediate node as well as the signature of the destination node. In Figure 20 (a), for example, when node *B* receives a route request from MS, if node *B* has a valid route leading to the destination node, node *B* creates a route reply message (*R_Reply)* and sends the message to the source node. After checking the signature of node *B* and *X*, source node *MS* constructs the route from MS to *X* via *B*.

9. *B receives R_Request from A*
R_Reply: [R_Reply, MS_{HM}, MS_{HMx}, N_{MSb}, *t, [Rx_Reply]* K^{-1}_{MSx}, *hop_cnt]* K^{-1}_{MSb}, $Cert_{MSb}$; *where Rx_Reply is the routing reply issued earlier by destination node x.*

10. *A receives R_Request from B*
R_Reply: [R_Reply, MS_{HM}, MS_{HMx}, N_{MSb}, *t, [Rx_Reply]* K^{-1}_{MSx}, *hop_cnt]* K^{-1}_{MSb}, $Cert_{MSb}$

4.4.2.3 Routing Maintenance

A route error message (*R_Error*) is generated by a node to report the failure of a link due to the movement of nodes. Based on *R_Error* messages, the source node reconstructs a new route to destination node. All *R_Error* messages must be signed by the issuing node. For example, when the link from node *A* to node *B* is broken in Figure 20 (a), node *A* creates *R_Error* to notify node *MS* the link failure. Upon receiving the *R_Error* and knowing the link failure at node *A*, node NN issues a new route discovery to search other route to the destination. The route error message (*R_Error*) issued by node *A* is of the form: *[R_Error, MS_{HM}, MS_{HMx}, MS_{HMb}, N_b, t]* K^{-1}_{MSb}.

4.5 Security Analysis and Network Performance

This section evaluates the security protocol for the integrated Internet and MANET. The evaluation has two parts: security and performance analysis. The security analysis illustrates the security achievements on the Internet connectivity as well as the integrated MANET. The performance analysis includes the effectiveness of key and certificate management as well as the communication cost for maintaining the security protocol. Appendix also includes a discussion on key creation, signature and verification algorithm.

4.5.1 Security Analysis on Internet Connectivity

The long-term shared secret key S_{MS-HA} is used for authentication at the MS's home network. We assume that S_{MS-HA} is kept secret meaning that a malicious MS cannot obtain S_{MS-HA}. Moreover, the long term private keys held by FA and HA cannot be obtained by an attacker. We also assume the security operations at FA and HA are strong enough from being compromised. For example, HA correctly authenticates the credentials of the registering MS. The certificates issued by FA are resilient to brute-force and cryptanalysis attacks. The private keys of FA and MSs are properly kept secret by themselves from stealing and guessing. It means that the signatures of FA and MS cannot be broken by an attacker. According to above assumptions, the security protocol achieves the goals of preventing the attacks of bogus registration, reply attacks, unauthorized routing, and forged FA. In order to obtain services from the Internet or MANET a MS first issues a FA discovery to establish a path between the MS and the FA by using authenticated nodes. This avoids unregistered malicious nodes to mislead route or drop registration messages with the intention of hindering MS registration. Then, by trusting HA for authentication, the security protocol ensures that the MS's registration request is legitimately created. Moreover, the security protocol also ensures that the MS's registration request has not been changed during the forwarding from MANET to FA/HA. In the end, the certificate issued for the successful registered MS, protects the ad hoc route discovery.

In this security protocol, all the route creations and communications take place among the trusted nodes verified by FA (Sections 4.4.1 and 4.4.2). The registered MS ignores all the control messages from an unauthorized MS. Thus, an unauthorized node cannot participate in the routing activity of this protocol due to lacking a validated certificate. The only exception is the route request in the process of FA discovery. The genuineness of this request is authenticated by verifying the MS's identity during the registration.

Bogus Registration: When a malicious node in the integrated MANET makes a fake registration by masquerading itself as someone else, the malicious node issues a forged registration message with an invented or spoofed address. The forged registration will be stopped at the step (*R3*) of Section 4.4.1.3 (Validate $< M_2 > S_{MS\text{-}HA}$ using $S_{MS\text{-}HA}$) because the malicious node does not have the knowledge of secret key ($S_{MS\text{-}HA}$) associated between the invented or spoofed address and the related HA.

Replay Attacks: The nonce (e.g., $N_{MS}, N_{HA},$ *and* N_{FA}) and timestamp (i.e., *t*) is used in all mobile IP and routing messages to ensure that a registration or routing message contain a unique data to prevent replay attacks. Each registration or routing request has a nonce, and a new nonce in the registration or routing reply message indicates the next nonce for the next request.

Forged FA: When a MS advertises itself as a fraudulent FA, there are two possibilities for MSs that are under the coverage of the forged FA

i. the MSs that have not yet registered with HA via a correct FA,

ii. the MSs that have successfully registered with FA and HA.

Assume MS *A* is in the first case and it starts FA discovery by providing its public key in the route request message. The forged FA then replies the MS *A* directly because the MS *A* is under the coverage of the forged FA. Then MS *A* tries to register with the forged FA. A registration message with a MAC association created by using MS *A*'s secret key ($S_{MS\text{-}HA}$) is sent to the forged FA. The forged FA cannot reply MS *A* a registration reply message with a correct MAC because the forged FA has no knowledge of the secret key of MS *A* ($S_{MS\text{-}HA}$). If the forged FA sends the registration message to MS *A*'s HA, the registration is declined at the step of (*R3*) in IV.A.2 (Validate $Cert_{FA}$ based on existing PKI at HA, Validate $[M_1]\ K^{1}_{FA}$ using authenticated K_{FA}). If the forged FA uses an earlier registration reply message in attempt to cheat the MS *A*, the MS *A* can know the trick due to the protection of nonce. Therefore the forged FA cannot cheat MS *A* and its HA. Then MS *A* tries another FA until it registers with a correct FA.

Let's consider the second case. When MS *B* receives an advertisement from the forged FA, if MS *B* doesn't want to leave the integrated MANET, MS *B* will not go to register with the forged FA since the MS *B* has successfully registered with a correct FA. The worse situation is that MS *B* tries to make a handoff to the forged FA by registering with the forged FA. However, MS *B* cannot successfully register with the forged FA for the same reasons as explained in the first case. Therefore, MS *B* will keep its registration with the correct FA.

In the next cycle, if MSs *A* and *B* receive the fraudulent advertisements again, MSs *A* and *B* just ignore it because there was no successful registration with the forged FA earlier. The MSs, which are located in the direct transmission range of the forged FA, will not forward the fraudulent advertisement to other MSs. So the forged FA has no negative effect on the MSs that are outside the radio coverage of the forged FA.

4.5.2 Security Analysis on Integrated MANET

The important principle in the security protocol is that unauthorized MSs are excluded from the integrated MANET. Unauthorized nodes cannot obtain services from the Internet and MANET since the packets from unauthorized nodes will be discarded by registered MSs. Before participating in ad hoc routing protocol, a MS must register with FA to obtain a certificate from FA. During registration, the HA authenticates the registering MS and its visiting FA. Once the MS receives a successful registration reply from the HA, it is assured that the FA is valid. Meanwhile if MS successfully registers with FA, it is an authorized ad hoc MS. The proposed registration plays two main roles: (i) unique mobility and security certificate binding at the HA/FA, and (ii) establishment of trust relationships amongst ad hoc MSs.

Unique mobility and security certificate binding at HA/FA: Each MS has a unique home address. Each MS's home address is bound with a public key in a certificate by FA. No node other than FA can create a correct certificate on behalf of FA because no other node has the private key of FA. Hence the identity binding in each certificate is unique for each MS.

Establishment of trust relationships amongst MSs: Figure 21 illustrates the trusted model for integrated MANET, Mobile IP, and AAA (authentication, authorization, and accounting). If there is a security association (SA) between the two entities (e.g., HA, FA, MS), and the SA has been validated, then the two entities are trusted with each other. For example, in Figure 21, SA2 can be established after the mutual authentication between node *A*'s HA and its AAAH server. In Figure 21, node *A* and HA AAA server of node *A* (A's AAAH server) have the SA1. If the SA1 is verified, then the trust relationship between node *A* and A's AAAH server is established. The proposed protocol verifies the SA1 with the process of registration. Node *A* starts a security authentication by issuing a registration request with a MAC calculated by using the secret key ($S_{MS\text{-}HA}$). The registration request messages are protected from being modification during the ad hoc and internet delivery due to the MAC protection. Therefore, the HA can validate SA1 after receiving the registration request. Also, the *A*'s AAA H checks whether node *A* has the acceptable credentials or not. In the wired network, a series of security associations (SA2, SA3, and SA4) are checked based on the peer trust relations [74] [75]. The FA authentication server (AAAF) obtains the authenticating result from the external home authentication server (AAA H of node A). Finally, if the node A's registration request is approved, then the node *A* and FA are mutually trusted. It is because the security associations constructs a trust chain (SA5) between node *A* and FA (in Figure 21, if SA1, SA2, SA3, SA4, then SA5).

In the same way, when node *B* is successfully registered, the trust relationship between node *B* and FA (SA10) is established (in Figure 21, if SA6, SA7, SA8, SA9, then SA10 for node B). Further, node *A* and *B* both are authorized nodes, and can be trusted by each other (in Figure 21, if SA5, SA10, then SA11). Similarly, if other nodes (e.g., *C*, *D*, and *E*) are authorized nodes, these nodes are trusted nodes too. The trust relations are associated by way of certificates issued by FA.

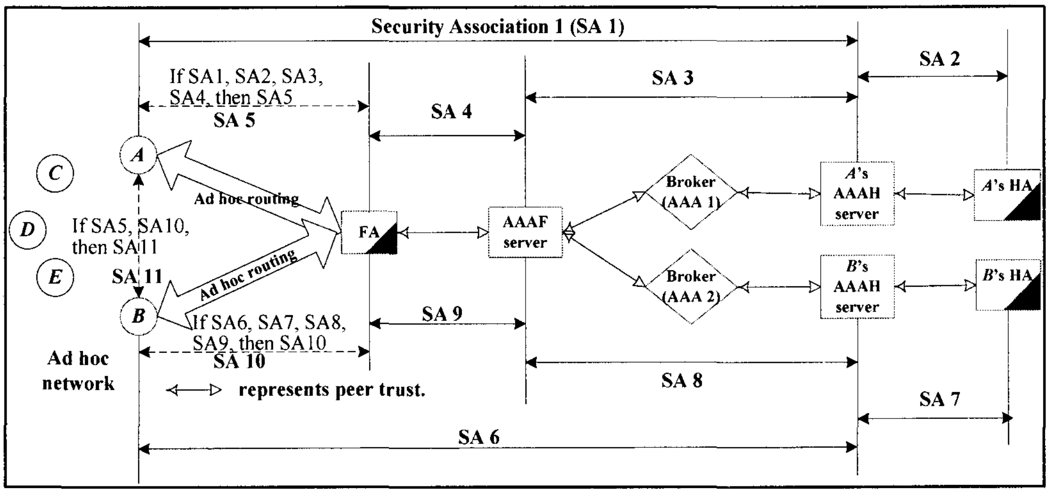


Figure 21. The Trust Model for Integrated MANET, Mobile IP, and AAA.

The proposed ad hoc route discovery prevents attacks in terms of integrity, impersonation, confidentiality and cooperation.

Integrity: Each ad hoc routing message is signed by using the private key of each sender. The receiver verifies the certificate and signature of the sender. Each authorized node keeps its private key secretly. Therefore, the signature and verification prevent anti-integrity attacks in the ad hoc routing protocol. The attacks of modification and routing loop can be prevented by the integrity protection of routing messages.

Impersonation: In the integrated network, the proposed approach binds the MSs' home addresses with public key in the ad hoc network. The binding is unique because of the uniqueness of MS's home address. The secret key encryption prevents impersonation on registration by way of MAC. The private key and the certificate prevent impersonation on ad hoc routing by signing and verifying. Therefore, it becomes difficult for any MS to masquerade itself by spoofing or inventing an address either in registration or in ad hoc routing. Fabrication can be avoided by protecting the identity of each MS.

DOS: The access control is achieved by issuing a certificate to each MS by FA in step IV.A.5 of the protocol. Based on the assumption stated at the beginning of Section 4.4, trusted nodes with valid certificates will participate properly in ad hoc routing and communication protocol. On the hand, we can consider all MSs are selfish and are not willing to relay traffic for other MSs in order to save their own resources. Thus, schemes like [66] [67] [68] [69] [70] are needed to stimulate the active participation of packet forwarding and deter malicious actions through manipulating the billing or credit account. Since each MS has a certificate, all nodes on a multi-hop path can be authenticated by the source or destination. Based on the reports from multi-hop network, the BS can verify the communication path and keep track of each register MS. According the traffic on the path, the BS can decide which accounts should be charged or credited [66] [67]. The low credit node is given less priority in network services such as low internet bandwidth. Even if a registered MS acts as a malicious node, malicious behaviors can be detected with the complaints of other MSs.

4.5.3 Performance Analysis

Analysis and simulations are conducted to evaluate the effectiveness of the proposed security protocol in terms of computation overhead, communication cost, and security achievements. The computation overhead depends on many factors, for instance, the algorithms for key creation, signing and verifying as well as security level. A MS can pre-create a pair of private and public keys. Currently, there are a number of asymmetric key cryptosystem in literature, e.g., RSA, DSA, ElGamal, and Elliptic Curve DSA. The RSA (Rivest, Shamir, and Adelman) is based on the difficulty of factoring large integers. The ElGamal is the cryptosystem based on the difficulty of solving the discrete logarithm in the multiplicative group of a field. The ECC (Elliptic curve cryptosystem), which is based on the difficulty of solving a discrete logarithm problem in the group of points on an elliptic curve, is a competing system because it offers equal security with a smaller key size. An implementation of the key creation, signature and verification based on the DL/EC signature techniques is implemented for evaluating the proposed security approach for integrated network (see appendix).

The communication cost incurred by the proposed security protocol includes two parts (i): communication cost of FA discovery, registration, and distribution of certificates, (ii) the communication cost for maintaining the certificates. The overhead of ad hoc routing is not considered because it is not caused by the proposed security protocol. We assume that the ad hoc network uses AODV protocol, and there are **n** MSs in the integrated MANET, in which **n'** is the node outside the coverage of the FA. Therefore, it has **(n-n')** nodes located inside the coverage of the FA). Let A_{nh} is the average number of hops from all MSs to FA. Let A'_{nh} is the average number of hops from the MSs, which is outside the coverage of FA, to FA ($A'_{nh} \geq A_{nh}$). Let A_{ps} be the average package size of all kinds of control packets, e.g., R_Request, R_Reply, Registration request and reply, and Certificate.

The communication cost for initiating an integrated MANET: It has two stages for registration and authentication during initiating an integrated MANET. In the first stage, (**n- n′**) nodes hear the advertisement from FA, and register with the FA. For each registration, it has 5 messages (advertisement solicitation, advertisement with CoA, registration request, registration reply, and certificate issuing of the MS). The number of wireless transmissions needed for (**n- n′**) registrations is 5*(**n- n′**). In the second stage, the **n′** nodes outside the coverage join the MANET one by one. Considering k (k ≤ **n′**) nodes have already joined the integrated MANET, the (k+1) th MS joins the integrated MANET according to the steps illustrated in Figure 18:

- The number of transmissions for a route request during FA discovery is *((n-n') + k + 1)* for the route request to reach a FA. In this case, the route request floods through the whole integrated MANET.
- The route reply message returns to the (k+1)th node via A'_{nh} transmissions because the route from the MS to FA has been constructed.
- Each message including advertisement solicitation, advertisement with CoA, registration request, registration reply, and certificate issuing of the MS, needs A_{nh} transmissions.

Therefore, the number of transmissions (Initiating_MANET [(k+1) th]) caused by the (k+1) th MS to join the integrated MANET is:

$$\text{Initiating\_MANET } [(k+1)^{th}] = ((n-n') + k + 1) + 6\, A'_{nh} = 6\, A'_{nh} + (n - n') + k + 1$$

The number of transmissions (Initiating_MANET (n')) for initiating an integrated MANET (**n** nodes) is:

$$\text{Initiating\_MANET (n')} = \sum_{k=0}^{n'-1}[6A'_{nk} + (n - n') + k + 1] = \text{n'}(6\ A'_{\text{nh}} + \text{n} - 0.5\ \text{n}' + 0.5)$$

The total number of transmissions (Initiating_MANET (n)) for initiating an integrated MANET (**n** nodes) is:

Initiating_MANET (n) = 5 * (n-n') + Initiating_MANET (n') = 5n + n' (6 A'_{nh} + n – 0.5 n' - 4.5)

The total communication cost (Initiating_MANET$_{\text{cost}}$ (n)) for initiating an integrated MANET with *n* nodes is:

Initiating_MANET$_{\text{cost}}$ (n) = Initiating_MANET (n) * A_{ps} = (5n + n' (6 A'_{nh} + n – 0.5 n' - 4.5))A_{ps}

The communication cost for maintaining certificates: For simplicity, we assume each MS has the same certificate updating interval. Let **t** is the certificate updating interval for each node. Before a certificate expires, its MS issues an ad hoc route discovery for constructing a route from the MS to FA if the MS has no fresh route to FA. The routing packets are not counted because it depends on routing protocols and mobility of mobile hosts. After the route from the MS to FA is constructed, the MS issues a certificate request message to FA. FA broadcasts a new certificate with a new issuing time and an extended expiration time to other nodes in the integrated MANET.

- A certificate request message is sent to FA via the constructed route. It causes A'_{nh} transmissions.
- A new certificate is sent back the MS. It also causes A'_{nh} transmissions.

Therefore, the number of transmissions for reissuing a certificate in an integrated MANET with n nodes (Certificate_MANET (1)) is:

$$\text{Certificate\_MANET}(1) = A_{nh} + A_{nh} = 2A_{nh}$$

The communication cost for reissuing a certificate in an integrated MANET with n nodes (Certificate_MANET$_{cost}$ (1)) is:

$$\text{Certificate\_MANET}_{cost}(1) = \text{Certificate\_MANET}(1) * A_{ps} = 2A_{nh}A_{ps}$$

The total number of transmissions for reissuing **n** certificates in an integrated MANET with n nodes (Certificate_MANET (n)) is:

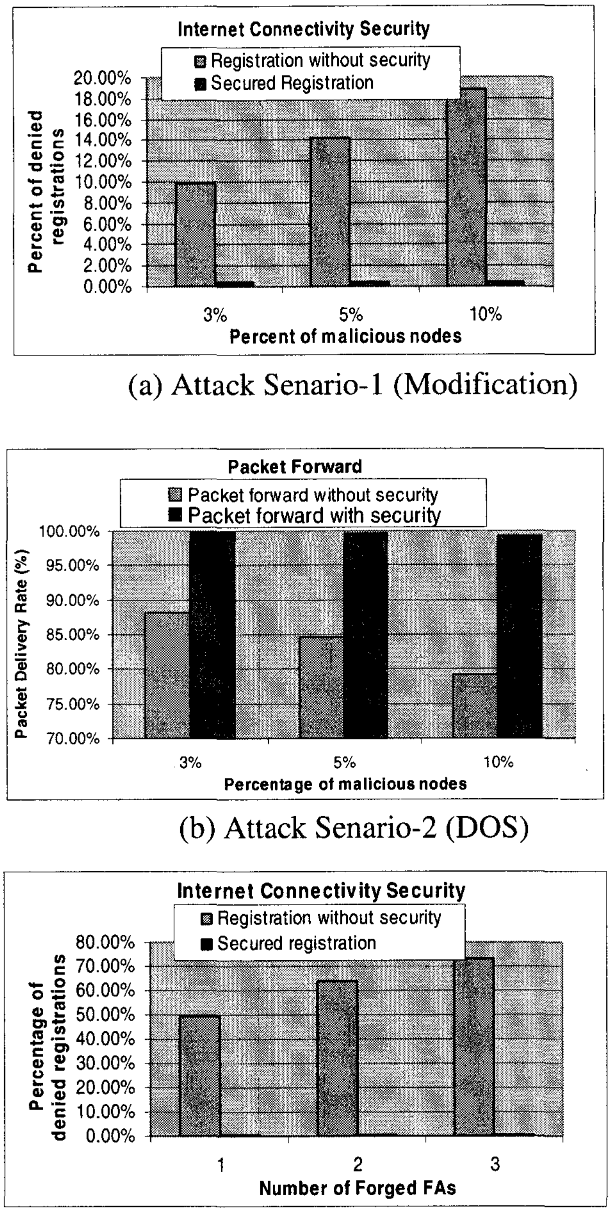
$$\text{Certificate\_MANET}(n) = \sum_{k=1}^{n} \text{Certificate\_MANET}(k) = 2nA_{nh}$$

The communication cost for reissuing **n** certificates in an integrated MANET with n nodes (Certificate_MANET$_{cost}$ (n)) is:

$$\text{Certificate\_MANET}_{cost}(n) = 2nA_{nh}A_{ps}.$$

Experimental result of Security Achievement:

In order to verify the proposed approach, experiments were carried out to evaluate the security achievements and the communication cost on the proposed security protocol. The experiments were conducted by using NS-2 [73] and the tested integrated network is configured as shown in Figure 2. An integrated network, which has a FA and 100 ad hoc nodes are tested for various security attacks. The FA was located in the center of simulation areas and connected to HA by a router. MSs are randomly located in the simulation area. The AODV protocol coordinates with Mobile IP protocol for providing global internet connectivity.



(a) Attack Senario-1 (Modification)

(b) Attack Senario-2 (DOS)

(c) Attack Senario-3 (Forged FA Attack)

Figure 22. Security Attacks and Security Achievements.

The scenarios with the dimensions 1000m x 1000 m are created with different percentage of malicious nodes that are randomly distributed in the simulation area. For the first scenario, the malicious nodes modify registration request from a genuine MS so that the MS cannot obtain the correct Internet connectivity. As shown from Figure 22(a), when the percentage of malicious nodes increases, the more MSs cannot obtain the

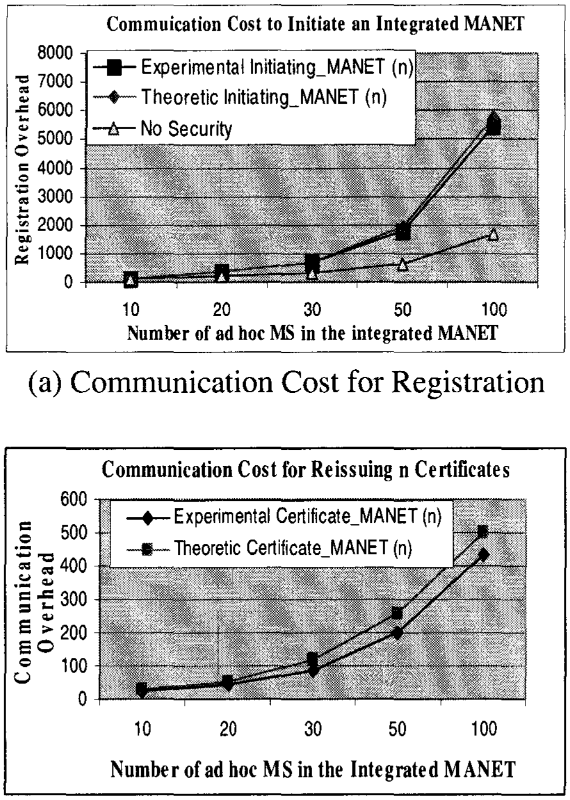
Internet connectivity. However, in the secure registration process, most of the MSs obtain the internet connectivity because of secure registration scheme in the IV.A. Only a few MSs are isolated and cannot reach the FA with correct intermediary nodes.

In the second scenario, we measure the DOS attack, as described in III.B, with the metric of packet delivery ratio, which is the percentage of CBR packets received by the destination in comparison with the number of the CBR packets generated by the source MS. In the DOS attack, the malicious nodes drop the data packets from a MS when they are selected an intermediary node for ad hoc communication. In the experiments, five connections were initiated between five pairs of randomly selected MSs. CBR packets were sent on each connection with the rate of 10 packets per second. If there is no route discovery security, as shown in Figure 22 (b), the packet delivery ratios decrease when the percentage of malicious nodes increases. However, a secure route discovery in Section 4.4.1.2 circumvents the malicious nodes. In this case, the packet delivery ratios are almost not affected by malicious nodes as can be seen from Figure 22 (b).

Also, experiments were designed to test the Internet connectivity with the attack of Forged FA. In the experiments, one, two or three forged FAs are created and randomly distributed in the simulation area. Each forged FA send advertisements to allure MSs to register as illustrated in Section 2.2.6. It can be observed from Figure 22 (c) that a very higher percentage of MS cannot obtain the correct Internet connectivity because of the forged FAs. By following the secure registration in Section 4.4.1, each MS authenticates the forged FA, and finally creates the correct connections to the correct FA. Similar to scenario-1, it also has a few MSs that are temporally isolated and cannot reach any correct intermediary nodes.

Experimental result of communication cost: In the following experiments, different sizes of ad hoc networks (10 to 100 nodes) are tested. The experimental scenarios were set with the following dimensions (i): 670m x 670m for 10, 20 and 30 nodes, (ii) 1000m x 1000 m for 50 nodes and 100 nodes. The communication cost is analyzed based on the following stages:

- Initiating the integrated Internet and MANET with different sizes of networks,
- Certificate reissuing for MSs after initiation.



(a) Communication Cost for Registration

(b) Certificate Issuing

Figure 23. Communication Cost in Different Sizes of Ad hoc Network

Each theoretic value in Figure 23 (a) and (b) is calculated based on an estimated value of A_{nh} and A'_{nh}. For example, observed from a scenario of 20 nodes in a 670 x 670 domain, A_{nh} is larger than 1 but less than 1.5 hops. In the theoretic calculation, 1.5 hops are used for A_{nh} in this scenario. Furthermore, in the theoretic calculations, the A_{nh} values are set as 1.3, 1.5, 2, and 2.25 hops for 10, 30, 50, and 100 nodes respectively. In all experiments, it is less than 20% experimental area covered by FA. In the theoretical calculation, we set A'_{nh} to $1.25A_{nh}$. In Figure 23 (a) and (b), it shows that the overhead either for initiating an integrated network or reissuing **n** certificates increases when the size of MANET increases. As shown in Figure 23 (a), to initiate an integrated MANET without a security has less overhead than the secure scenario because MSs in an unsecured network may hear, even eavesdrop an advertisement from a neighboring without going through the FA. With security, as can been seen from the calculation of Initiating_MANET (**n**), the communication complexity for initiating an integrated network is O ($\mathbf{n}^2$). However, it could not significantly affect the scalability because each MS only registers with FA once when it enters the network. The network maintains the certificates with the communication complexity of O (**n**). This also can be seen from Figure 23 (b) that the overhead for certificate issuing does not show significant additional overhead of security. In addition, the certificate for each MS typically is updated just before the expiration of its certificate.

As seen from Figure 23 (a) and (b), the experimental results are less than theoretic values. It is because the estimated A_{nh} and A'_{nh} values in theoretical calculation are larger than experimental values.

4.6 Conclusion

The integrated Internet and MANET provides the Internet connectivity for MSs and ad hoc communication. The proposed security protocol protects the Internet connectivity and ad hoc route discovery from various attacks. Compared to the security solution for pure ad hoc networks, the proposed security protocol takes advantage of infrastructure-based Internet authentication to enhance the trustworthiness amongst MSs. On the base of overall authentications among MSs, FA and HA, malicious nodes can be effectively excluded from participating ad hoc routing activities. The ad hoc routing security is achieved by the certificates issued by FA. The certificates exclude those unauthorized nodes from misleading FA discovery and ad hoc route discovery. The extensive experiments show the security achievements and the efficiency of the protocol.

PART B

ON SECURE MOBILITY MANAGEMENT IN HETEROGENEOUS MULTI-HOP WIRELESS NETWORKS

CHAPTER V

MULTI-HOP CELLULAR IP: A NEW APPROACH TO HETEROGENEOUS WIRELESS NETWORKS

This chapter proposes a new Heterogeneous Multi-hop Cellular IP (MCIP) network that integrates multi-hop communication with Cellular IP. MCIP increases the coverage of the wireless network and improves the network robustness against adverse propagation phenomena by supporting communication in dead zones and areas with poor radio coverage. MCIP includes three components: location management, connection management and route reconfiguration. Location management is responsible for maintaining the location information for Mobile Stations (MSs) in a local domain. Connection management establishes an initial path for data transmission and a route reconfiguration mechanism is proposed to take advantage of various multi-hop connection alternatives available based on terminal interfaces, network accessibility and topology. Our simulation results show that MCIP performs well in networks of various sizes including scalability, throughput, and packet delay.

6.1 Introduction

In the near future, a large number of Mobile Stations (MSs) will be equipped with multiple radio interfaces for wireless access to the Internet. A multi-mode MS with multiple air interfaces (cellular interface, Bluetooth, IEEE 802.11 and IEEE 802.16 etc) and different data rates will be able to access cellular Base Stations (BSs), WLAN or WMAN Access Points (APs). In this scenario, the integration of multi-hop ad hoc communications with infrastructure based (or single-hop) wireless networks, such as wireless WANs (e.g., 2.5G, 3G, and 4G), wireless LAN (e.g., IEEE 802.11 a/b/e/g and HiperLAn/2) and wireless MANs (e.g., IEEE 802.16), is fundamental to improving the coverage and performance of the integrated network [2]. In addition, multi-hop communications can be used to increase the utilization and capacity of a BS by decreasing the co-channel interference via lowering the transmission power either of the BS or of the MSs [44] [49]. Also, the integration can be useful in achieving load-balancing by forwarding part of the traffic from an overloaded cell to a free neighboring cell [45] [49]. Many communication scenarios may benefit from heterogeneous multi-hop networks with the desirable properties of ubiquitous coverage and higher data rates. The integration of multi-hop communications with infrastructure-based networks is not a simple task, however. It involves numerous challenges including efficient spectrum utilization, integrated routing, Quality of Service (QoS) support, security, and mobility management 3. In this chapter we consider the mobility management and routing problems, which are two basic building blocks required to support the desired seamless handoffs across heterogeneous networks. The proposed Multi-hop Cellular IP (MCIP) protocol integrates multi-hop relaying with Cellular IP [36] [37] and Mobile IP [34] [35]. As in Cellular IP, the MCIP protocol differentiates between local and global domains. A local domain is a local wireless network consisting of MSs, cellular BSs and/or WLAN APs and an Internet Gateway. In order to simplify the description of MCIP, we use the term BS on generic sense to refer to any network attachment point that provides wireless access to MSs, i.e., a cellular BS or a WLAN AP. Also, we define single-hop MSs as the MSs that are able to communicate directly, in a single-hop, with a BS, while multi-hop MSs are the ones that connect to the BS through multi-hop ad hoc routes. MCIP provides mobility support for single or multi-hop MSs inside a local domain, while mobility between local domains or networks is handled through legacy Mobile IP. Moreover, the MCIP maintains the basic features of Cellular IP, as it provides fast and smooth mobility for single hop MS in heterogeneous multi-hop networks. They key contributions of MCIP are:

- A micro-mobility management for heterogeneous multi-hop networks;
- A connection and multi-hop routing configuration scheme for increasing network coverage and capacity.

The rest of this chapter is structured as follows: the background and related work are discussed in Section 6.2. Then, the heterogeneous MCIP network model is proposed in Section 6.3. Section 6.4 describes the details of location management, connection management and route reconfiguration in the proposed MCIP protocol. The MCIP implementation issues are identified in Section 6.5 and experimental results are provided in Section 6.6. Finally, the chapter is concluded in Section 6.7.

6.2 Background and Related Work

5.2.1 Mobility Management Protocols

Mobile IP [34] [35] performs its mobility management with two entities: home agent (HA) and foreign agent (FA). HA and FA advertises their existence by periodically sending advertisement on the home network and foreign network respectively. When a

MS visits a foreign network, it picks a care-of-address (COA) from the beacon message advertised by the FA and initiates a registration procedure by sending a registration request to FA. The FA then forwards the registration request to the HA. After creating the mobility binding for the MS, HA responds to FA with a registration reply message carrying out the registration result. Then, FA forwards the registration result to the MS so that the MS can receive the packets from the correspondent node (CN) via the HA, Internet, and FA.

Mobile IP provides a continuous Internet accessibility for a MS at its visiting foreign network, but it is inadequate in terms of seamless handoff support in micro-mobility environment where transitions between network attachment points could be frequent. Every time a MS changes its attachment point, Mobile IP requires the MS to register with the network and create a mobility binding which causes significant delay during migration. The Cellular IP [36] [37] protocol provides efficient mobility and handoff support for frequently moving MSs. In Cellular IP network, a gateway separates cellular access network from the Internet. In a local cellular access network, there is no need for a MS to register with FA when MS moves from a BS to another, thereby supporting fast handoff using HLR-VLR.

5.2.2 Mobility Management in Heterogeneous Networks

Consider the communication scenario shown in Figure 24. As can be noted, MS3 can not obtain services from the Internet via standard Cellular IP because: (i) MS3 is located outside the direct transmission range of any BS or it is in a poor coverage area; and (ii) standard Cellular IP only supports single hop communication from a MS to a BS. In fact, if MS3 and other intermediate MSs (e.g., MS 2 and MS 1 in Figure 24) were equipped with dual-mode radio interface (e.g., cellular and IEEE 802.11), MS3 could connect to the Internet via a multi-hop route (MS 3-2-1-BS-Internet, as shown in Figure 24). In order to allow such a communicating scenario, a new architecture is needed for providing efficient mobility management and routing schemes that support heterogeneous multi-hop communication.

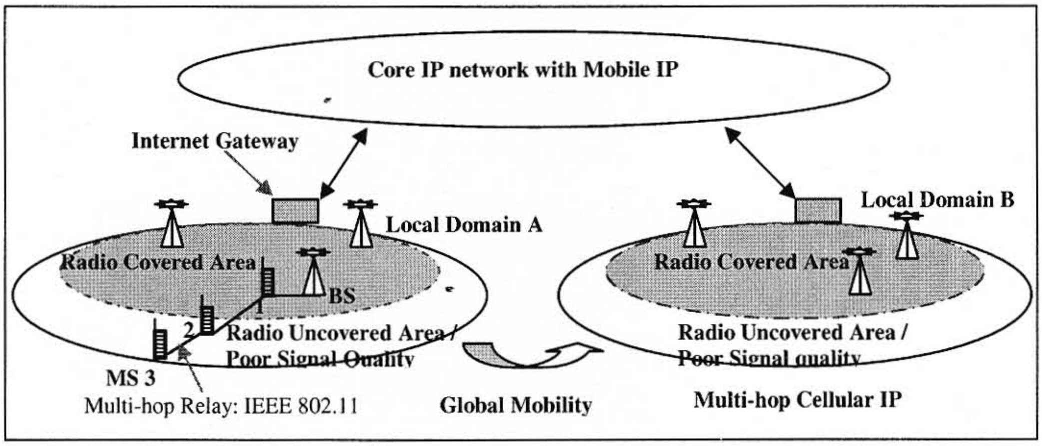


Figure 24. Heterogeneous Multi-hop Access Network and Mobile IP.

Existing mobility management protocols are not optimized for multi-hop heterogeneous networking scenarios. Mobile IP can support global mobility, but it is not designed for micro-mobility management in a local domain as shown in Figure 24. On the other hand, Cellular IP suffers from two key drawbacks in a multi-hop heterogeneous scenario: (i) it does not support communication between heterogeneous wireless networks; (ii) it cannot provide service to MSs that are out of the coverage or have poor signal quality from a BS. Therefore, it is necessary to design new mobility management architectures and protocols that exploit the multi-hop communication paradigm and support seamless mobility in ubiquitous multi-hop heterogeneous scenarios.

If a MS moves outside the coverage of a BS or moves into an area with poor signal quality, the mobility management protocol should have the capability of detecting the possibility of reaching a BS through a multiple-hop path. In general, the integration of multi-hop communications imposes important challenges including location management, route creation and maintenance. The location management requires that single hop or multi-hop MSs periodically report their locations to the network so that they can be reached when data packets are received from the Internet. In addition, a single hop or multi-hop MS must be able to detect and construct a path to a network attachment point (BS) in order to have connectivity "anytime and anywhere". However, the multi-hop routes between MSs and BSs could be easily broken due to the mobility of intermediate MSs. Therefore multi-hop route creation and maintenance are important issues for efficient integration of multi-hop communication and Internet accessibility. The key challenges in this environment can be summarized as follows:

- How to track multi-hop MSs at the fixed network components (BSs)?
- How to create communication route from a multi-hop MS to a BS?
- How to handle route reconfiguration if intermediate nodes in multi-hop route move away?
- How to improve connection quality and performance (which is changing due to mobility) by route reconfiguration?

5.2.2 Mobility Management Related Work

Several routing protocols have been proposed for integration of multi-hop communication in wireless networks. These protocols can be divided into two categories: multi-hop cellular networks and multi-hop WLAN networks. The architectures for the cellular multi-hop networks include A-GSM [43], MCN [44], iCAR [45], MADF [46], UCAN [47], ODMA [48], and SOPRANO [49], and Two-Hop-Relay [50]. On the other hand, architectures for multi-hop WLAN networks include HWN [51], 1-hop and 2-hops Direct Transmission [52].

Multi-hop cellular network (MCN) [44] and Ad hoc GSM (A-GSM) [43] extend the coverage of a BS by using multi-hop relaying. MCN enables two MSs to communicate with each other via multi-hop relaying or through a BS. Ad hoc GSM (A-GSM) supports communication in the dead zones or poor radio coverage areas. Opportunity driven multiple access (ODMA) [48] breaks a single CDMA transmission from a MS to a BS into several multiple wireless hops, thereby reducing the transmission power and co-channel interferences. In UCAN [47], MSs with a low data rate in the downlink channel with the BS can constructs a multi-hop route using relay nodes with better downlink rates to connect to the BS. The iCAR [45]and MADF [46] address the problem of cellular congestion due to unbalanced traffic in a cellular system. The iCAR system diverts the traffic from an overcrowded cell to a neighboring cell that has lower load by making use of dedicated stationary relay stations. Similarly, MADF diverts the traffic of MSs in a hot spot BS to its neighboring cold spot BSs by the relaying through other MSs. SOPRANO [49] advocates self-organization at the physical, data link, and network layers for the purpose of optimizing the capacity of multi-hop cellular network. The Two-Hop-Relay architecture [50] exploits the availability of dual-mode terminals that can act as ad hoc relaying station between a single hop and multi-hop domains.

In 1-hop and 2-hops Direct Transmission [52], besides the option to communicate through the AP, two MSs can directly communicate (one-hop direct transmission), or can use an intermediate relaying station in a multi-hop transmission. The Hybrid Wireless Network (HWN) [51] architecture allows each cell (BS) to select the operation mode between the typical single-hop or in the ad hoc mode. The motivation for this scheme is that a single-hop mode performs better for sparse topologies, while the ad hoc mode is well suited for dense topologies, where in network connectivity can be ensured. The BS runs an algorithm to decide the operation mode that maximizes the throughput based on the topology information received from the nodes. A detailed comparison of multi-hop architectures can be found in 3.

If the Mobile IP protocol [34] [35] is used to support handoffs between BSs in all the above multi-hop cellular and multi-hop WLAN networks, an update message must be sent to the MS's home agent (HA). After creating a new location binding for the MS through the FA, the HA will send the result to the MS via the FA and the visiting BS. This incurs unacceptable delay in the Internet environment. With a large number of MSs, the process of registration at every move generates an unacceptable signaling load across the network. Also, Cellular IP cannot be used for the mobility management scheme integrated networks. A BS in a standard Cellular IP network can only record single hop radio mapping of MS, but it cannot do much for multi-hop MS such as recording, constructing, and maintaining a route.

6.3 Multi-hop Cellular IP Architecture and Design Challenges

5.3.1 Multi-hop Cellular Network Model

This section discusses the architecture of a multi-hop heterogeneous MCIP network, which is shown in Figure 25. The architecture includes HA/FA, Internet Gateway Router (IGW), BSs, and MSs. For clarity, the network is divided into global and local domain. The FA or HA address are used as local domain identifier and IGW address is the Care-Of-Address (COA) for local domain. In the global domain, Mobile IP supports the global mobility for MSs with the granularity of wireless access networks. An IGW, a FA and a set of BSs constitute a local wired network. A local domain is connected to global Internet via an IGW and the BSs are interconnected through wired links and routers. BSs serve as the communication bridges for MSs. Cellular BSs and WLAN APs can be co-located in the hot-spot area as shown in Figure 25. The deployed MS can be a single radio mode or multi-mode with multiple radio interfaces. A dual- mode MS can access a cellular BS by the cellular interface and a WLAN AP by using the WLAN interface. Furthermore, a dual-mode MS can directly communicate with other WLAN capable MS in the ad hoc mode. MSs at a given location could be either within or outside the coverage of cellular BSs and/or WLAN APs. As shown in Figure 25, MS3 is located at the radio-uncovered-area and a multi-hop route (path MS3-MS2-MS1) is required for MS3 to communicate with a BS and obtain the Internet service.

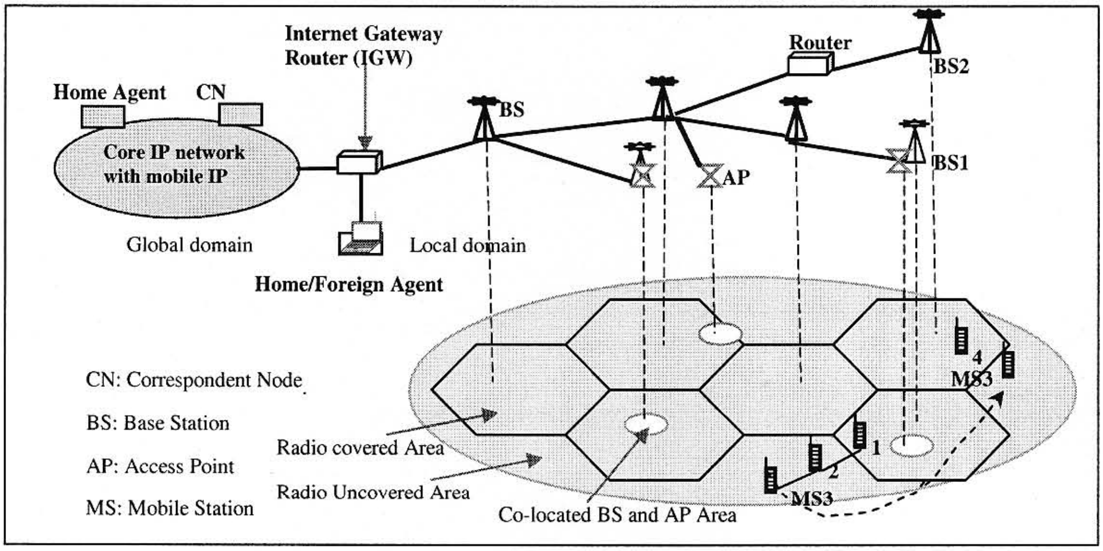


Figure 25. Heterogeneous Multi-hop Cellular IP Network Model.

5.3.2 Design Challenges

There are three critical issues for providing Internet service for MSs in the MCIP network of Figure 25, namely: (i) location management for idle MSs, (ii) connection management, and (ii) multi-hop route reconfiguration for active MSs. Based on the location management scheme, the fixed network, i.e., BSs, can track the location of each MS in the local domain, while each MS can record the accessible BSs for Internet connection. Connection management provides bi-directional Internet accessibility for MSs. It allows the network to locate a single hop or multi-hop MS when the IGW receives the first data packet from the Internet to delivery to a MS. It also allows a single hop or multi-hop MS to set up upstream connection with a CN in the Internet. However, the bi-directional Internet accessibility cannot be achieved in a straightforward way due to the mobility of MSs and multi-hop route. For example, in Figure 25, MS3 constructs a multi-hop route (MS3-MS2-MS1) to BS1 and registers with the FA at a given instant. After some time, MS3 moves to another location as shown in Figure 25, and the established multi-hop route (MS3-MS2-MS1) is broken. At this moment, data packets coming from the Internet for MS3 will be delivered to BS1 and these packets will be dropped because the multi-hop route has been broken. In order to maintain up-to-date location and routing information about all MSs, a significant overhead will incur due to the mobility of MSs that causes frequent changes in the multi-hop routes. Thus, an efficient location and connection management scheme is required to maintain bi-directional Internet connectivity for MSs without excessively overloading the network with location update messages.

Due to MS mobility in a local domain, there is not only the possibility of having broken routes, but also the possibility of finding an improved relaying path for communication. In both cases the original route has to be re-configured to provide improved connectivity. Thus, a route reconfiguration approach is essential for providing efficient communication with higher quality route for data transmission, i.e., less delay, higher speed, or less network congestion. Multi-hop route reconfiguration requires multi-hop routing detection, construction, and maintenance as MSs are communicating with the Internet and moving in a local domain.

Routes between MSs and BSs are needed for location reporting for idle MSs, and are also required for data packet transmission for active MSs. The route used for reporting MS location may be different from the route used for transmitting data packets. Smaller overhead in the network and less delay in reporting location can be achieved if the MS uses a larger range operating air interface. For example, in Figure 27, dual-mode MSs 2 can report its location to BS using cellular radio interface due to longer operating range and shorter delivery latency. However, MS2 can transmit data packets to BS2 via MS1 using the IEEE 802.11 air interfaces to achieve higher communication speed. This separation of routes for location reporting/paging and data packet transmission is proposed in MCIP to provide improved performance and is described in the following section.

6.4 Heterogeneous Multi-hop Cellular IP Protocol

In MCIP, MSs operate in one of two states: idle and active. In the idle MSs are inactive while the active MSs are receiving or sending data packets. Figure 26 illustrates the state machine of a MS. If there is data to be transmitted or received, the MS immediately moves to the active state after starting up the connection management. In the active state, the MS starts the procedure of route reconfiguration as needed. After finishing data transmissions, the MS returns to the idle state again, and it only executes the location management. Figure 26 also shows the basic functionalities provided by MCIP, namely, location management, connection management, and route reconfiguration. In addition, MCIP uses the concepts of multi-hop paging and routing caches to implement location management and route reconfiguration, respectively. Each block in Figure 26 is described in details in the remaining of this section.

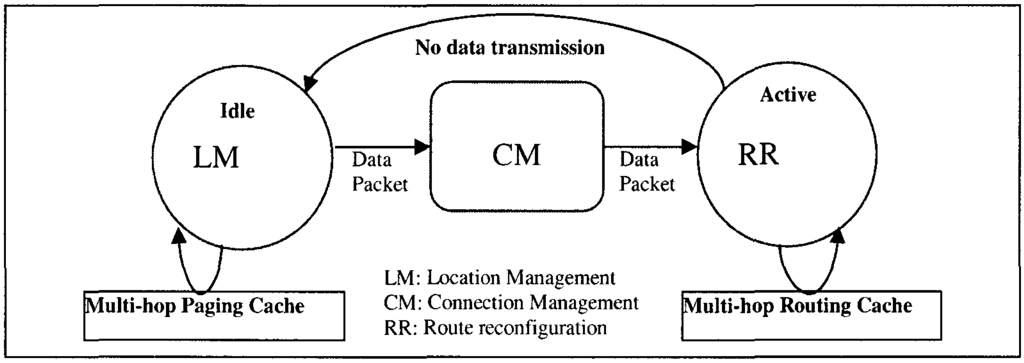


Figure 26. MS State Machine and MCIP functionalities.

5.4.1 Multi-hop Paging/Routing Cache

MCIP uses *multi-hop paging caches* for maintaining the location of MSs and uses *multi-hop routing caches* for keeping the multi-hop routes for data packet transmission. A multi-hop paging/routing cache stores a single or multi-hop route between the IGW and a multi-hop MS. The IGW maintains a paging cache with a route to each of its associated MS. The IGW also maintains a routing cache for each active MS where the route is used to send data packets. At the same time, each MS keeps paging and/or routing cache information, depending on its current state. An idle MS periodically reports its location to its IGW and updates the multi-hop paging cache with the multi-hop route. On the other hand, an active MS keeps an additional cache, the multi-hop routing cache, with the route used for data transmission. Each paging and routing caches have associated timers called Paging Interval Timer (PIT) and Routing Interval Timer (RIT), respectively. A PIT/RIT is initialized at paging/routing cache creation. When a PIT or RIT expires, the corresponding paging or routing cache is cleared. Figure 27 illustrates the multi-hop paging and routing caches stored at the IGW for MS2. As shown in Figure 27, when MS2 is idle, the corresponding multi-hop paging cache stored by the IGW has the path: IGW-BS1-BS2-MS2. In contrast, when MS2 moves to the active state, MS2 or BS2 initiates a route discovery by IEEE 802.11 interface (if available) for constructing a multi-hop route for data transmission. Then, a multi-hop routing cache for MS2, with the route IGW-BS1-BS2-MS1-MS2, will be created after the process of route discovery is completed.

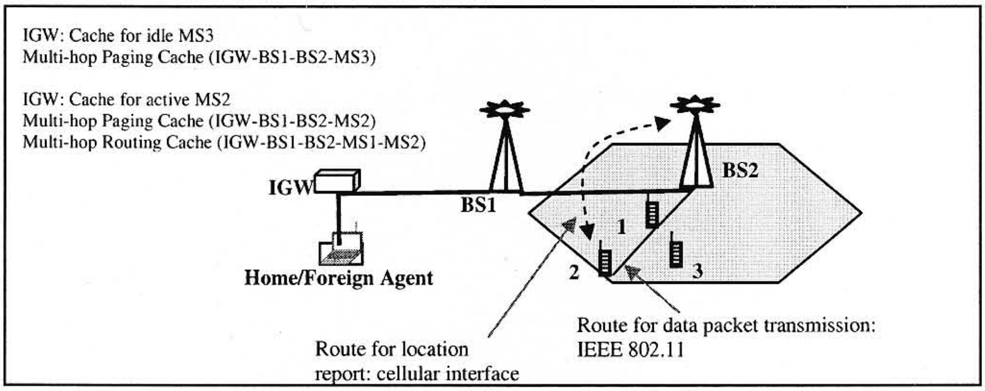

IGW: Cache for idle MS3
Multi-hop Paging Cache (IGW-BS1-BS2-MS3)
IGW: Cache for active MS2
Multi-hop Paging Cache (IGW-BS1-BS2-MS2)
Multi-hop Routing Cache (IGW-BS1-BS2-MS1-MS2)
IGW
BS1
BS2
1
2
3
Home/Foreign Agent
Route for location report: cellular interface
Route for data packet transmission: IEEE 802.11

Figure 27. Multi-hop Paging and Routing Caches.

Each MS creates its multi-hop paging cache when it enters a local domain. In addition, each MS periodically sends page-update packets to the IGW and updates its multi-hop paging cache so that the MSs and the network can track each other. The multi-hop paging cache provides location information for the purpose of downlink connection establishment between the Internet and the MS. Based on the multi-hop paging cache for a MS, the IGW has the knowledge of

- the wired route from IGW to the BS that the MS is associated with,
- the single or multi-hop wireless route from the associated BS to the MS and,
- the number of hops from the associated BS to each MS.

On the other hand, the multi-hop paging cache stored at the MS provides the following information:

- the BS the MS is associated with,
- the number of hops between the MS and its associated BS and,
- the multi-hop route between the IGW and the MS.

The processes of creating and maintaining the multi-hop paging and routing caches are discussed in Section 5.4.3 and Section 5.4.4.

5.4.2 MCIP Overview

MCIP supports several connectivity scenarios and involves different procedures for creating connections in each case. We provide an overview of the protocol's operation in this section, while the details are described in the following sections.

There are two types of connection establishment for a MS: (i) Downlink (Internet->MS), and (ii) Uplink (MS -> Internet). In the first case, when the IGW receives the first data packet from the Internet for a MS, the IGW, based on its paging cache, sends the packet to the BS, where the MS is currently associated. Then, the BS initiates the process of connection establishment towards the MS (see Section 5.4.4). In the second case, a MS initiates the process of connection establishment with the Internet if the multi-hop route from the MS to the Internet stored in multi-hop paging cache is stale. The MS constructs a route using the range-based search process described in Section 5.4.2. After the connection is established, the MS creates a multi-hop routing cache by sending a route-update packet to IGW. Then, the MS enters into an active state and starts up the route reconfiguration scheme.

The route reconfiguration scheme is achieved by updating the multi-hop routing cache for data transmission between the IGW and an active MS. Therefore the path from the Internet to the MS is kept fresh for data transmission. The following events trigger the update of the multi-hop routing cache

- periodic update timer expires,
- the MS performs a handoff between neighboring BSs and
- a new higher quality multi-hop route is identified.

5.4.2 BS/MS Search Algorithm

MCIP involves two important route discovery procedures: the BS search algorithm for locating a MS, and, the MS search algorithm to find the nearest available BS. The proposed search algorithms reduce the search overhead by limiting the search range through a TTL (time to live) field in each search packet. When a MS (BS) detects a broken route to BS (MS) during the transmission of a page-update or data packet, the MS (BS) starts a search process for a new route to the destination. A search-request packet is transmitted by the source through the available air interface (cellular and/or IEEE 802.11) with a search range set by the TTL field. The initial value of TTL is set to the number of hops recorded in the paging/routing cache. If the first attempt fails, the next search range is incremented by 1 until the TTL reaches a maximum hop number (M_{hop}). The interval between issuing two continuous search-request packets is called the Search Interval. Upon receiving the search-request packet, an intermediate MS responds with a search-reply packet if it has a fresh route to a BS. Otherwise, it decreases the TTL by 1, and then forwards the packet as long as TTL is greater than zero. The search-request packet floods through the network within the TTL range. When the destination receives the search-request packet, it returns a search-reply packet with the reverse route.

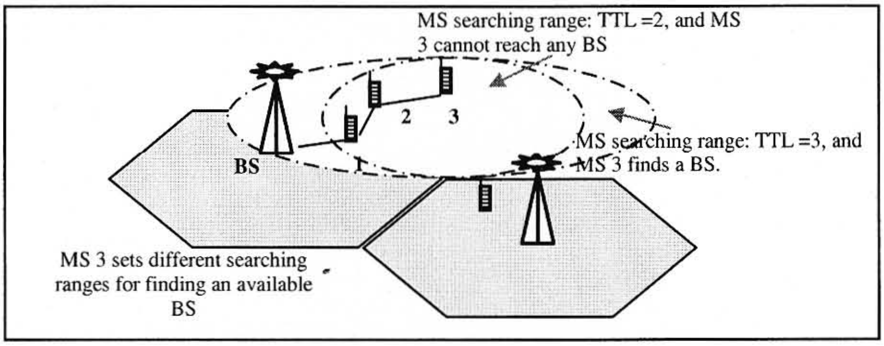

MS searching range: TTL =2, and MS 3 cannot reach any BS
2
3
BS
MS searching range: TTL =3, and MS 3 finds a BS.
MS 3 sets different searching ranges for finding an available BS

Figure 28. MS Search Process for Locating an Available BS.

Figure 28 describes a MS search process. When MS3 detects that it has lost connectivity with the BS in Figure 28, it sends a search-request packet with TTL=2, which is the number of hops recorded in its multi-hop paging/routing cache. If the search-request packet cannot reach any available BS, as shown in Figure 28, MS3 then enlarges its search range by setting a larger TTL (TTL=3). The searching process will continue until the largest search range (M_{hop}) or a reply message is received from a BS. The BS closest to the MS (in number of hops) responds with a search-reply packet. The search-reply packet will be forwarded to the MS3 via the reverse multi-hop route. After the MS search process, a multi-hop path will be established between the MS and the responding BS. The multi-hop path can then be used to update multi-hop paging and routing caches.

Figure 29 illustrates a similar search process that is executed by a BS to locate a destination MS3. As the path stored in the multi-hop paging/routing cache for a MS becomes stale and the BS needs to send data to the MS, the BS transmits the first search-request packet with the search range (TTL) set to 2, as recorded in the multi-hop paging/routing cache for MS3 (see Figure 29). If the search-request cannot reach the destination MS, the BS enlarges the search range (TTL = 3 hops). As MS3 responds with a search-reply, the BS updates the downlink path to MS 3 (MS1-MS2-MS3).

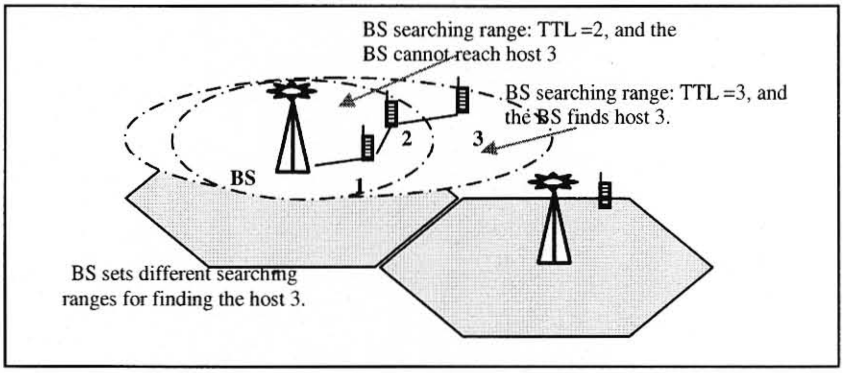

BS searching range: TTL =2, and the BS cannot reach host 3
BS searching range: TTL =3, and the BS finds host 3.
BS
1
2
3
BS sets different searching ranges for finding the host 3.

Figure 29. BS Search Process to Locate a Destination MS.

5.4.3 Location Management

Location management allows the fixed network (IGW and BSs) to track each MS in its local domain. If the MS has not yet registered with the HA via the FA, it initiates the Mobile IP registration process and then sends a page-update packet to the IGW, which replies with page-reply packet after creating the multi-hop paging cache for the MS. As illustrated in Figure 30, MS2 constructs a multi-hop route to BS3 via MS1 after the search process is completed. The page-update packet is forwarded to IGW through the path MS1-BS3-Router-BS2-BS1. After creating a multi-hop paging cache for MS2, a page-reply packet is sent by the IGW to MS2 via the reverse path.

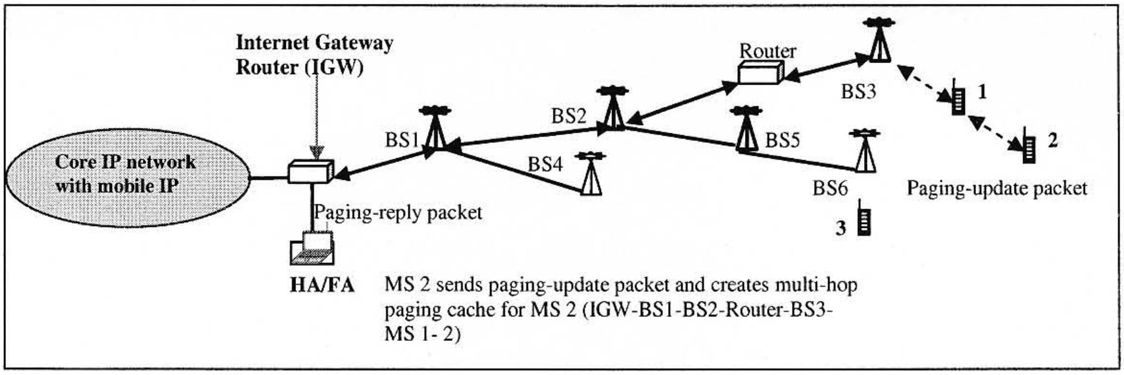

Internet Gateway
Router (IGW)
Router
BS3
1
2
BS2
BS1
BS5
BS4
Core IP network with mobile IP
BS6
Paging-update packet
Paging-reply packet
3
HA/FA
MS 2 sends paging-update packet and creates multi-hop paging cache for MS 2 (IGW-BS1-BS2-Router-BS3-MS 1- 2)

Figure 30. A Page-update Packet Creates a Multi-hop Paging Cache.

When an idle MS moves and the path established via a previous BS is lost, the MS initiates a new search process to locate a new nearest BS and a new multi-hop paging cache will be created by a page-update packet. The old multi-hop paging cache will be cleared after the expiration of the PIT. For instance, in Figure 31, when MS2 moves away from BS3, the route from MS2 to BS3 is broken. Hence, MS2 initiates a search process, and then receives a reply from BS6, which has the shortest route to MS2. A new page-update packet will be forwarded to the IGW via BS6 as shown in Figure 31. A page-reply packet will be returned to MS2 from the IGW through BS6 after updating its multi-hop paging cache. The paging routing cache at BS3 and router will be automatically cleared after the RIT expires.

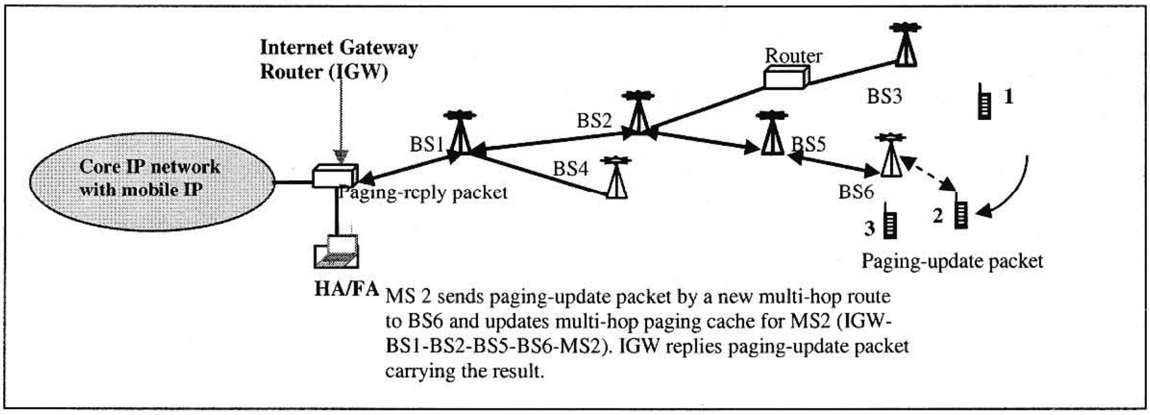
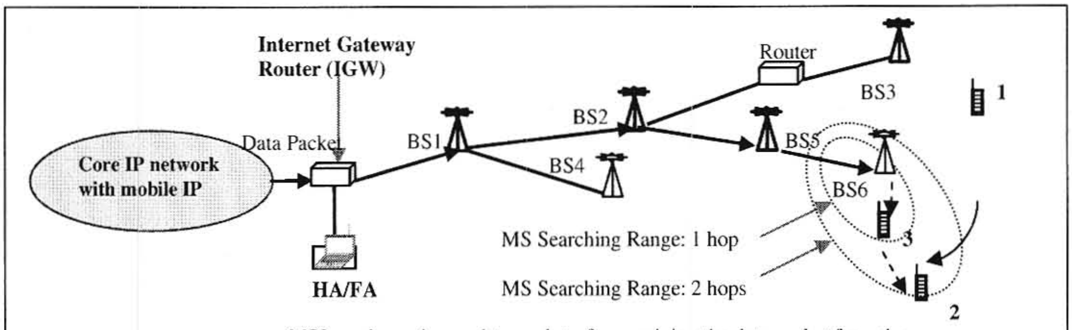

Internet Gateway
Router (IGW)
Router
BS3
1
BS2
BS1
BS5
BS4
Core IP network with mobile IP
Paging-reply packet
BS6
3
2
Paging-update packet
HA/FA
MS 2 sends paging-update packet by a new multi-hop route to BS6 and updates multi-hop paging cache for MS2 (IGW-BS1-BS2-BS5-BS6-MS2). IGW replies paging-update packet carrying the result.

Figure 31. Update of Multi-hop Paging Cache for a Moving MS

5.4.4 Connection Management

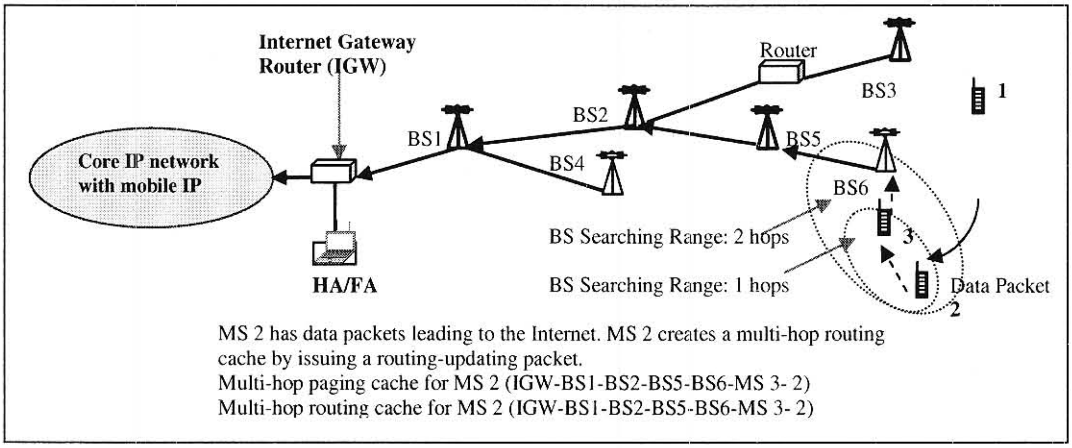
Downlink Connection (Internet → MS): When a data packet for a MS arrives at IGW via the Internet for the first time, the location information in the multi-hop paging cache is used to establish the downlink connection with the MS. First, the IGW forwards the data packet to the BS through which the MS is currently accessing the network. Then, the BS delivers the packet to the destination MS by following the multi-hop paging cache route for the MS. However, the route in the multi-hop paging cache for the MS may be stale, because of the mobility of MSs. In this case, the BS initiates a BS search process, as described in Section 5.4.2, to find a new route to the destination MS. For example, in Figure 31, the data packets for MS2 are initially forwarded to BS6. However, MS2 moves to a new position before the connection is established, as shown in Figure 32, and BS6 cannot reach MS2 directly as recorded in its multi-hop paging cache for MS2. Then, BS6 uses the search procedure to find a new route to MS2. Eventually, the data packets will be delivered to the destination as a new multi-hop route is created via MS3. Then, after receiving the first data packet from the Internet, MS2 turns to active state and sends a route-update packet to the IGW for creating multi-hop routing cache for MS2. Meanwhile, MS2 starts up the route reconfiguration scheme described in the next section for exploring an improved path.


Internet Gateway Router (IGW)
Router
BS3
1
BS2
Data Packet
BS1
BS5
Core IP network with mobile IP
BS4
BS6
MS Searching Range: 1 hop
3
HA/FA
MS Searching Range: 2 hops
2

MS2 sends routing-update packet after receiving the data packet from the Internet. A multi-hop routing cache for MS 2 is created.
Multi-hop paging cache for MS 2 (IGW-BS1-BS2-BS5-BS6-MS 3- 2)
Multi-hop routing cache for MS 2 (IGW-BS1-BS2-BS5-BS6-MS 3- 2)

Figure 32. Creating Multi-hop Routing Cache at the First Data Packet from the Internet.

Uplink Connection (MS → Internet): In the case of uplink connections, the MSs forward the packets to their IGW by following its multi-hop paging cache. The IGW forwards the packet to the destination by following the Mobile IP protocol. Because of the mobility of MSs, the route in the multi-hop paging cache for the MS may have been broken since the last update. In this case, the MS initiates a search process to find a path to the nearest BS. After a new route is found, the MS turns to active state and creates its multi-hop routing cache by issuing a route-update packet to the IGW. The IGW responds to the MS with a route-reply packet after creating the multi-hop routing cache for the MS. Consider the example in Figure 32, where MS 2 tries to establish an uplink connection with the Internet, but its paging routing cache has stale information, which indicates BS6 as the next hop to forward uplink packets (MS2 was under coverage of BS6, as shown in Figure 31, but it has moved to the new position shown in Figure 32). Because of the stale routing information, BS6 cannot receive the packets sent by MS2, and the link layer of MS2 can detect the link breakage and report it to network layer. Then MS2 initiates a search process to find a path to an available BS. After a new route has been constructed as illustrated in Figure 33, the data can be delivered to the IGW, which forwards the data to the corresponding node in the Internet. Meanwhile, MS2 creates a multi-hop routing cache by issuing a route-update packet, to which, the IGW responds with a route-reply packet that is forwarded through the reversed path obtained from the route-update packet. Thereafter the multi-hop routing cache is maintained by route reconfiguration scheme during the period of packet transmission.



MS 2 has data packets leading to the Internet. MS 2 creates a multi-hop routing cache by issuing a routing-updating packet.
Multi-hop paging cache for MS 2 (IGW-BS1-BS2-BS5-BS6-MS 3- 2)
Multi-hop routing cache for MS 2 (IGW-BS1-BS2-BS5-BS6-MS 3- 2)

Figure 33. Creating Multi-hop Routing Cache at the First Data Packet Sending to the Internet.

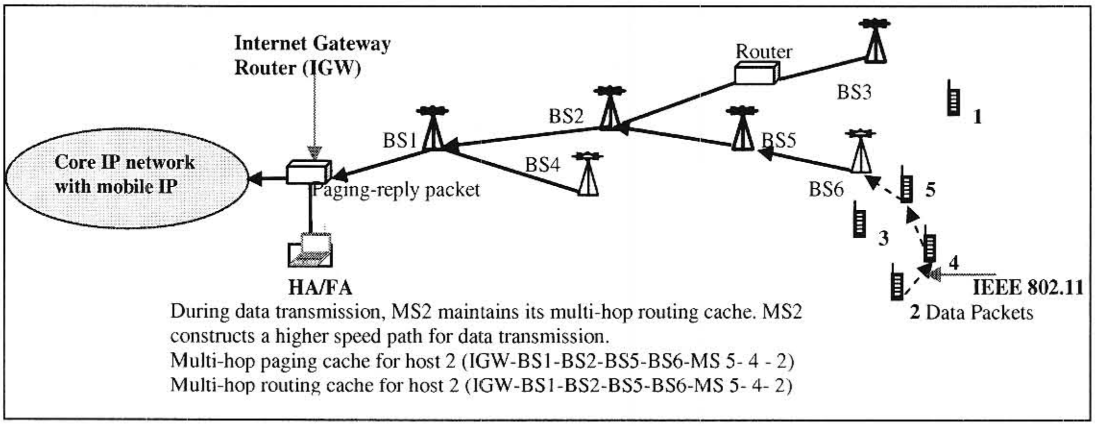
5.4.5 Route Reconfiguration

A multi-hop route may include wired and multi-hop wireless links. The wired part of the route is the path between the IGW and the BS that the MS is currently associated with, and the wireless part of the route is the path between the BS and the MS. The route reconfiguration procedure periodically updates both the wired and wireless parts of a route in the multi-hop routing cache. To update a multi-hop routing cache, a MS sends out a route-update packet to the IGW before the expiration of the RIT. After updating its multi-hop routing cache, the IGW replies to the MS with a route-reply packet. The route reconfiguration has two main goals:

- To keep the multi-hop routing cache fresh and,

- To improve the quality of connection between the multi-hop MSs and BSs if possible.

Recall that the multi-hop route stored in multi-hop routing cache is used for data transmission. If this multi-hop route becomes stale, data packets will be lost. The route reconfiguration has the responsibility of rediscovering a new path if the current route is broken. The multi-hop routing cache is updated after the route discovery process and the data packets are redirected to the new path. The discovery of higher performance path during route reconfiguration is based on BS's accessibility, terminal interfaces, and network topology. For instance, if a dual-mode MS detects a higher speed transmission path via IEEE 802.11 interface, then it can construct a new and improved path for communication. As shown in Figure 32 and Figure 33, MS2 establishes an initial connection through path MS2-MS3-BS6. When a route reconfiguration scheme is performed after the arrival of MS4, MS2 detects a higher speed route through MS2-MS4-MS5-BS6 as shown in Figure 34.



During data transmission, MS2 maintains its multi-hop routing cache. MS2 constructs a higher speed path for data transmission.
Multi-hop paging cache for host 2 (IGW-BS1-BS2-BS5-BS6-MS 5- 4 - 2)
Multi-hop routing cache for host 2 (IGW-BS1-BS2-BS5-BS6-MS 5- 4- 2)

Figure 34. Update of Multi-hop Routing Cache During Packet Transmission.

In order to maintain the connection for data transmission, the following events result in an update process of the multi-hop routing cache of a MS:

Link Broken: When a link breakage toward a BS (MS->Internet) is detected at the link layer and reported to the network layer, the MS suspends the process of packet transmission and initiates a MS search process to construct another multi-hop path to an available BS. After receiving the search-reply from a BS, the MS sends out a route-update packet for the purpose of updating its multi-hop routing cache. If new BS is different than the old BS, the new data migration process implements handoff in local domain. On the other hand, if a link toward a destination MS (Internet->MS) is lost, the BS suspends the process of data transmissions and issues a BS search process to locate the destination MS. The MS responds the BS search process with search-reply. At the same time, the MS initiates the process of updating its multi-hop routing cache by sending a route-update packet. In the end, the BS restarts the process of packet transmission.

New BS: A moving MS keeps a record of available BSs, even though the current data transmission path may be fresh. For a single hop MS, it keeps listening to the beacons from BSs and determines whether a migration to a different BS should be encouraged or not. For a multi-hop MS, it detects an available BS by a MS search process. If a BS with fewer numbers of hops is found, the MS updates the multi-hop routing cache so that the data packets could be forwarded through the new BS. If a multi-hop MS receives a beacon directly from a BS, it switches the current connection from multi-hop to single hop. At the same time, the MS updates its multi-hop routing cache. In this case, a multi-hop hop communication has been migrated to single-hop. The processes of BS detection and data packet migration implement the hand-off in the local domain.

Higher quality multi-hop route: Due to the dynamic nature of multi-hop networks, the initial path determined by the connection management may not provide the desired performance. Therefore, it is essential to have a route reconfiguration protocol for MCIP, which will be responsible for: (i) exploring the path with higher performance; (ii) redirecting dataflow from a path to another. Four possible connection alternatives between a MS and a BS are proposed for higher performance in the MCIP network. Figure 35 illustrates various connection alternatives (I-IV) for data transmission.

I. MSs located inside the radio coverage of a BS may communicate with the BS directly, e.g., MS1, MS2, MS3, MS11, and MS14 in Figure 35.

II. MSs (e.g., in Figure 35, MS4, MS5, and MS6) located outside the radio coverage of a BS can established multi-hop routes to the BS with the help of other MSs. This kind of multi-hop route extends the service coverage of the BSs. For example, in Figure 35, MS5 contacts BS1 via the path <MS5, MS4, MS2, and BS1>.

III. A single hop connection from a MS to a BS can be broken into several wireless hops. For example, in Figure 35, MS7 is under the coverage of BS1; and the path from MS7 to BS1 can be broken into a multi-hop route <MS7, MS3, BS1>. This connection alternative could reduce the transmission power and co-channel interferences.

IV. A MS located inside BS's coverage area may communicate with a neighboring BS by using multi-hop relaying. For example, in Figure 35, if BS1 is overloaded and BS2 has a small traffic load, MS8 could communicate via BS2 using the multi-hop path <MS8, MS9, MS10, BS2>. Otherwise, the communication of MS8 would be blocked in the BS1. The intermediate nodes (e.g., in Figure 35, MSs 9 and 10) could be some user terminals or some stationary relays. This feature relives the cellular congestion by diverting traffic from an overcrowded cell to a neighboring free cell.

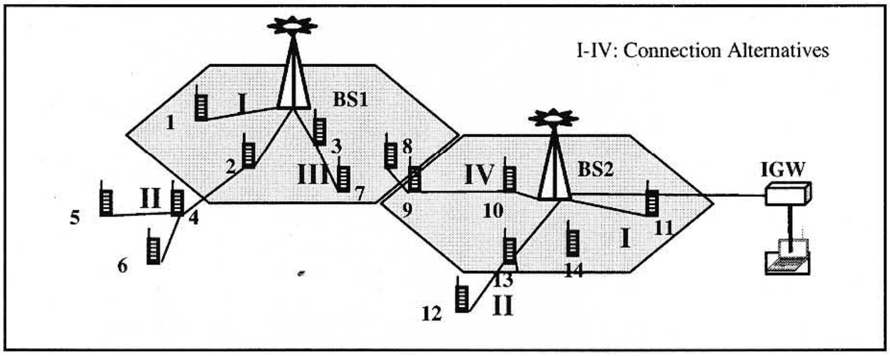


Figure 35. Connection Alternatives.

A MS may migrate from one connection alternative to another depending on the quality of the connection. When a MS detects a connection alternative with higher performance, the MS issues a route-update packet for the purpose of updating its multi-hop routing cache. Then, the MS or BS redirects the data packets to the new connection. If a dual-mode MS experiences low performance using its cellular interface, the MS initiates a MS search process using another radio interface (e.g., IEEE 802.11). After a new path is found, the MS updates its multi-hop routing cache and redirects its dataflow to its new interface. For instance, in Figure 36, MSs 1 and 2 are dual-mode MSs having cellular and IEEE 802.11g interface. Initially connection management for MS1 establishes the single hop connection to the BS as shown in (Connection Alternatives I). As MS1 moves, the route reconfiguration protocol detects that an AP can be reached through MS2 via its IEEE 802.11 interface. Then, MS1 constructs a multi-hop route to the AP (Connection Alternatives II). After updating the multi-hop routing cache, the data flow of MS1 migrates from its cellular interface to IEEE 802.11 interface.

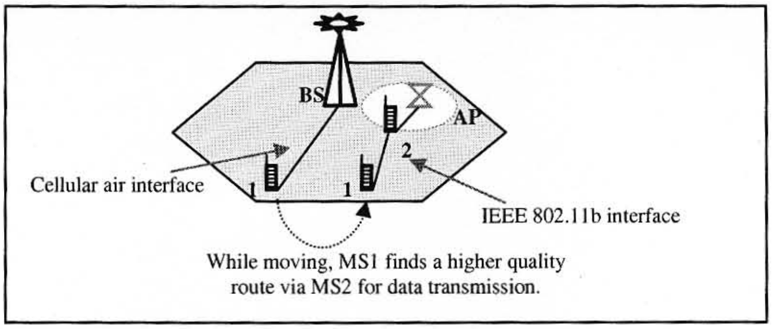

BS
AP
Cellular air interface
1
1
2
IEEE 802.11b interface
While moving, MS1 finds a higher quality route via MS2 for data transmission.

Figure 36. Connection Alternative Migration.

5.5 Protocol Implementation Issues

In MCIP, the IGW maintains a multi-hop paging cache for each idle MS at all time. On the other hand, an active MS maintains its multi-hop routing cache fresh. The MS also takes advantage of various communication alternatives for interacting with the heterogeneous MCIP environment. The implementation of such architecture is influenced by many design factors. The design factors considered here include the interaction with BS beacon advertisements, control packet overhead, mobility and PIT/RIT selection. This section provides an overview of these design factors and their impact on our implementation.

5.5.1 BS Beacon Advertisement

A BS in the MCIP network advertises its presence periodically with beacons carrying FA/HA address, IGW address, and its BS address with a sequence number. The MSs in the transmission range of a BS can directly receive the beacons. The MS outside the radio coverage could obtain beacons by flooding. However, the flooding of beacon may result in the overhead of relaying beacons periodically. Instead of this proactive solution, in MCIP, MSs obtain beacons in an on-demand manner. MSs outside the radio coverage of any BS obtains the BS information by querying its neighboring MSs. For instance, when a MS outside the any BS is switched on, it sends a BS-query packet out. The neighboring MSs having the fresh BS information responds the MS with BS-query-reply packet and the route to the BS.

5.5.2 Control Packet Overhead

MCIP uses eight control packets: page-update packet, page-reply packet, route-update packet, route-reply packet, search-request packet, search-reply packet, BS-query packet, and BS-query-reply packet. These packets are regular IP packets with new options so that the packets can be understood in a local MCIP domain. The packets never reach outside the local domain. These packets have the MS identifier, which can include the MS's IP home address. Because the search-request and search-reply are broadcast packets, they have a broadcast identifier to avoid duplicate transmission and reception during forwarding. When an intermediate MS receives search-request packet, it appends its address in the packet for the purpose of constructing the reverse multi-hop route to the destination. The overhead of control packets in a multi-hop communication is influenced by many factors, including mobility of MS, PIT/RIT, BS/MS search algorithm and its parameters (e.g., the maximum search range), the network topology, and density of MSs. Following strategies are used in our implementation to control the overhead including:

- Not only paging/route-update packets but also registration/routing/data packets are used to create or update multi-hop paging/routing caches.
- A single paging/routing/data packet updates multiple paging and routing cache not only for the sending MS but also for all the intermediate relaying MSs. For example, in Figure 30, the page-update packet sent by MS2 can result in the

multi-hop paging updates for both MS2 and MS1. In this case, MS1 has no need to send an extra page-update packet until the next PIT expiration.

- When a MS is powered on without any BS information (e.g., beacon), it starts up a query process to obtain BS information. In order to reduce the number of MS searching process, a MS learns BS information from control/data packets and update its multi-hop paging/routing cache cither sent by the MS or other MSs. For instance, in Figure 32, the BS search process at BS6 results in the update of multi-hop paging/routing cache of MS2. In addition, MS3 obtains the path to BS6 from this process and the multi-hop paging/routing cache for MS3 can also be updated since the MS3 acts as the intermediate relaying MS for reaching MS2.

5.5.3 Mobility and Paging/Routing Interval Timer (PIT /RIT)

PIT and RIT are two key parameters that determine the performance of the MCIP protocol. The MSs having a higher value of Paging/Routing Interval report their locations to IGW with less frequency causing smaller overhead in the wireless network. If the Paging/Routing Interval is large for frequently moving MSs, it is more difficult for the network to locate the MSs when required. In case of a MS moving away from the maximum search range of BS without reporting its location, the Internet connectivity will be lost until the next update of multi-hop paging cache. Compared with a higher value of PIT/RIT, a lower PIT/RIT value will result in heavier overhead in the wireless network because of frequent control packet transmissions. Mobility is the key factor in deciding the best values for PIT and RIT. The higher the mobility of MS, the lower the PIT/RIT should be.

5.5.4 Hand-off

In the local domain of a heterogeneous MCIP network, the handoff of MS from a BS to another is called intra-domain-handoff. The processes illustrated in the Figure 30 and Figure 31, where MS2 migrates from BS3 to BS6, implement the intra-domain-handoff. In this case, the MS updates its multi-hop paging cache through a new BS. On the other hand, for an active MS, the intra-domain-handoff is implemented by route reconfiguration scheme that is illustrated in Section 5.4.5. In this case, the MS updates its multi-hop routing cache and redirects its dataflow via the new BS. If a MS moves from a local domain to another local domain, the handoff is called inter-domain-handoff. The inter-domain-handoff is implemented through regular Mobile IP. In a dual-model MS, the service migration between two different radio interfaces, e.g., from a cellular interface to IEEE 802.11 interface, is called vertical handoff. The basic procedure of a vertical handoff has three steps:

- Improved path detection: A dual-model MS communicates with a BS (BS1) with an air interface (address: IP1) and detects a single or multi-hop route to another BS (BS2) in a new air interface (address: IP2), which provides higher performance such as data rate and power consumption.
- Address & route binding: The MS issues a routing-paging packet to BS2 including IP1 and IP2. If the BS2 accepts the handoff request, BS2 forwards the request to the IGW so that the IGW can create a mapping from IP1 to IP2. Meanwhile, after creating a multi-hop routing cache that is associated with IP2, the IGW replies a route-reply packet to the MS through BS2. The multi-hop

routing page will be cleared after the expiration of the PIT that is associated with IP1.

- Flow redirection: After the process of address and route binding, all packets from the Internet with IP1 will be forwarded to the MS through the new interface (IP2) and the new BS (BS2).

As illustrated in Figure 35, a MS redirects a connection from its cellular interface to its IEEE 802.11 interface when the MS obtain a route to an AP.

5.6 MCIP Performance

We have performed a simulation-based analysis of the MCIP protocol, in which we have measured three commonly used performance metrics:

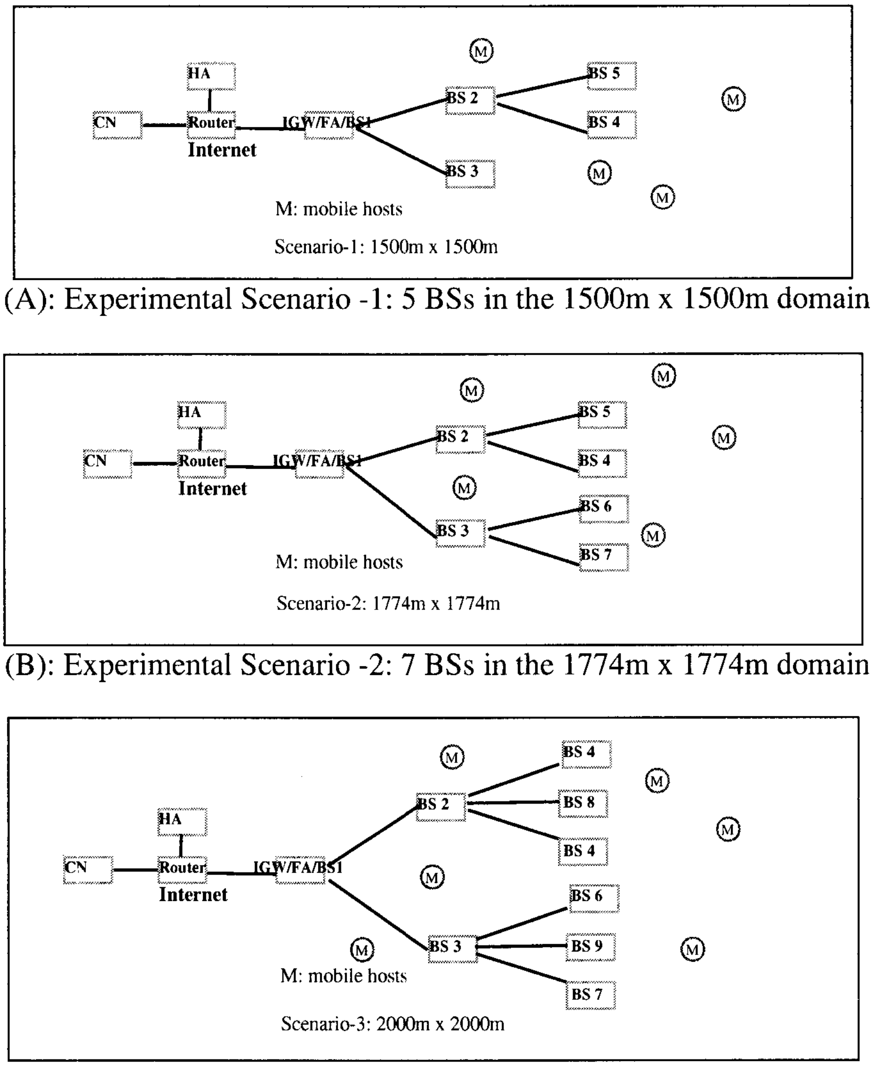
- **Routing Overhead**: The average number of transmitted control packets for maintaining the bi-directional Internet accessibility per MS in the network during a given period.
- **Packet Delivery Fraction:** The percentage of data packets received by the destination in comparison with the number of data packets generated by the source.
- **Average Packet Delay:** This is the average packet delay between a MS and the IGW, which includes the wired routing delay and wireless transmission delay in a local domain.

5.6.1 Experimental Configuration

The experiments were conducted in the NS-2 [73] simulator. In the experimental implementation, a link is considered as broken when three continuous transmissions fail over the link. The link layer detects the link state and reports the network layer when a link is broken. Figure 37 illustrates the experimental scenarios considered. The dimensions of each scenario are as follows:

- Scenario -1, 1500m x 1500m simulation area with 5 BSs,
- Scenario -2, 1774m x 1774m simulation area with 7 BSs,
- Scenario -3, 2000m x 2000m simulation area with 9 BSs.

As shown in Figure 37 (A), (B) and (C), a router represents the Internet with certain delay. Each router connects to HA, CN and IGW/FA/BS1. In all scenarios, the BS1 is not only configured to act as an access point, but it also implements the functionalities of the IGW and the FA. In each local domain, the BSs are interconnected by wired interfaces with a fixed delay. The three different domains (scenarios) represent the different sizes of multi-hop networks. Each domain has several BSs (5, 7, and 9) and different number of MSs (75, 105, and 133) so that the three cases have the same MS density. The MSs move according to the random waypoint mobility model and the maximum speed of a MS is set between 1 to 25 m/s during the lifetime of simulation runs according to a uniform random distribution. The pause time is consistently 10 seconds between each movement. The transmission ranges of each MS and BS are set as 250m. Table 6 shows the summary of some simulation parameters used in the experimentation. Each data point shown in the following figures is averaged over five runs with different seeds and random node distributions. Each hop-wise wireless transmission of a packet is counted as one transmission in all experiments. In all experiments the simulation runs for 600 seconds.



(A): Experimental Scenario -1: 5 BSs in the 1500m x 1500m domain

(B): Experimental Scenario -2: 7 BSs in the 1774m x 1774m domain

(C): Experimental Scenario -3: 9 BSs in the 2000m x 2000m domain

Figure 37. The Experimental Configurations.

Table 6: Experimental parameters

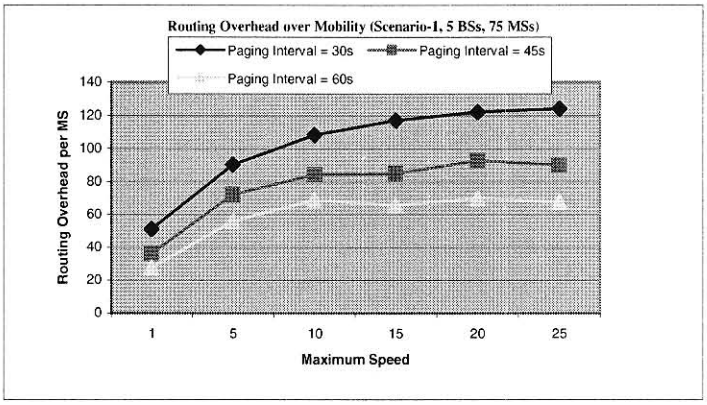
Parameters	Setting
BS beacon Interval	10.0 s
BS/MS Search Interval	1.5 s
Paging Interval Timer (PIT)	30s, 45s, 60s
Paging timeout (for clearing multi-hop paging cache)	60s, 90s, 120s
Routing Interval Timer (RIT)	3s
Raging timeout (for clearing multi-hop routing cache)	6s
Number of MSs	75, 105, 133
Maximum Search Range (M_{hop})	4 hops

5.6.2 Overhead for Bi-directional Internet Connectivity

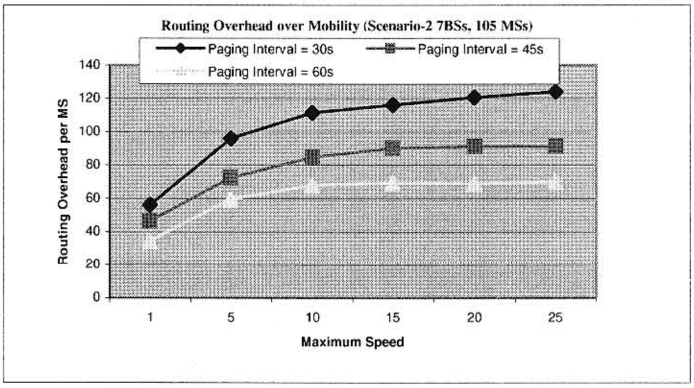
The first goal of this experiment is to measure the overhead for maintaining bi-directional Internet accessibility under different mobility levels (maximum speed varying from 1m/s to 25m/s). It can be observed from Figure 38 (a), (b) and (c) that when mobility increases the routing overhead per MS increases. For the same paging interval in each scenario, the routing overhead per MS increases sharply at lowers mobility. It can also be observed that when MSs move faster than 15 m/s, the routing overhead saturates. This occurs because at higher speeds, the topology changes significantly during the paging intervals. Thus, there is no significant different in control packet overhead between speeds of 15 m/s and 25 m/s.

The next goal of this experiment is to test the effect of paging interval on the routing overhead of maintaining the bi-directional accessibility. In the three scenarios (see Figure 38), the paging cost is highest when the paging interval is 30 seconds while the paging cost is lowest when the paging interval is 60 seconds. Mobility is the key factor affecting the paging interval. Higher mobility MSs should have a smaller paging interval, while lower mobility MSs can have a larger paging interval. The value of the PIT of a MS is typically on the MS mobility time scale.

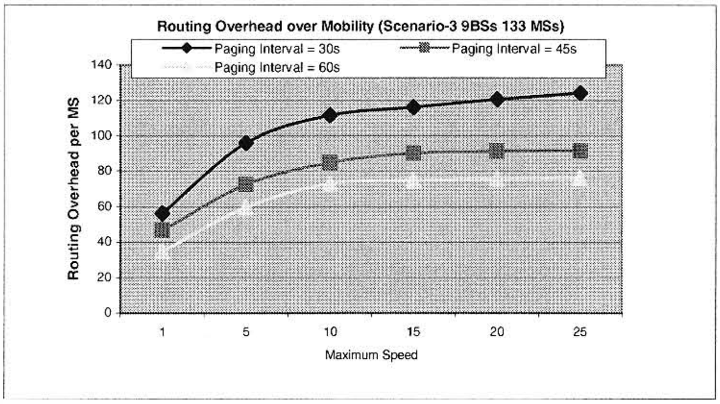
The third goal of this experiment is to test protocol scalability. Scalability refers to the property of maintaining the routing overhead per MS constant, as the size of a MCIP network increases. In a local domain, there may be many BSs but the overhead for each MS to maintain its route should not increase with the growth of the network. Because the proposed MCIP limits each BS/MS search range by setting a TTL value, the MSs outside the search range have no impact on the BS/MS search overhead. In the experiment, different sizes of networks (1500m x 1500m < 1774 m x 1774 m < 2000m x 2000m) were used and the MSs were randomly distributed over the simulated area. In order to test the network scalability, these networks are set with same MS density and BS radio coverage rate. Also these networks run with the same MCIP parameters (e.g., Paging Interval). It can be observed from Figure 38 (A) (B) (C), the average number of transmission per MS in each network is almost the same when the PIT and maximum speed are the same. It can be seen that when the maximum speed is 5 m/s and the PIT is 60s, the average number of transmission is about 60 in each of the three scenarios in spite of the different sizes of networks.



(A) Scenario -1



(B) Scenario -2



(C) Scenario -3

Figure 38. Overhead for Bi-directional Internet Connectivity.

5.6.3 Performance of Data Transmission

In order to observe the performance of data transmission in the MCIP network, connection alternatives I, II and IV (see Figure 35) are analyzed in each scenario. Five constant bit rate (CBR) traffic flows from a source to a destination node were simulated. The CBR data packets size is 512 bytes and the sending rate is 10 packets per second. In the five connections, two connections are initiated by the CN and three randomly selected MSs start three connections. As shown in Figure 39, the packet delivery fraction in each scenario drops when mobility increases. High mobility causes more frequent connection breaks and routing paths changes. However, in general, multi-hop connections achieve a very high packet delivery rate. It can be observed from Figure 39 that there is not a significant difference in packet delivery fraction for the three scenarios of our experiments. This shows the scalability of the protocol.

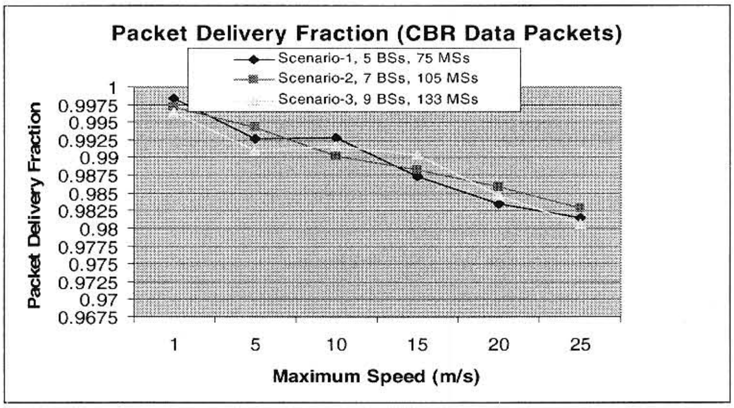


Figure 39. Packet Delivery Fraction.

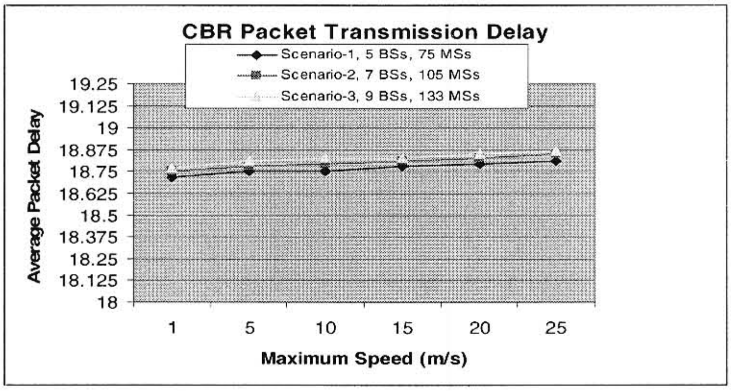


Figure 40. Packet Delay.

Recall that the packet delay refers to the average transmission latency of CBR a data packet between a MS and the IGW. This includes the wired routing delay between the IGW and the BS that the MS is associated with. This also includes the multi-hop wireless transmission delay between the associated BS and the MS, which is affected by buffering during multi-hop route discovery, queuing at the interface queue, retransmission latency at the MAC layer, and propagation delay. Figure 40 shows that the average delay increases slightly as the mobility increases. As described in Section 5.4.5, the MS always uses the BS having best metrics (the shortest path in the experiments) for communication with the Internet. When the MS moves, the route reconfiguration detects a new BS (a BS having shortest multi-hop path) for communication. Therefore, the movement of the MS causes no significant increase of the average packet delay. The increase shown in Figure 40 is primarily caused by the buffering during connection establishment. Higher mobility causes more frequent broken link during the period of CBR transmissions, and the MSs with higher mobility run the route reconfiguration scheme more frequently. On the other hand, as can be seen from Figure 40, the scenario 1 has least delay and scenario 3 has highest delay when MSs have the same mobility. It is because the local wired routing delay from a BS to IGW in the three scenarios increases when the number of BSs increases. The wired routing delay is much lower than the Internet delay because all BSs are within a local domain. The increment of the local wired routing delay is not substantial in comparison with the Internet delay.

5.7 MCIP Conclusion

The wireless communication in the near future will take advantage of dual-mode MSs, heterogeneous network accessibility, and multi-hop communications. Heterogeneous MCIP represents a new approach to support micro-mobility in wireless network. This chapter defines the general framework for Multi-hop Cellular IP and identifies some key implementation issues. Also, the chapter presents some important experimental results that show the applicability of MCIP to integrate multi-hop and single hop networks in local domains. The design of heterogeneous MCIP could be further improved and some of the open issues include: how to reduce control packet overhead and efficiently detect available BS during data transmission; and how to efficiently support various connection alternatives. We are currently looking at these problems.

CHAPTER VI

SECURED MACRO/MICRO-MOBILITY PROTOCOL FOR MULTI-HOP CELLULAR IP

A Multi-hop Cellular IP (MCIP) network differentiates global and local domains in terms of macro/micro-mobility in heterogeneous multi-hop communication. However, a MCIP network is vulnerable to various attacks and compromises during macro/micro-mobility process in an adversarial environment. Existing MCIP protocol does not provide macro/micro-mobility security protection for mobile stations. In this chapter, we introduce and evaluate a secure macro/micro-mobility protocol (SM^{3}P). In the proposed SM^{3}P, mobile IP security has been extended for supporting macro-mobility across local domains through the process of multi-hop registration and authentication. In a MCIP local domain, a certificate-based authentication achieves the effective routing and micro-mobility protection from a range of potential security threats. Our evaluation and simulation demonstrates the effectiveness of the SM^{3}P.

The proposed secure macro/micro-mobility protocol (SM^{3}P) provides MCIP network protection from various network attacks that could occur due to macro and micro mobility in heterogeneous multi-hop communication. When a MS visits a MCIP network, it needs to execute the secure global registration and authentication only once. In the MCIP local domain, the certificate-based authentication protects the multi-hop route discovery and micro-mobility of a MS from one BS to another. In the SM^{3}P, we assume that a MS and the visiting MCIP network don't have any prior security association. The key contributions of the SM^{3}P are:

- Secure single-hop and multi-hop registration and authentication in support of macro-mobility;
- Secure multi-hop paging/routing cache update for maintaining micro-mobility;
- Secure multi-hop route discovery for micro-mobility.

The rest of this chapter is organized as follows: the background of macro/micro-mobility for MCIP networks is discussed in Section 6.1. Then the security threat model for MCIP is presented in Section 6.2. After defining the security assumptions and key issues in Section 6.3, Section 6.4 describes the details of the SM^{3}P for securing macro/micro-mobility. The security achievements are discussed in Section 6.5. The implementation and simulation results are also presented in Section 6.5. Some of the related work is discussed in Section 6.6. Section 6.7 concludes the chapter.

6.1 Security Thread of MCIP

Security support is a key issue in deploying MCIP network for any application. In this section, the macro/micro-mobility security threats for MCIP network are discussed. The MCIP macro-mobility between local domains could be exposed to three types of attacks including: forged BS, unauthorized network access, registration attacks. The micro-mobility within a local domain could also suffer from various attacks including multi-hop paging/routing poisoning and multi-hop routing attacks.

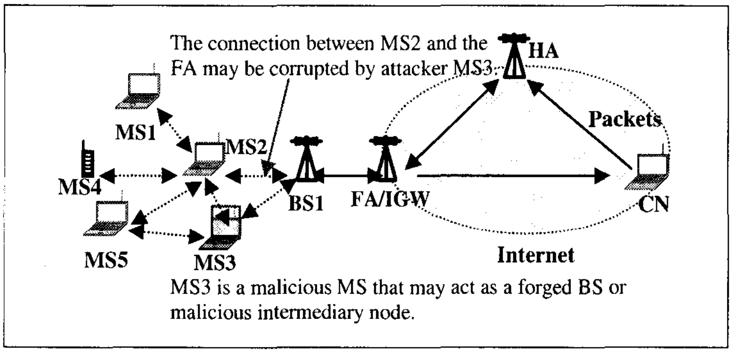
6.1.1 Macro-mobility Attacks

- **Forged BS**

In a MCIP network, an attacker can attack the multi-hop wireless network by advertising itself as a genuine BS using some forged messages or duplicate beacons recorded from a correct BS or obtained by eavesdropping. When the MS hears the fraudulent beacons from the forged BS, the MS assumes that it is within the radio coverage of a genuine BS and then initiates a registration procedure. A registration request is issued from the MS to the forged BS. Furthermore, the forged BS replies the MS with a bogus registration reply carrying the acceptance of the registration request. After receiving the registration result, the MS further assumes that it has obtained the Internet connection through the forged BS and disconnects its communication from the genuine BS. One by one, the forged BS could entice a number of MSs to disconnect with the genuine BSs and establish connections with the forged MS either by single hop or multi-hop route. However, the MSs cannot obtain any Internet service correctly from the forged BS. For instance, in Figure 41, malicious MS3 (i.e., forged BS) advertises a high-speed connectivity to the Internet by sending bogus beacons to its neighbor. After hearing the beacons from the forged BS (i.e., MS3) but without realizing the fraud, MS2 and MS5 register with MS3 by a single hop. After registrations, MS2 and MS5 believe they have connected with a BS with a higher speed, and thus disconnect the connection with the genuine BS (i.e., BS1). An attacker using forged BS achieves the following:

- The forged BS captures registration information of MSs such as their home IP addresses, home network IP addresses. It can also break a proper Internet connection and can cause unwarranted registration delay.

- The forged BS can act as the gateway to the Internet and can seize the data packet or capture sensitive personal or network data of MSs.



The connection between MS2 and the FA may be corrupted by attacker MS3.

MS3 is a malicious MS that may act as a forged BS or malicious intermediary node.

Figure 41. Macro/Micro-mobility Attacks in a MCIP Network.

- **Unauthorized Network Access**

A malicious MS may access the MCIP network and enjoy free network usage by way of single hop or multi-hop communication. After entering the Internet, the attacker may use some techniques like Medium Access Control address spoofing to gain access to the network infrastructure. For instance in Figure 42 attacker MS1 connects to the Internet and sends a malicious ARP (address resolution protocol) reply to the Router associating MS1's Medium Access Control address with MS2's IP address. At this moment, the Router thinks the MS1 is MS2. Next, MS1 sends a malicious ARP reply to MS2 associating MS1's Medium Access Control address with the Router's address. In this case, the MS2 believes MS1 is the Router. Finally, MS1 can access the session between MS2 and MS3. In this case, all data packets from MS2 will be redirected to MS1 and MS1 forwards the received data packet from MS2 to Router again. In the end, Router forwards the packets to MS3. In the opposite direction, all packets from MS3 will be forwarded to MS1 and MS1 sends the packets to MS2. MS1 implements the man-in-the-middle attack. In this manner, MS1 intercepts the traffic between MS3 and MS2.

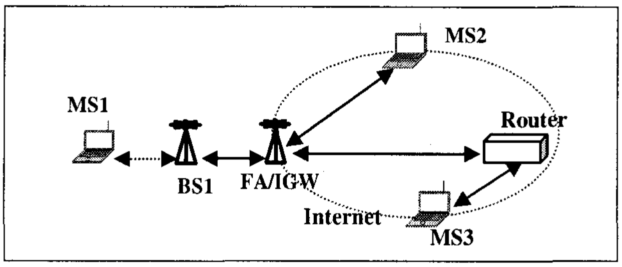


Figure 42. Unauthorized Network Access and Man-in-the-middle Attack.

- **Registration Attacks**

The registration procedure imposes several serious security threats. Three typical types of registration attacks include registration poisoning, bogus registration, and registration replay.

- **Registration Poisoning**

A malicious MS in the MCIP network can poison the registration procedure. For example, in a MCIP network, a multi-hop MS registers with the MCIP network through a multi-hop route with the help of some intermediate MSs. A malicious MS can entice a multi-hop MS to choose it as an intermediate MS by claiming a short or fast route to a BS. When a malicious MS is selected as the intermediate MS for a registration MS, the malicious MS can modify or drop the MS's registration request/reply before forwarding to the next hop. When the malicious MS modifies or rejects the registration request, the MS cannot correctly register with the foreign network. If the malicious MS modifies the registration result in a registration reply, the MS cannot access the Internet. For instance, in Figure 41, when MS5 moves to the MCIP network, it initiates a registration by sending out a registration request to MS3. When MS3 receives the registration request from MS5, MS3 modifies the address of MS5's home network in the registration request before forwarding the request to BS. Based on the modified request, the FA will forward the request to a wrong HA so that the MS cannot obtain a successful registration reply from the correct HA. Registration poisoning prevents multi-hop MSs from obtaining services from the wired network.

- **Bogus Registration**

It occurs when a malicious MS does a fake registration by masquerading itself as someone else using a spoofed or invented IP address. The bogus registration causes a wrong mobility binding at its HA so that all packets are tunneled to a illegitimate MS instead of the correct MS. By a bogus registration, the attacker obtains the right to access the Internet so that it can implement further attack on the MCIP network or the Internet. For instance, in Figure 41, MS3 does a forged registration by masquerading itself as MS5, and then all the packets coming from Internet for MS5 are forwarded to MS3. In this case, MS5 cannot receive any packet from the Internet.

- **Registration Replay**

In replay attack, an attacker records a legitimate registration packet and replays the packet for the purpose of creating a false registration. In Figure 41, MS3 may repeatedly forward a copy of registration request originated from MS5 and cause BS/FA and MS5's home network to initiate the process of registration many times. Without protection, the attacker could perform valid but unwanted operations afterwards by sending old messages.

6.1.2 Micro-mobility Attacks

- **Multi-hop Routing Attacks**

Multi-hop route discovery is responsible for detecting the multi-hop routes between MSs and BSs in the local domain. An attacker may exhibit its actions in the form of refusing to participate fully and correctly in multi-hop route discovery according to the principles of integrity, authentication, non-duplication, confidentiality, and cooperation. Therefore, multi-hop routing attacks can be grouped into five categories: anti-integrity, impersonation, duplication, and anti-confidentiality.

- Anti-integrity is the action of breaking the integrity of a message. Modification is a typical example of anti-integrity. The malicious MSs modify, inject or delete some fields of a routing packet, and then forward the packet with falsified values in the packet fields. These fields may include the source or destination address, hop count, sequence number, etc.
- Impersonations are those actions in which a malicious station spoofs an existing or forged IP address, or uses broadcast address to generate or duplicate one or more messages, and then forwards them to other MSs. In Figure 41, MS3 may masquerade itself by spoofing an invented address or an IP address of a MS. With the spoofed address, MS3 could entice MS5 to use MS3 as the shortest route for reaching BS1. And then MS3 can compromise MS5's communication.
- In a duplication attack, a malicious MS sends a legitimate message more than once. These duplicated messages cause multiple receptions and processing overhead on adjacent MSs.
- In an anti-confidentiality attack, an attacker may reveal sensitive information, such as private key. Stealing, eavesdropping, guessing, brute-force and cryptanalysis are some ways for an attacker to identify sensitive information.

- **Multi-hop Paging/Routing Cache Poisoning**

The BS in a MCIP network suffers from multi-hop paging/routing cache poisoning due to multi-hop communication. To support the micro-mobility and multi-hop communication, the routing cache is needed at each BS to record the multi-hop routes between the BS and each attached multi-hop MS. When a MS moves from one BS to another, the multi-hop paging/routing cache of the MS should be updated as illustrated in Chapter 5. The multi-hop paging/routing cache of a MS may be poisoned in several ways. For instance, when a multi-hop MS sends a multi-hop page/route-update packet for creating or updating its multi-hop paging cache, the malicious MS may modify the packet that could result in multi-hop paging/routing cache poisoning. Also, a malicious MS may send a wrong page/route-update packet on behalf of a genuine one. And the IGW and BS update the routing information for the genuine MS with the wrong information sent by the malicious MS. To locate a multi-hop MS, the BS finds the first hop MS in multi-hop paging/routing cache that can reach the destination MS. The data packet from the BS will forwarded, hop-by-hop, to the destination in accordance with the multi-hop route. When a multi-hop route is poisoned, the BS cannot locate the destination MS by following the multi-hop route provided in the BS cache. For instance in Figure 41, MS3 sends a multi-hop page/route-update packet to IGW on behalf of MS5, and IGW updates the multi-hop paging/routing cache from MS5-MS2-BS1-IGW to MS5-BS1-IGW. In this case, the packets of MS5 from the Internet will be lost due to the incorrect routing information until MS5 updates its multi-hop paging/routing cache next time.

6.2 Assumptions and Key Definitions

In a MCIP network, the infrastructure servers (i.e., HA, FA) act as the authentication and key management center for the multi-hop wireless communication, which is fundamental to MCIP mobility security. For a MS, its HA is supposed to be the most trusted server [39] [40] [41]. Its HA has the capability to manage the MS's billing and credentials when the MS has any malicious action [74][75]. Due to mobility, we assume that there is no a priori security association (SA) between the MS and the visiting FA. Otherwise, it will impose heavy burden on the key management if each MS shares a permanent key with each FA [41]. Table 7 depicts the notations used for describing SM^{3}P. In the global domain, the SM^{3}P takes advantage of existing security associations (SAs) and key management solutions:

- A MS and its HA share a secret key: ($S_{MS\text{-}HA}$) [39] [41] . This key establishes SA between the MS and its HA.
- The HA and the FA have their public-private keys (K_{HA}/K^{-1}_{HA}, K_{FA}/K^{-1}_{FA}). These keys establish SA between the HA and the FA [74][75].

In the local MCIP domain, the SM^{3}P defines the following MCIP domain keys.

- Public/Private Key. Each BS or router in a MCIP local domain (e.g., BS1, BS2 in Figure 43) has a pair of public and private keys. For instance, the BS1 in Figure 43 has its public and private keys, i.e., K_{BS1}/K^{-1}_{BS1}. The private key is kept secret by each BS or router, and the public key is wrapped up with its identification by FA with a certificate (i.e., $Cert_{BS1\text{-}FA} = <<BS1_{id}, FA_{id}, K_{BS1}, T_{issue}, T_{expire}>> K^{-1}_{FA}$). T_{issue} is the issuing time, and T_{expire} is the expiration time.
- Neighboring Shared Secret Keys: Each fixed component in a MCIP network, including BS, router and FA, shares a secret key with its neighboring fixed components. For example, in Figure 1, BS2 connects with BS1 and FA by wired cable and shares the secret keys $S_{BS1\text{-}BS2}$ and $S_{BS2\text{-}FA}$ with the two neighbors respectively. The symmetric neighboring keys are used to calculate the message

authentication code (MAC[1]) during the exchange of control packets within the local wired domain.

Table 7: Notations used for describing SM^{3}P

M, N	Concatenation of two messages *M* and *N* in the order specified
Rx, Rx'	A registering step with an issued message Rx. Rx' is the variant of Rx
A->B	A message is forwarded from *A* to *B*
MS_{HM}	MS home IP address
HA_{id}, FA_{id}, IGW_{id},	HA, FA and Gateway IP address
<<M>>K	Digital signature of the message *M* by using the private key K
<M>K	MAC value of the message *M* by using the secret key K
{M}K	Encrypt *M* by using the key *K*
$Cert_X$, $Cert_{X\text{-}Y}$	Certificate of *X*, Certificate of *X* issued by *Y*
$S_{MS\text{-}HA}$	Shared secret key between *MS* and *HA*
K_X K^{-1}_X	Public and private key of *X*
N_x	Nonce issued by X, e.g. HA, MS, FA
t	Timestamp

6.3 Secured Macro/Micro-mobility Protocol Design

The SM^{3}P has been designed to address macro/micro-mobility security including mobile IP authentication for multi-hop MS, paging/routing integrity, and multi-hop route discovery at MCIP networks. The security protocol requires that each MS provides authentication information upon entering the local domain of a MCIP network. If a MS does not have the security binding with the local domain, the MS must register with its HA and the visiting FA to get its identity verified before using the MCIP network. At the same time, during its registration, each MS also authenticates the MCIP network (i.e., FA). After the mutual authentication, each MS obtains a certificate from the FA. Each MS discards the multi-hop control packets from a MS without a valid certificate. The FA can filter data packets and discard those that don't belong to a registered MS.

[1] In this chapter, MAC refers message authentication code

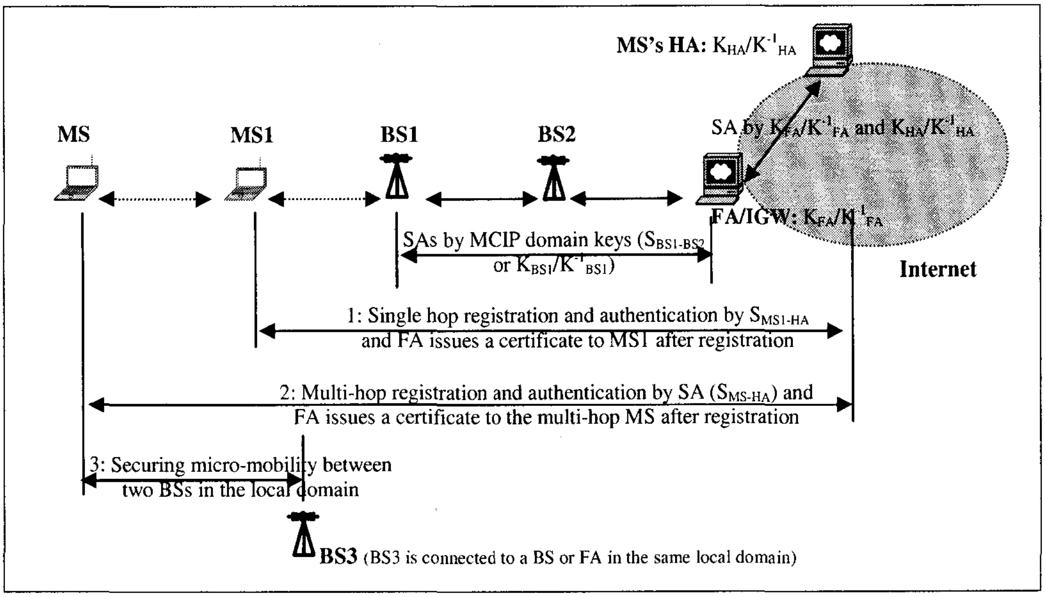

MS's HA: K_{HA}/K^{-1}_{HA}
MS
MS1
BS1
BS2
SA by K_{FA}/K^{-1}_{FA} and K_{HA}/K^{-1}_{HA}
FA/IGW: K_{FA}/K^{-1}_{FA}
Internet
SAs by MCIP domain keys ($S_{BS1-BS2}$ or K_{BS1}/K^{-1}_{BS1})
1: Single hop registration and authentication by S_{MS1-HA} and FA issues a certificate to MS1 after registration
2: Multi-hop registration and authentication by SA (S_{MS-HA}) and FA issues a certificate to the multi-hop MS after registration
3: Securing micro-mobility between two BSs in the local domain
BS3 (BS3 is connected to a BS or FA in the same local domain)

Figure 43. Security Associations (SAs) in a MCIP Network

When MCIP control packets (mobile IP or MCIP packets) is transmitted between two components in the local MCIP domain, the MCIP domain keys protects the integrity of the packet as shown in Figure 43. There are two types of registration and authentication: single hop and multi-hop, as shown in Figure 43. During registration, the single or multi-hop MS, FA, and HA authenticate one another. With a secured registration and authentication, a MS can establish a pair of temporary public and private keys (K_{MS}/K^{-1}_{MS}) during the period of its visit to the MCIP domain. The computation of public and private keys is performed by the HA when a MS registers with the HA through the FA. The computed public key of the MS is forwarded to the FA and the MS with a signature of the HA (R6 in Section 6.3.1 or 6.3.2). And the private key of the MS is encrypted and delivered to the MS secretly (in Section 6.3.1 or 6.3.2).

When a single or multi-hop MS receives a successful registration reply from the FA, the MS obtains an authenticated public key of FA (K_{FA}). The public key of the successfully registered MS is enveloped into the MS's certificate ($Cert_{MS\text{-}FA}$ = <<MS_{HM},

FA_{id}, K_{MS}, T_{issue}, T_{expire}>> K^{-1}_{FA}) (R7 in6.3.1 or 6.3.2). The certificate may read: "it is certified that the public key (K_{MS}) of the MS (MS_{HM}) is issued by FA (FA_{id}) starting from the issue time (T_{issue}) until the expiration time (T_{expire}). Each certificate has the signature of FA. Since each registered MS has an authenticated public key of FA, they can validate the certificate using the public key of FA (K_{FA}). Before the expiration of a certificate, MS requests to update its certificate if the MS is still in the MCIP local domain. The certificate can be used as the security token for micro-mobility security. For instance, if a registered MS (e.g., MS2, which is a neighboring MS of MS1 in Figure 41) receives a control packet from the neighboring MS1 such as: <<*page-update packet*>> K^{-1}_{MS1}, $Cert_{MS1\text{-}FA}$, MS2 first validates the legality of the certificate of MS1 ($Cert_{MS1\text{-}FA}$ = <<$MS1_{HM}$, FA_{id}, K_{MS1}, T_{issue}, T_{expire}>> K^{-1}_{FA}) by verifying the signature of FA with the public key of FA (K_{FA}) and checking the issuing and expiration times of the certificate. If the certificate is validated, the MS reads the public key of MS1 (K_{MS1}) from the certificate, and uses it to verify the authenticity of control packet. In the next cycle, if MS2 receives another control packet from MS1 with the same certificate, the MS2 can verify the packet directly using the public key of MS1 without validating the certificate again.

During the MS's registration, the FA creates the multi-hop paging/routing cache to record the single or multi-hop route between the MS and the IGW. Thereafter, when the MS changes its Internet attachment in the MCIP local domain (e.g., MS3 moves from BS1 to BS2), the migration is implemented by micro-mobility. The micro-mobility maintains the multi-hop paging/routing cache so that the MS and the network can reach each other when required. In the proposed SM^{3}P, the certificates of MSs protect the integrity of paging/routing-update packets and prevent unauthorized multi-hop paging/routing cache poisoning. Figure 43 illustrates the need for following security modules to be implemented for SM^{3}P protocol:

- Secure single-hop registration and authentication in support of macro-mobility;
- Secure multi-hop registration and authentication in support of macro-mobility;
- Secure micro-mobility in local domain
  - Secure searching of BS by a MS and secure route creation
  - Secure multi-hop paging/routing cache update for micro-mobility

6.3.1 Secure Single-hop Registration

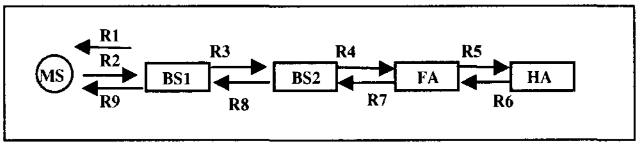


Figure 44. Registration for a Single-hop MS.

The process of the single-hop registration implements the macro-mobility for single hop MSs. The registration for a single-hop MS, which is under the radio coverage of a BS, creates a secured macro-mobility binding at the MS's HA. At the same time, during the process of secured registration, the FA issues a certificate for the single hop MS that is valid in the MCIP local domain. A BS in the MCIP local domain advertises its presence periodically to MSs with beacons. The FA on behalf of each BS creates the beacons and each beacon has a certificate of the FA. If a single-hop MS has not yet registered with the MCIP network, it follows the steps of registration and authentication outlined in Figure **44**. For instance, MS1 in Figure **44** uses this procedure for registration.

Beacon:

(R1) BS->MSs: <<M_1>> K^{-1}_{FA}, $Cert_{FA}$

Where M_1 is beacon, FA_{id}, IGW_{id}, BS_{id}, Seq;

Where FA_{id} is the FA address; IGW_{id} is the CoA address; BS_{id} is the BS address; Seq is the sequence number

Registration for a single-hop MS:

(R2) MS -> BS1: < M_3> S_{MS-HA};
Where M_3=Registration Request, FA_{id}, IGW_{id}, HA_{id}, MS_{HM}, N_{MS}, N_{HA}, R1

(R3) BS1 -> BS2: < R2, $BS1_{id}$>$S_{BS1-BS2}$

BS2 validates R3 using $S_{BS1-BS2}$
(R4) BS2 -> FA: < R2, $BS1_{id}$,$BS2_{id}$>S_{BS2-FA}

FA validates R4 using S_{BS2-FA}
FA records the route for the MS
(R5) FA->HA: R2, N_{FA}

HA (upon receipt of R5):
Validates R2 using S_{MS-HA}
Checks whether FA_{id} in M_1= FA_{id} in M_3
Validates $Cert_{FA}$ based on existing PKI at HA
Validates <<M_1>> $K^{-1}{}_{FA}$ using the authenticated K_{FA}
Computes a pair of public and private keys (K_{MS}, $K^{-1}{}_{MS}$)
Continues with the steps in [33] [34](Perkins, Mobile IP support), i.e., updating the mobility binding
(R6) HA->FA: M_4, << M_4>> $K^{1}{}_{HA}$, $Cert_{HA}$;
Where M_4 = M_5, M_6, N_{FA};
Where M_5 = Reply, Result, FA_{id}, HA_{id}, BS_{id}, MS_{HM}, K_{FA}, K_{MS} N'_{HA}, N_{MS}
Where M_6 = < M_5, {$K^{-1}{}_{MS}$} S_{MS-HA} > S_{MS-HA}

FA (upon receipt of R6):
Validates N_{FA}
Validates $Cert_{HA}$ based on existing PKI at FA
Validates << M_4>> $K^{-1}{}_{HA}$ using the authenticated K_{HA}
Creates multi-hop paging / routing cache at IGW for the MS
Issues a certificate for the MS: $Cert_{MS-FA}$ = <<MS_{HM}, FA_{id}, K_{MS}, T_{issue}, T_{expire}>> $K^{-1}{}_{FA}$,
Logs this message as the history trace of the MS,
Continues with the steps in [33] [34] (Perkins, Mobile IP support), i.e., creating mobility binding
(R7) FA->BS2: <M_6, $Cert_{MS-FA}$> S_{BS2-FA}

BS2:
Validates R7 by using S_{BS2-FA}
Creates multi-hop paging / routing cache for the MS
(R8) BS2->BS1: <M_6, $Cert_{MS-FA}$> $S_{BS1-BS2}$

BS1:
Validates R8 by using $S_{BS1-BS2}$
Creates multi-hop paging / routing cache for the MS
(R9) BS1->MS: M_6, $Cert_{MS-FA}$

MS (upon receipt of R9):
Validates < M_6 > S_{MS-HA} using S_{MS-HA}
Obtains $K_{FA}$from M_5
Validates $Cert_{MS-FA}$ using K_{FA}
Obtains private key ($K^{-1}{}_{MS}$) by decrypting {$K^{-1}{}_{MS}$} S_{MS-HA} using S_{MS-HA}
Continues with the steps in [33] [34] (Perkins, Mobile IP support), i.e., recording the registration result

On receiving a beacon from a BS, the MS1 compares the FA address, the CoA address, the BS address, and the sequence number with those of previously received beacons. The sequence number is incremented every time a new beacon is issued by the BS. The MS discards the duplicate beacons. If it is a fresh beacon, the MS records the beacon. An unregistered MS cannot validate the authenticity of the beacon. This task is left to registration and authentication process for unregistered MSs. First the MS creates a registration request (R2) and then sends it to a BS (e.g., BS1 in Figure 45) directly. The MAC, which is created by using the secret key ($S_{MS\text{-}HA}$), maintains the integrity of the registration request. Then the registration request (R2) is forwarded to the FA hop by hop with the relays of BSs (R3 and R4). Each intermediate BS appends its address so that the FA can construct the reverse route between the IGW and the MS. The FA then forwards the registration request to the MS's HA after appending the FA's nonce (R5). While receiving the registration request, the HA checks the integrity of the registration request by using $S_{MS\text{-}HA}$. Then FA is verified for preventing any camouflaged FA. After the series of verifications, the HA computes a pair of public and private keys ($K_{MS,}$ $K^{1}{}_{MS}$) for securing micro-mobility of the MS in the MCIP local domain. The private key is sent to the MS secretly with encryption (R6).

Upon receiving the registration reply from the HA, the FA validates the HA using the HA's public key. After the authentication of the HA and the registration reply (R6) from HA, the FA records the MS's public key. The FA now creates a multi-hop paging cache for the MS. The FA returns a registration reply (R7) back to the MS. When the MS receives the registration reply from the FA via the reversed route, the MS validates the registration reply message using $S_{MS\text{-}HA}$. The MS is then sure that the HA is valid and the public key of FA (K_{FA}) is authenticated because of the authentication of the registration reply message. In the end, the MS obtains the registration result, its private key computed by its HA and certificate issued by FA. Once the MS receives a successful reply from its HA, it is guaranteed that FA and keys are valid. The beacon is verified at its HA.

6.3.2 Secure Multi-hop Registration

The process of a multi-hop registration implements the macro-mobility for a multi-hop hop MS. A multi-hop MS, which is outside the transmission range of all BS in the MCIP local domain, obtains beacons with the help of other registered MSs. The beacon forwarded by a registered MS carries the original beacon issued by a BS and, the home address of the registered MS as well as its signature. For instance, MS obtains beacon from MS1 in Figure 45, in which MS1 sends MS with (R1'), which is << R1, MS1 >> K^{-1}_{MS1}, $Cert_{MS1-FA}$. However, as for an unregistered MS, it cannot verify the FA and its neighbors because the MS does not have the authenticated public key of the FA and the authenticated public key of the neighboring MS. The task is left to the process of the secure multi-hop registration of the MS. The multi-hop registration can be used to create a secured macro-mobility binding at the HA for the multi-hop MS. The steps of securing multi-hop registration are illustrated in Figure 45.

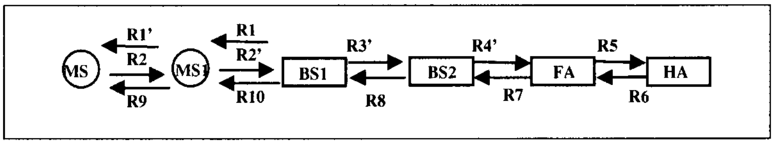


Figure 45. Registration for a Multi-hop MS.

Beacon Relaying:

(R1) BS->MSs: <<M_1>> K^{-1}_{FA}, $Cert_{FA}$;

Where M_1 is beacon, FA_{id}, IGW_{id}, BS_{id}, Seq;

Where FA_{id} is the FA address; IGW_{id} is the CoA address; BS_{id} is the BS address; Seq is the sequence number

(R1') MS1->MS: << M_2 >> K^{-1}_{MS1}, $Cert_{MS1-FA}$;

Where M_2 is R1, MS1

Registration for a Multi-hop MS:

(R2) *MS -> MS1: <* M_3*>* S_{MS-HA}, *R1'*
Where M_3*=Registration Request,* FA_{id}, IGW_{id}, HA_{id}, MS_{HM}, N_{MS}, N_{HA}, *R1*

(R2') *MS1 -> BS1: <<R2,* $MS1_{HM}$ *>>* K^{-1}_{MS1},$Cert_{MS1-FA}$

BS1 (upon receipt of R2'):
Validates R2' using K_{MS1} in $Cert_{MS1\text{-}FA}$
Validates R1' using K_{MS1} in $Cert_{MS1\text{-}FA}$
(R3') BS1 -> BS2: $<R2, MS1_{HM}, BS1_{id}>S_{BS1\text{-}BS2}$

BS2 validates R3' using $S_{BS1\text{-}BS2}$
(R4') BS2 -> FA: $< R2, MS1_{HM}, BS1_{id}, BS2_{id}>S_{BS2\text{-}FA}$

FA validates R4' using $S_{BS2\text{-}FA}$
FA records the multi-hop route for the MS
(R5) FA->HA: R2, N_{FA}

HA (upon receipt of R5):
Validates R2 using $S_{MS\text{-}HA}$
Checks whether FA_{id} in $M_1 = FA_{id}$ in M_3
Validates $Cert_{FA}$ based on existing PKI at HA
Validates $<<M_1>> K^{-1}_{FA}$ using authenticated K_{FA}
Computes a pair of public and private keys (K_{MS}, K^{-1}_{MS})
Continues with the steps in [33] [34] (Perkins, Mobile IP support), i.e., updating the mobility binding
(R6) HA->FA: $M_4, << M_4 >> K^{-1}_{HA}, Cert_{HA};$
Where $M_4 = M_5, M_6, N_{FA};$
Where M_5 = Reply, Result, $FA_{id}, HA_{id}, BS_{id}, MS_{HM}, K_{FA}, K_{MS} N'_{HA}, N_{MS}$
Where $M_6 = < M_5, \{K^{-1}_{MS}\} S_{MS\text{-}HA} > S_{MS\text{-}HA}$

FA (upon receipt of R6):
Validates N_{FA}
Validates $Cert_{HA}$ based on existing PKI at FA
Validates $<< M_4 >> K^{-1}_{HA}$ using authenticated K_{HA}
Creates multi-hop paging / routing cache at the IGW for the MS
Issues a certificate for the MS: $Cert_{MS\text{-}FA} = <<MS_{HM}, FA_{id}, K_{MS}, T_{issue}, T_{expire}>> K^{-1}_{FA}$
Logs this message as the history trace of the MS
Continues with the steps in [33] [34] (Perkins, Mobile IP support), i.e., creating the mobility binding
(R7) FA->BS2: $<M_6, Cert_{MS\text{-}FA}> S_{BS2\text{-}FA}$

BS2:
Validates R7 by using $S_{BS2\text{-}FA}$
Creates multi-hop paging / routing cache at BS2 for the MS
(R8) BS2->BS1: $< M_6, Cert_{MS\text{-}FA} > S_{BS1\text{-}BS2}$

BS1:
Validates R8 by using $S_{BS1\text{-}BS2}$
Creates multi-hop paging / routing cache at BS1 for the MS
(R10) BS1->MS1: $<< M_6, Cert_{MS\text{-}FA} >> K^{-1}_{BS1}, Cert_{BS1}$

Validates R10 by using K_{BS1}
(R9) MS1->MS: $M_6, Cert_{MS\text{-}FA}$

MS (upon receipt of R9):
Validates $< M_6 > S_{MS\text{-}HA}$ using $S_{MS\text{-}HA}$
Obtains K_{FA} from M_5
Validates $Cert_{MS\text{-}FA}$ using K_{FA}
Obtains private key (K^{-1}_{MS}) by decrypting $\{K^{-1}_{MS}\} S_{MS\text{-}HA}$ using $S_{MS\text{-}HA}$
Continues with the steps in [33] [34] (Perkins, Mobile IP support), i.e., recording the registration result

As seen from Figure 45, the basic operations in multi-hop registration are similar to single-hop registration. The FA and HA has the same steps either in single hop or multi-hop registration (R5 and R6). Also, upon receiving the registration reply (R9), the multi-hop MS has the same operations as single hop registration such as verifying the registration reply and obtaining its private key. However, the multi-hop registration has three differences compared to single hop registration:

i. The registration request of the multi-hop MS is forwarded to a BS (BS1 as shown in Figure 45) by multi-hop. Each intermediate MS has a signature and its next hop MS validates the signature before signing and forwarding until the request is received by a BS (R2-R2'),

ii. Before forwarding the registration request to FA (R3'), the BS validates the intermediate MSs and the beacon. If there is any error during the process of verifying the beacon or intermediate relaying MS, the BS denies the registration request, and

iii. The registration reply message is returned back to the multi-hop MS by the secured multi-hop forwarding (R10) with signature and verification.

Each MS only registers with MCIP network once either by single hop or multi-hop. After registration, the migration from one BS to another and the change of multi-hop route are implemented by the process of micro-mobility as explained in the next section.

6.3.3 Secure Micro-mobility

As a MS moves in local domain, it may experience low data rate at current connection or lose the connectivity to the current BS. In such cases, the MS initiates a route discovery to find a BS with better connectivity by issuing a search-request packet with its signature and the attachment of its certificate. The search-request packet is sent out on an available radio interface (e.g., cellular or IEEE 802.11). The certificate of the MS acts as the security token during the forwarding of the route discovery packet. Upon the reception of a routing packet, the intermediate MS verifies the MS from which the searching-request packet is received, and then forwards the route discovery packet again after appending its address, its signature, and its certificate. When a BS receives the searching-request packet, it verifies the intermediate MSs. Then it sends a search-reply packet with its signature. The search-reply packet will be forwarded to the source MS with the reversed route without modification. The MS initiated searching process for locating a BS is depicted in Figure 46. Before moving, the MS has its multi-hop routing cache (Path: MS-MS3-MS2-BS3-BS4-IGW) through BS3 with its cellular radio interface (address: IP_C). When MS moves to the new location, MS detects a new BS (R11-R14) with its IEEE 802.11 radio interface (address: IP_W). In R11, MS_{HM} is the address IP_C. In the end, BS1 replies with a route of (MS-MS1-BS1). When MS receives the search–reply packet (R14) from BS1, the MS initiates the process of micro-mobility from BS3 to BS1 as illustrated in the next subsection.

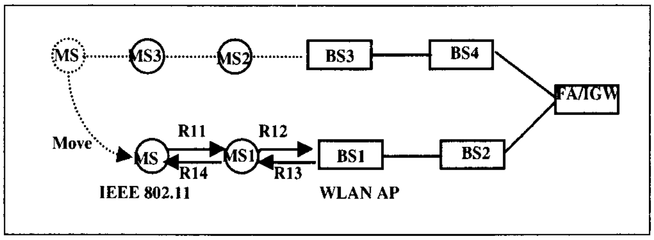


Figure 46. Secure MS Searching Process for Locating BS.

Secure MS Searching Process:

(R11) *MS broadcasts:* $<< M_7 >> K^{-1}_{MS}, Cert_{MS}$
Where M_7= *searching-request packet,* $FA_{id}, IGW_{id}, MS_{HM}, IP_W, t, N_{MS}$

MS1 validates R11 using K_{MS}
(R12) *MS1 broadcasts:* $<<R11, MS1_{HM}>> K^{-1}_{MS1}, Cert_{MS1}$

BS1 validates R12 using K_{MS1}
(R13) *BS1->MS1:* $<< M_8 >> K^{-1}_{BS1}, Cert_{BS1}$
Where M_8= *searching-reply packet,* $FA_{id}, IGW_{id}, MS_{HM}, IP_W, MS1_{HM}, BS1_{id}, t, N_{MS}$

MS1 validates R13 using K_{BS1}
(R14) *MS1->MS:* $<< R13 >> K^{-1}_{MS1}, Cert_{MS1}$

In a MCIP local domain, a MS implements the micro-mobility by updating its multi-hop paging/routing cache, from which a MS can change its network attachment from one

BS to another in a heterogeneous multi-hop environment. Figure 47 shows the steps to update a multi-hop route, which prevents unauthorized multi-hop cache update. The private key of the source MS protects the integrity of the page/route-update packet. In Figure 47, while communicating with its cellular interface (address: IP_C), MS moves to a new location and detects a new BS with its IEEE 802.11 interface (address: IP_W). Due to higher performance of IEEE 802.11, the MS initiates a micro-mobility switch from BS3 to BS1 with a vertical handoff from its cellular radio interface with IP_C to its IEEE 802.11 radio interface with IP_W. The MS updates its multi-hop paging/routing cache through BS1, which is a WLAN AP.

- The MS issues a page/route-update packet (e.g., R15 as shown in Figure 47) with the destination of the IGW carrying its migrating multi-hop route (IP_W, $MS1_{HM}$, $BS1_{id}$).
- The paging/routing update packet is forwarded to the FA via BS1 hop by hop. Each intermediate MS and BS validates the authenticity of the packet.
- The FA authenticates the MS by using the MS's certificates. After authentication, the FA finally updates paging/routing cache and replies with a paging/route-reply packet (R16) to the source MS. The FA creates a mapping between the two IP addresses (IP_C and IP_W) so that data packets from the Internet with IP_C can be redirected to IP_W [76].
- After updating the multi-hop paging/routing cache, the data packets of the MS will be delivered according to the multi-hop route in the updated multi-hop paging/routing cache.

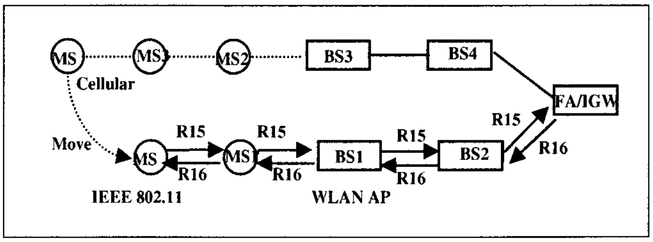


Figure 47. Securing Micro-mobility.

Secure Multi-hop Paging/Routing Update:

(R15) $MS\text{->}FA: << M_9 >> K^{-1}_{MS}, Cert_{MS}$
Where M_9 *= page/route-update packet,* $FA_{id}, IGW_{id}, MS_{HM}, IP_W, MS1_{HM}, BS1_{id}, N_{MS}, N_{FA}$

FA validates $<< M_9 >> K^{-1}_{MS}$ *using* K_{MS}
FA updates the multi-hop paging/routing cache for the MS and creates a mapping from MS_{HM} (IP_C) *to* IP_W
(R16) $FA\text{->}MS: << M_{10} >> K^{-1}_{FA}$
Where M_{10} *= paging/route-reply packet,* $FA_{id}, IGW_{id}, MS_{HM}, IP_W, N_{MS}, N'_{FA}$

6.4 Security Analysis

This section evaluates the security achievements and performance of the proposed SM^{3}P. The computation overhead of the proposed protocol depends on many factors, for instance, the algorithms of key creation, signing and verifying as well as the key length and security level. Currently, there are a number of asymmetric key cryptosystems in the literature, e.g., RSA, ElGamal and Elliptic Curve DSA [77]. An implementation of the key creation, signature and verification based on the elliptic curve cryptography is implemented for evaluating the performance of the proposed approach.

The proposed SM^{3}P achieves the goals of preventing the attacks in macro/micro-mobility of:

1. Forged BS
2. Unauthorized network access
3. Registration attacks (registration poisoning, bogus registration, and registration reply attack)
4. Multi-hop paging/routing cache poisoning

5. Multi-hop routing attacks (anti-integrity, impersonations, and anti-confidentiality)

This section describes the above security achievements in detail.

Forged BS: After a forged BS broadcasts a fraudulent beacon (a forged R1 = $<<M_1>>$ K^{-1}_{FA} $Cert_{FA}$) to declare itself as a genuine BS, the beacon could be received by either a registered MS or an unregistered MS. The registered MS can identify that the beacon is incorrect by validating the signature of the beacon since the genuine BS has an authenticated public key of FA (K^{-1}_{FA}), which is obtained during its registration and authentication (*Validates* $< M_6 >$ $S_{MS\text{-}HA}$ *using* $S_{MS\text{-}HA}$, *Obtains* K_{FA} *from* M_5 in Section 6.3.1 or 6.3.2).

An unregistered MS cannot verify the forged beacon. However, if the MS uses the forged beacon to register with the forged BS, the forged BS cannot provide a correct registration reply (R6 or R7) due to lack of $S_{MS\text{-}HA}$. Therefore, the unregistered MS knows that the registration reply from the forged BS is incorrect at the step: *Validates* $< M_6 >$ $S_{MS\text{-}HA}$ *using* $S_{MS\text{-}HA}$ in Section 6.3.1 or 6.3.2. If the forged BS forwards the registration to a correct FA, the correct FA will reject the beacon because the forged BS does not have a legal public key. If the forged BS forwards the registration to the MS'HA, the HA verifies the beacon in the step: "*Validates* $Cert_{HA}$ *based on existing PKI at FA*" in Sections 6.3.1 and 6.3.2 and will discard it. If the forged BS replays an outdated beacon, it can be checked out by the registered MS, BS, or FA due to the non-duplication protection at each beacon (timestamp). When a malicious intermediate MS creates a fraudulent beacon (R1': M_2= << forged-R1, forged-MS>> K^{-1}_{MS}, forged-$Cert_{MS\text{-}FA}$), the registered MS will reject the beacon using the authenticated public key of FA. As for an unregistered MS, the fraudulent beacon will be identified in the step: *BS1 verifies the* $Cert_{MS\text{-}FA}$ *by* K_{FA} in Section 6.3.2.

Unauthorized network access: An unregistered MS cannot access the Internet and MCIP network due to lack of proper certificate ($Cert_{MS1\text{-}FA} = \ll MS1_{HM}, FA_{id}, K_{MS1}, T_{issue}, T_{expire} \gg K^{-1}_{FA}$) that is issued by FA (R7 in Section 6.3.1 and 6.3.2). Registered MSs will discard all packets issued or forwarded by the unregistered MS.

Registration attacks (registration poisoning, bogus registration, and registration replay):

(A) Registration poisoning: The secret key ($S_{MS\text{-}HA}$) between a MS and its HA maintains the integrity of registration against poisoning during the forwarding of registration packets (registration request or registration reply). The MAC value at R2 (M_3) and R6 (M_6) prevents any modification.

(B) Bogus registration: When a malicious MS does a fake registration by masquerading itself as someone else with an invented or spoofed address, the malicious MS issues a forged registration request (forged R2). The forged registration will be identified by the HA at the step: *Validates R2 using* $S_{MS\text{-}HA}$ in Section 6.3.1 or 6.3.2. This is because the malicious MS has no knowledge of secret key ($S_{MS\text{-}HA}$) which is associated with the invented or spoofed address.

(C) Registration replay: Nonce (e.g., N_{MS}, N_{HA}, *and* N_{FA}) is used in all registration and control packets to ensure that these packets contain a unique identification to prevent replay attack. Each registration or control packet issued by an originator has a nonce, and a new nonce in the reply packet indicates the next nonce for the next request.

Multi-hop paging/routing cache poisoning: The multi-hop paging/routing cache for a MS is created after its secured registration is carried out as discussed in Section 6.3.1 or 6.3.2 (*Creates multi-hop paging / routing cache at IGW for the MS*). Then, the paging/routing cache is updated as discussed in Section 6.3.3 (R15-R16). The page/route-update packets are issued with a signature by the source MS. The signature of the source MS protects the page/route-update packet from modification during the forwarding from the source MS to the FA. The FA authenticates the source MS and the paging/routing packet before updating the multi-hop paging/routing page (*FA updates the multi-hop paging/routing cache for the MS*) as discussed in Section 6.3.3.

Multi-hop routing attacks (anti-integrity, impersonation, anti-confidentiality, and duplication):

(A) Anti-integrity: The search-request packet is signed by using the private key of each sender. The receiver verifies the certificate and the signature of the sender (R11- R14 in Section 6.3.3). Each registered MS keeps its private key secret. Therefore, the signature and verification prevent anti-integrity attacks. The corresponding search-reply packet is signed by the destination, which maintains the integrity of the search-reply packet during the forwarding from the destination to the source MS.

(B) Impersonation: The certificate, which is issued and signed by FA, prevents impersonation during multi-hop forwarding through signing and verifying (R11-R14 in

Section 6.3.3). In the certificate, the MS's address and its public key are bound together. The binding is unique because of the uniqueness of the MS's IP address. The certificate can not be forged because of the signature of FA in each certificate. All registered MSs, which have the authenticated public key of the FA, can verify the certificate. Therefore, it is difficult for any MS to masquerade itself by spoofing or inventing an address in multi-hop route discovery.

(C) Anti-confidentiality: Each MS has a pair of public and private keys after registration and keeps its private key secret. Sensitive data can be transmitted after encrypting with the receiver's public key at the sender. Only the receiver can decrypt the messages by using its private key.

(D) Duplication: A nonce and timestamp in each routing packet prevent duplication.

6.4.2 Performance Analysis

6.4.2.1 Experimental Configuration

The experiments were conducted in the NS-2 simulator [73] to evaluate the effectiveness of the proposed SM^{3}P. The configuration of an experimental MCIP network is shown in Figure 48. A router in the global domain represents the Internet with certain delay (0.01 seconds). The functionalities of BS1, the FA, and the IGW are integrated as a node. A HA as well as an IGW/FA/BS1 is connected to the router. In the local domain, seven BSs are connected by wired cable. The dimension of the local domain is 1770m x 1770m with 133 MSs, which are randomly located in the domain. All the functionalities of above components are implemented under a Pentium IV 2.8 G with 516 MS of memory.

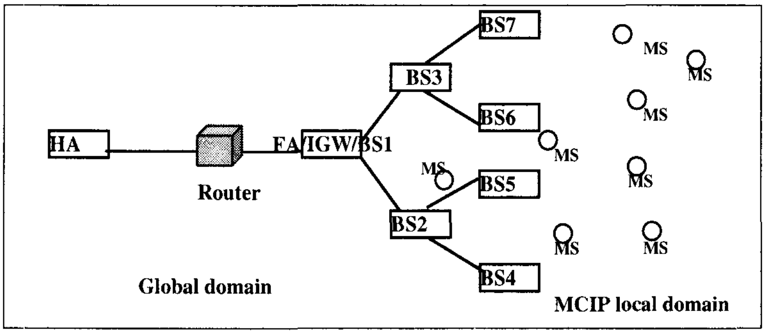


Figure 48. Experimental Configuration.

To simplify our experimental configuration, the experiments ignore the authentication delay at PKI authorities (AAA) [74]. As for the asymmetric operations, the experiments based on elliptic curve cryptosystem (ECC) [77] are carried out to evaluate the computation overhead for key creation, signature, and verification. The elliptic curves are usually defined over binary fields F_{2m} ($M \geq 1$) or prime fields F_p, p> 3. The experiments here implement the elliptic curves defined over prime fields. The implementation of ECC is based on the software ECC library in [77] because of its portability and ease of use. The HA and FA can generate system parameters, a prime **P**, elliptic curve **E**, base point **G** = (x,y), and order **r** of the point **G** so that they can compute a pair of public and private keys as required. The prime **P** is generated based on security level. Since the size of **P** is equal to 160 bits long, the ECC has the same level of security with RSA and DSA with a 1024-bit modulus [77]. The experiments choose **P** as 163 and 175 bits long. In the following, let **P** = x donate the size of prime **P**. The HA and FA generate an elliptic curve having suitable order **r** by using the complex multiplication method with a discriminant **D** [77]. The value of **D** is 40 in the experiments. In the stage of initiation, the FA and HA and local domain BSs are configured with public and private keys according to the proposed SM^{3}P. Each data point shown in the following figures is averaged over five runs with different MS distribution. The number of hops in multi-hop communication is the key factor in increasing the transmission delay. Thus, the experiments are conducted based on different number of hops from a MS to a BS (a single-hop or multi-hop MS).

6.4.2.2 Impact of Security on Macro and Micro mobility

The macro/micro-mobility delay is primarily impacted by communication and computation delay. The communication delay of a packet includes all the latencies of multi-hop wireless forwarding, MCIP local wired network, and the Internet. The local network has less delay for a packet transmission than that of the Internet. In the SM^{3}P, the computational delay is primarily caused by the cryptographic operations: symmetric or asymmetric. The symmetric cryptographic operation is very efficient in generating and verifying authenticator. In general, it is one-thousandth of the asymmetric cryptographic operation [41]. For instance, an AMD Opteron 1.6GHz processor under Linux 2.4.21 can perform a stream cipher encryption/decryption using Panama Cipher (little endian) algorithm at 344.781 Megabytes/second [78]. It can perform a MAC computation (and verification) with HMAC/MD5 algorithm at 152.381 Megabytes/second [78]. Based on these observations, symmetric cryptographic operations are not considered for mobility delay in the simulations.

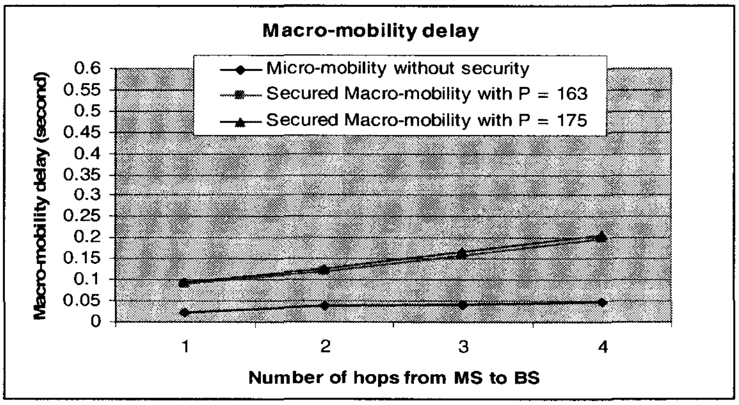


Figure 49. Macro-mobility delay for single hop and multi-hop MSs.

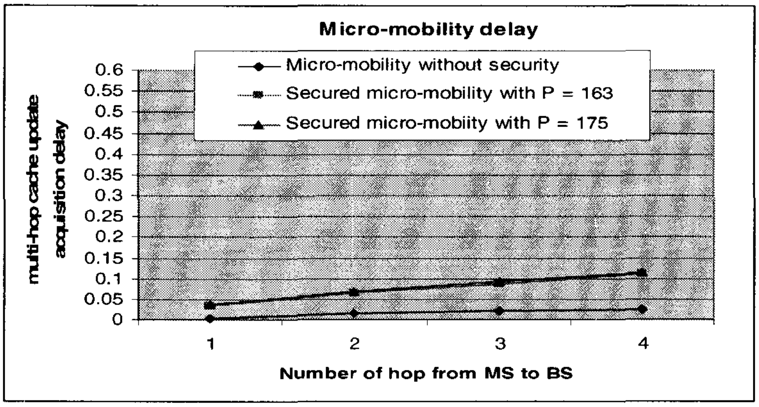


Figure 50. Micro-mobility delay for single hop and multi-hop MSs.

The macro-mobility delay is the interval from the issue of registration request to the completion of all registration and authentication operations. As shown in Figure 49, for secured or unsecured macro-mobility, registration delays increase when the number of hops from MS to network attachment (BS) increases. It is because the increase of hops results in more wireless transmission and computation delay at the intermediate MSs. Security never comes for free. The registration delays caused by security operations are higher than that of registrations without security because the security feature involves computation overhead during registration. The macro-mobility delay also increases when the size of the prime field increases because the timings for ECC operations increase for a larger prime field (175 >163 bits long).

Micro-mobility delay for a MS is the interval from the initiation of paging/routing-update packet to the verification of paging/routing-reply packet. During this period, the MS updates its multi-hop paging/routing cache in a safe manner and migrates from one BS to another. As seen from Figure 49 and Figure 50, the micro-mobility delay is less than that of macro-mobility when the MS has the same number of hops to the BS. The secured micro-mobility delay increases linearly and does not show significant additional overhead of security. It is because the process of a macro-mobility not only has the communication and computation delay at multi-hop wireless network and local network, but also has the delay in the Internet and the significant delay in the HA and FA authentications. When a registered MS migrates from a BS to another in the MCIP local domain, there is no need for the MS to registers with its HA again. Thus, the MCIP micro-mobility protocol significantly reduces the delay.

6.4.2.3 Impact of Security on Route Acquisition Delay

Figure 51 shows the delay of MS search process for locating the neighboring BS due to micro-mobility based on different hops from a source MS to BS. The delay is the interval from the issue of searching-request packet to the moment of finishing the verification of the searching-reply packet from a BS. It is clear from Figure 51, the security authentication results in extra route discovery delay because of the computation overhead at each intermediate MS for signature and verification. For single hop MSs, it learns the BS information by listening to the beacon from the BS and checks the authenticity of beacon by using the public key of FA. This is why the route delay is much smaller when hop count is one.

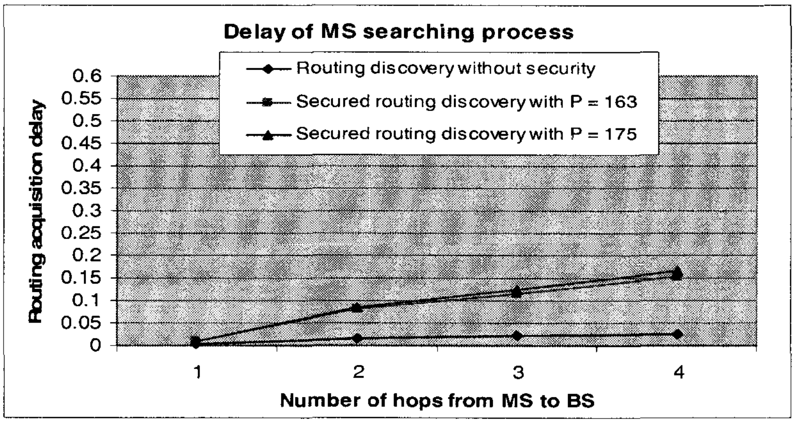


Figure 51. Routing Acquisition Delay for BS Detection.

After the process of route discovery, MS obtains the route to the new BS, but it can not use the route for internet connection immediately because of the stale mobility binding. Otherwise, the packets from the Internet will be directed to the previous associated BS rather than the new BS. As shown in Figure 50, the higher delay occurs if the MS updates its new network attachment (new BS) with a macro-mobility process. In the proposed SM^{3}P, the migration using the process of micro-mobility has lower delay compared to the scheme using macro-mobility process as shown in Figure 51. Meanwhile, some schemes such as packet buffering can be used to reduce the packet drop during the period of link breakage and migration.

In the proposed SM^{3}P, instead of using secret key-based (symmetric) primitive, the certificate-based scheme (asymmetric) achieves the scalability. There may be a large number of MSs in a MCIP local domain but only a small percent of MSs are active sending or receiving packets. In the case of symmetric cryptography, each MS should have a secret key with every other MS, which makes it unappealing for key establishment and management. In our protocol, each MS only need to maintain its private key, its certificate, and the public key of FA. By using the public key of the FA, the MS can verify all the certificates from other MSs, and further validate the authenticity of the MSs and their packets.

6.4.1 Conclusion

In this chapter, we have presented the SM^{3}P to secure macro/micro-mobility for MCIP network. The SM^{3}P extends the mobile IP security for registration and authentication for single hop or multi-hop MS. During the process of registration, the SM^{3}P prevents forged BS and registration attacks. After a mobile IP registration, each

MS obtains a certificate for providing effective micro-mobility protection when the MS moves from a BS to another in the MCIP local domain without re-registration and re-authentication at the migrating BS. During the process of micro-mobility, the SM^{3}P provides the multi-hop routing prevention in accordance with anti-integrity, impersonation, anti-confidentiality and duplication. The performance of the proposed SM^{3}P approach shows that the macro-mobility delay is higher than micro-mobility delay.

CHAPTER VII

CONCLUSION AND RECOMMENDATION

The secure Internet connectivity and mobility management are two crucial issues in designing an integrated internet and heterogeneous multi-hop network. This chapter summaries our contributions in this dissertation and discusses and future works for secured communication for heterogeneous multi-hop networks. Section 7.1 illustrates the conclusion drawn from our works, and Section 7.2 discusses the future research.

7.1Conclusion

7.1.1 Secure Communication for Integrated Internet and MANET

In an integrated internet and heterogeneous multi-hop network, a multi-hop route discovery protocol is required to support multi-hop wireless communication. At the same time, as discussed in Chapter 3, the route discovery protocol should cooperate with mobile IP protocol to provide the global internet connectivity. A multi-hop MS located outside the coverage of the BS can effectively detect the availability of a BS by a process of a route discovery. The route discovery protocol further enables the MS with capability to reconfigure a new route to the Internet when current path is broken due to intermediary node mobility. It can be seen from our analysis and experimental results in Chapter 3, the effectiveness of such integration protocol should be:

- Providing efficient bi-directional connection,

- Lower overhead and higher throughput,
- Less connection delay for communication establishment and packet forwarding.

In an integrated network, an adversary may modify the routing packet from a neighboring MS. As discussed in Chapter 2, many attacks may happen in an integrated network. The security protocol proposed in Chapter 4 integrates mobile IP security and ad hoc security for providing security protection for integrated Internet and MANET. This security protocol requires that each MS provides authentication information upon entering the multi-hop network. If a MS does not have the security binding with the network, the MS must register with the HA and the FA to get its identity verified before using the multi-hop heterogeneous network. This prevents unauthorized access to multi-hop network. At the same time, during registration, each ad hoc host also authenticates the FA through the HA. In the proposed model, each MS only chooses the authenticated MSs as ad hoc relaying stations. The FA can filter data packets and discard those that do not belong to an authenticated MS. The security protocol achieves the following goals.

- Multi-hop heterogeneous network provides services only to authenticated MSs.
- Only authenticated MSs must be used in multi-hop routes. The route creation process must only involve trusted and authenticated MSs.
- Only authenticated routing information should be used in the updates of paging/routing caches.

The first goal is achieved by requiring the registration of all MSs. During registration, the MS, the FA, and the HA are authenticated on the whole. Our proposed mechanisms to establish distributed trust relationship for a multi-hop MS have three steps (i) the multi-hop MS sends its credentials to the network (BS) so that the BS can authenticate the MS,

(ii) the multi-hop MS validates the BS so that the MS can trust the BS for accessing the network, and (iii) after the mutual authentication, the MS can create a security binding with the integrated network. The second goal is achieved by enforcing signature and verification along a multi-hop route during the construction of the route. The signature verification prevents the attacks of the anti-integrity and impersonation. The third goal is met by authentication of a MS before BS updates a paging/routing cache.

7.1.2 Secure Mobility Management in Heterogeneous Multi-hop Wireless Networks

Since the standard mobile IP protocol suffers from limitations (e.g., latency, frequent handoff) as illustrated in Chapter 5, the proposed MCIP protocol provides multi-hop communication with micro-mobility support by integrating multi-hop communication with Cellular IP. The MCIP divides a multi-hop wireless network into separate domains and each domain has an IGW for providing internet accessibility. The BSs are connected with each other by wired or wireless links. Because of local domain, a MS doesn't need to register with its HA every time when it moves from a BS to another in a local domain. In order to support micro-mobility, the MCIP includes three components: location management, connection management and route reconfiguration. Location management is responsible for maintaining location information for idle MSs in a local domain. Connection management constructs an initial path for data transmission if a MS is moving to active state for sending or receiving packets. In the active state, a route reconfiguration mechanism is proposed to take advantage of various multi-hop connection alternatives available based on terminal interfaces, network accessibility and topology.

In our secure macro/micro-mobility protocol, each BS has a security association with its MCIP local domain (i.e., IGW) and shares a secret key with its neighboring BSs to facilitate the process of authentication. The proposed SM^{3}P for MCIP network includes two components: secure macro-mobility and secure micro-mobility. Before obtaining service via a local domain, the MS authenticates with the visiting local domain by a secure macro-mobility protocol, and creates a security association with the domain. Then, when the MS moves from a BS to another, a secure micro-mobility protects the process of multi-hop paging/routing cache update of the MS.

7.2 Future Recommendations

7.2.1 Secure Communication for Integrated Internet and MANET

Each MS has limited power and computational capability. Thus, a security protocol for interconnection between the Internet and MANET should be developed in a way that a multi-hop MS can effectively authenticate with the visiting network or other MSs. The private/public key-based protocols cause heavy computation delay compared to secret key-based cryptosystem. However, it is hard for the secret key-based scheme to achieve scalability due to key management. In the future, a hybrid scheme should be developed to facilitate the authentication process with high efficiency and scalability. At the same time, the selfishness is a key issue for multi-hop security that should be considered in the future. During the data transmission phase, an adversary that acts as the intermediary relay node, gets hold of the data packets and mistreats them. In a multi-hop network, the protection for routing security cannot guarantee each intermediary MS to forward the data packet in accordance with the routing table. It is possible to develop a collaborative scheme to encourage packet forwarding in the heterogeneous multi-hop network.

Meanwhile, protecting wireless and mobile infrastructure is anther topic that should be addressed in the future work. The BS/AP protection for future work will primarily be focused on:

- Access control: As illustrated in Chapter 2, a BS may be accessed and reconfigured according to the attacker's convenience. The sensitive information in the BS such as routing information and keys may be captured. Thereby the BS should be protected by some security policies to against unauthorized access and operations.
- Protecting personal information: The personal information of a MS should be properly protected in the device or on the internet relevant databases. The identification such as IP address, secret keys, certificates and other sensitive information should be stored in a proper manner so as to prevent getting compromised and running as a malicious MS.

7.2.1 Secure Mobility Management in Heterogeneous Multi-hop Wireless Networks

The objective of mobility management is to provide a fast hand-off scheme for MSs, which is supported by different layers of the network protocol stack, including (i): network layer, (ii) link layer, and (iii) cross layer design for handoff management. In the future, we can focus on the link layer and cross layer-design to improve our MCIP scheme proposed in Chapter 5.

As a very important issue, the link layer provides crucial information (e.g., signal) for a MS to determine whether it initiates a handoff or not. On the network layer, the topology and velocity information can be helpful for mobility management. The future mobility solutions in the integrated network should be developed based all these information to optimize location management, connection management, and reconfiguration scheme. More research should be done about the handoff management protocols including horizontal handoff, e.g., from one IEEE 801.11 AP to another IEEE 802.11AP, and vertical handoff, from an IEEE 802.11 AP to a cellular BS. A handoff process can be facilitated by cross layer design for the purpose of reducing the handoff delay by detecting the signal strength, or predicting the movement of the mobile station.

The secure macro-mobility and secure micro-mobility protocols can be developed in several ways. The proposed SM^{3}P for MCIP network uses the HA as the authentication server. In some cases, the HA on the Internet may be bogus and thus cannot be trusted for authentication. Therefore, new approaches (e.g., third party authentication) may be developed with a stronger scheme of macro-mobility authentication. During a micro-mobility authentication, the domain IGW maintains the security information (i.e., certificate) for the MS roaming in the domain in a secure manner. Besides a secure handoff process, a future work for secure micro-mobility is how to monitor and measure the local domain information (e.g., the credit or reputation of a MS in the domain) to prevent the attacks from inside or outside.

REFERENCES

[1] D. P. Agrawal and Q. A. Zeng, Introduction to Wireless and Mobile Systems. Brooks/Cole Publishing, 436 pages, ISBN 0534-40851-6, 2003.

[2] D. Cavalcanti, C. Cordeiro, D. Agrawal, B. Xie, and A. Kumar, "Issues in Integrating Networks, WLANs and MANET: A Futuristic Heterogeneous Wireless Network," *IEEE Wireless Communications Magazine*, June 2005.

[3] Sleem, A. Kumar, and K. Kamel, "Using Cell Topography in Predicting Intra-Cell Mobility in Wireless", in *Proc. SCI2003*, Orlando, 2003.

[4] S. Sharma, N. Zhu, and T. C. Chiueh, "Low latency Mobile-IP Handoff for Infrastructure-Mode WLANs", *IEEE Selected Areas in Communications*, Vol. 22, No. 4, 2004, pp. 643-652.

[5] H. Lei and C. E. Perkins, "Ad Hoc Networking with Mobile IP," in *Proc. Second European Personal Mobile Communication Conference*, 1997, pp. 197-202.

[6] D. B. Johnson and D. A. Maltz. "Dynamic Source Routing in Ad hoc Wireless Networks," *Mobile Computing*, Kluwer, ch.5, 1996, pp.1369–1379.

[7] U. Jonsson, F. Alriksson, T. Larsson, P. Johansson, and G. Maguire, Jr., "MIPMANET-Mobile IP for Mobile MANETs," in *Proc. IEEE/ACM Workshop on Mobile and MANETing and Computing*, 1999.

[8] P. Ratanchandani and R. Kravets, "A hybrid Approach to Internet Connectivity for Mobile MANETs," in *Proc. IEEE WCNC*, Vol. 3. Mar. 2003, pp. 1522-1527.

[9] J. Broch, D. A. Maltz, and D. B. Johnson, "Supporting Hierarchy and Heterogeneous Interfaces in Multi-Hop Wireless MANETs," in *Proc. I-SPAN*, June 1999, pp. 370-375.

[10] B. Xie, and A. Kumar, "A Protocol for Efficient Bi-directional Connectivity between Ad hoc networks and Internet," *Journal of Internet Technology, Special Issue on Wireless Ad hoc and Sensor Networks*, Vol.6, no.1 2005, pp. 101-109.

[11] Charles E. Perkins and Pravin Bhagwat, "Highly Dynamic Destination Sequenced Distance-vector Routing (DSDV) for Mobile Computers," *ACM SIGCOMM'94 Conference on Communications Architectures, Protocols and Applications*, 1994, pp. 234-244.

[12] Shree Murthy and J.J. Garcia-Luna-Aceves, "An Efficient Routing Protocol for Wireless Networks," Tech. Rep., Copmuter Engineering, University of California at Santa Cruz, Santa Cruz, CA 95064, 1996.

[13] John Sucec and Ivan Marsic, "Clustering overhead for hierarchical routing in mobile ad hoc networks," Tech. Rep., Rutgers University, 2002.

[14] Atsushi Iwata, Ching-Chuan Chiang, Guangyu Pei, Mario Gerla, and Tsu wei Chen, "Scalable Routing Strategies for Ad Hoc Wireless Networks," Tech. Rep., Department of Computer Science University of California, Los Angeles, 1999.

[15] M. Ng, I Lu, "A Peer-to-peer Zone-based Two-level Link State Routing for Mobile Ad Hoc Networks," IEEE JSAC, pp. 1415-1425, Aug. 1999.

[16] Samir R. Das Charles E. Perkins, Elizabeth M. Belding-Royer, "Ad Hoc On-demand Distance Vector (AODV) Routing," http://www.ietf.org/internet-drafts/draft-ietf-manet-aodv-10.txt, Jan. 2002, Work in progress.

[17] Yih-Chun Hu Jorjeta G. Jetcheva David B. Johnson, David A. Maltz, "The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks (DSR)," http://www.ietf.org/internet-drafts/draft-ietf-manetdsr-07.txt, Feb. 2002, Work in Progress.

[18] Mingliang Jiang, Jinyang Li, and Y.C. Tay, "Cluster Based Routing Protocol (CBRP)," Internet Draft ftp://ftp.leo.org/pub/comp/doc/standards/internet-drafts/manetcbrp-spec/draft-ietf-manet-cbrp-spec-01.txt.gz, August 1999, Work in progress, draft expired.

[19] Vincent D. Park and M.Scott Corson, "A Highly Adaptive Distributed Routing Algorithm for Mobile Wireless Networks," *Proceedings of INFOCOM1997*, 1997.

[20] C-K. Toh and Vasos Vassiliou, The Effects of Beaconing on the Battery Life of Ad Hoc Mobile Computers, chapter 9, pp. 299-321, 2001.

[21] Mario Gerla, Guangyu Pei, Xiaoyan Hong, and Tsu-Wei Chen, "Fisheye State Routing Protocol (fsr) for Ad Hoc Networks," Internet Draft http://www.ietf.org/internet-drafts/draft-ietf-manet-fsr-00.txt, November 2000, Work in progress.

[22] R. Dube et al., "Signal Stability based Adaptive Routing (SSA) for Ad-Hoc Mobile Networks," IEEE Pers. Commun., Feb. 1997, pp. 36–45.

[23] E. M. Royer, C-K Toh, "A Review of Current Routing Protocols for Ad Hoc Mobile Wireless Networks," IEEE Personal Communications April 1999 pp. 46-55.

[24] J. Broch, D. A. Maltz, D. B. Johnson, Y. C. Hu, and J. Jetcheva, "A Performance Comparison of Multi-Hop Wireless Ad Hoc Network Routing Protocols,"

ACM/IEEE MobiCom,1998.

[25] Y. Sun, E. M. Belding-Royer, C. E. Perkins, "Internet Connectivity for ad hoc Mobile Networks," *International Journal of Wireless Information Networks special issue on Mobile MANETs*, 9(2), April 2002.

[26] Papadimitratos and Z.J. Haas, "Secure Routing for Mobile Ad hoc Networks", Proceedings of SCS Communication Networks and Distributed Systems Modeling and Simulation Conference, San Antonio, USA, 2002.

[27] Y.C. Hu, D.B. Johnson, and A. Perrig. SEAD: Secure Efficient Distance Vector Routing for Mobile Wireless Ad Hoc Networks. Ad Hoc Networks Journal, 1 (2003):175-192.

[28] Y-C. Hu A. Perrig, and D. Johnson, "Ariadne: A Secure On-Demand Routing Protocol for Ad Hoc Networks", in the 8th ACM International Conference on Mobile Computing and Networking (MobiCom), September 2002.

[29] A. Perrig, R. Canetti, J. Tygar, and D. Song, "Efficient Authentication and Signing of Multicast Streams over Lossy Channels", 2000 IEEE Symposium on Security and Privacy, pp. 56-73.

[30] B. Dahill, B. Levine, E. Royer, and C. Shields, "A Secure Routing Protocol for Ad Hoc Networks", Univ. of Massachusetts Tech. Rep. 01-37, 2001.

[31] Charles E. Perkins, Elizabeth M. Belding-Royer, and Samir Das, "Ad Hoc On Demand Distance Vector (AODV) Routing." IETF Internet draft, draft-ietf-manet-aodv-10.txt, March 2002 (Work in Progress).

[32] H. Yang, H. Luo, F. Ye, S. Lu, and L. Zhang, "Security in Mobile Ad hoc Networks: Challenges and Solutions," *IEEE Wireless Communication*, Vol. 11, No 1, 2004, pp. 38-47.

[33] C. E. Perkins, "IP Mobility Support for IPV4," Revised, IETF Internet Draft, draft-ietf-mobile-rfc2002-bis-08.txt, Sept. 2001.

[34] D. B. Johnson and C. Perkins, "Mobility Support in IPv6," Internet draft, draft-ietf-mobileip-ipv6-13.txt, November 2000, work in progress.

[35] D. Saha, A. Mukherjee, I. S. Misra, and M. Chakraborty, "Mobility Support in IP: A Survey of Related Protocols," IEEE Network, 2004, pp. 34-40.

[36] A.G. Valko, "Cellular IP: A New Approach to Internet MS Mobility," ACM Com. comm. Rev., 1999.

[37] A. T. Campbell, J. Gomez, S. Kim, A. G. Valkó, C-Y. Wan, and Z. Turányi, "Cellular IP Internet draft," draft-ietf-mobileip-cellularip-00.txt, January 2000, Work in progress.

[38] S. Das, A. Misra, and PAgrawal, "TeleMIP: Telecommunication-enhanced Mobile IP Architecture for Fast Intra-domain Mobility," IEEE Personal Communications, vol.7, issue 4, Aug. 2000, pp 50-58.

[39] J. Zao, S. Kent, J. Gahm, G. Troxel, M. Condell, P. Helinek, N. Yuan, and I. Castineyra, "A Public-Key Based Secure Mobile IP," Wireless Networks, Vol. 5, pp. 373-390, October1999.

[40] S. Jacobs, "Mobile IP Public Key Based Authentication," http://search/ietf.org/internet-drafts/dreaft-jacobs-mobileip-pki-auth-02.txt, 1999.

[41] Sufatrio and K. -Y. Lam, "Mobile-IP Registration Protocol: A Security Attack and New Secure Minimal Public-key Based Authentication," Proceedings of the ISPN, June 1999.

[42] B. Xie, and A. Kumar, D. P. Agrawal, and S. Srinivasan "Securing Macro/micro mobility for Multi-hop Cellular IP," *Elsevier Special Issue of Pervasive and Mobile Computing (PMC) Journal on Security in Wireless Mobile Computing*, Volume 2, Issue 2, Pages 111-136, 2006.

[43] N. Aggelou and R. Tafazolli, "On the relaying capability of next-generation GSM cellular networks," *IEEE Personal Communications*, Vol.8, 2001, pp. 40-47.

[44] Y.-D. Lin and Y.-C. Hsu, "Multihop Cellular: A New Architecture for Wireless Communications," in *Proc. INFOCOM*, Vol.3, 2000, pp.1273 – 1282.

[45] H. Wu, C. Qiao, S. De, and Q. Tonguz, "Integrated Cellular and Ad hoc Relaying Systems: iCAR," *IEEE Journal Selected Areas in Communications*, Vol.19, No: 10, 2001, pp. 2105 - 2115.

[46] X. Wu, S.-H. Chan, and B. Mukherjee, "MADF: A Novel Approach to Add an Ad-hoc Overlay on a Fixed Cellular Infrastructure," in *Proc. IEEE WCNC*, Vol.2, 2000, pp. 549- 554.

[47] Luo, R. Ramjee, P. Sinha, L. (Erran) Li, and S. Lu, "UCAN: A Unified Cellular and Ad-Hoc Network Architecture," in *Proc. ACM MOBICOM'03*, Sept. 2003, pp. 353-367.

[48] T. Rouse, I. Band, and S. McLaughlin, "Capacity and Power Investigation of Opportunity Driven Multiple Access (ODMA) Networks in TDD-CDMA Based Systems," in *Proc. IEEE ICC*, Vol.5, 2002, pp. 3202–3206.

[49] Zadeh, B. Jabbari, R. Pickholtz, and B. Vojcic, "Self-organizing Packet Radio MANETs with Overlay (SOPRANO)," *IEEE Communications Magazine*, Vol. 40, no.6, 2002, pp.149 – 157.

[50] H-Y. Wei and R. Gitlin, "Two-Hop-Relay Architecture for Next-Generation WWAN/WLAN Integration," *IEEE Wireless Communications*, Vol. 11, no. 2,

April 2004, pp.24-30.

[51] H. Hsieh, K-H. Kim, Y. Zhu and R. Sivakumar, "A Receiver-Centric Transport Protocol for Mobile Hosts with Heterogeneous Wireless Interfaces," in *Proc. ACM MobiCom*, Sept. 2003, pp. 1-15.

[52] R. Chang, W. Yeh, and Y. Wen, "Hybrid Wireless Network Protocols," *IEEE Transactions on Vehicular Technology*, Vol. 52, no.4, July 2003, pp.1099-1109.

[53] R. Bruno, M. Conti, and E. Gregori, "Mesh Networks: Commodity Multi-hop MANETs," *IEEE Communication Magazine*, Vol. 43, no. 3, Aug. 2005, pp. 123-131.

[54] K. Ahmavaara, Henry Haverinen, and Roman Pichna, "Interworking Architecture between 3GPP and WLAN Systems," IEEE Communication Magazine, Vol. 41, No. 11, Nov 2003, pp. 74 – 81.

[55] M. Shi, X. Shen, and J.W. Mark, "IEEE 802 Roaming and Authentication in WLAN/Cellular Mobile Networks," IEEE Communication Magazine, Vol. 11, No. 4, Aug. 2004, pp. 66-75.

[56] A. Chakrabarti and G. Manimaran, "Internet Infrastructure Security: A Taxonomy," *IEEE Network*, Vol.15, No.6, Nov/Dec.2002.

[57] A. Fasbender, D. Kesdogan, and O. Kubitz, "Analysis of Security and Privacy in Mobile IP," in *Proc. International Conference on Telecommunication Systems*, March 1996.

[58] B. Awerbuch, D. Holmer, C. Nita-Rotaru, and H. Rubens, "An On-Demand Secure Routing Protocol Resilient to Byzantine Failures," in *Proc. ACM Wise* 2002, Sept. 2002, pp.21-31.

[59] G. M. Koien and T. Haslestad, "Security Aspects of 3G-WLAN Interworking," *IEEE Communication Magazine*, Vol. 41, No. 11, Nov. 2003, pp. 82-88.

[60] B. Xie and A. Kumar, "A Framework for Integrated Internet and Ad hoc Network Security", in *Proc. IEEE Symposium on Computers and Communications*, Vol. 1, June 2004, pp. 318-324.

[61] S. Marti, T.J. Giuli, K. Lai, and M. Baker, "Mitigating Routing Misbehavior in Mobile MANETs", in *Proc. MobiCom*, Aug. 2000, pp. 255-265.

[62] S. Buchegger and J.Y. Le Boudec, "Performance Analysis of the CONFIDANT Protocol," in *Proc. MobiHoc*, Jan. 2002. pp. 226-236.

[63] D. Nayak, N. Rajendran D. B. Phatak and V.P Gulati, "Security Issues in Wireless Local Area Networks", in *Proc. IEEE Canadian Conference on Electrical and Computer Engineering*, Vol. 3, May 2004, pp.1637-1640.

[64] X. Wu and B. Bhargava, "Improving Security in MANET through Integrated Architecture", In *Proc. 1st NSF/ NSA/AFRL Workshop on Secure Knowledge Management*, 2004.

[65] N. Ben Salem, L. Buttyan, J.-P. Hubaux, and M. Jakobsson, "Cooperation in Multi-hop Cellular Networks with Extended Security Analysis," in *Proc. MobiHoc*, June 2003.

[66] L. Buttyan and J.P. Hubaux, "Enforcing Service Availability in Mobile Ad-hoc WANs," in *Proc. MobiHoc*, Aug. 2000, pp.87-96.

[67] L. Butty´an and J.-P. Hubaux, "Stimulating Cooperation in Self-Organizing Mobile Ad Hoc Networks," *Mobile Networks and Applications*, Vol. 8, No. 5, pp. 579–592, Oct. 2003.

[68] B. Raghavan and A. C. Snoeren, "Priority Forwarding in Ad Hoc Networks with Self-Interested Parties," in *Proc. Workshop on Economics of Peer-to-Peer Systems* 2003.

[69] S. Zhong, J. Chen, and Y. R. Yang, "Sprite: A Simple, Cheat-Proof, Credit-Based System for Mobile Ad-Hoc Networks," in *Proc. IEEE Infocom* 2003, pp. 1987–1997.

[70] P.-W. Yau and C. J. Mitchell, "Reputation Methods for Routing Security for Mobile Ad Hoc Networks," in *Proc. Mobile Future and Symposium on Trends in Communications* 2003, pp. 130–137.

[71] E. Huang, J. Crowcroft, and I. Wassell, "Rethinking Incentives for Mobile Ad Hoc Networks," in *Proc. ACM SIGCOMM* 2004, pp. 191–196.

[72] R. Wakikawa, J. T. Malinen, C. E. Perkins, A. Nilsson, and A. J. Tuominen, "Global Connectivity for IPv6 mobile ad hoc networks," Internet Draft (Work in Progress), IETF, 2002.

[73] K. Fall and K Varadhan, "Ns Manual". The VINT Project, http://www.isi.edu/nasam/ns/doc, 2001.

[74] W. Liang, and W. Wang, "A Quantitative Study of Authentication and OoS in Wireless IP Networks," Proceedings of IEEE INFOCOM, Mar. 2005.

[75] P. R. Calhoun, C. E. Perkins - "Diameter Mobile IPv4 Application", draft-ietf-aaa-diameter-mobileip-08.txt, 2001.

[76] B. Xie, A. Kumar, D. Cavalcanti, D. P. Agrawal, and S. Srinivasan "Mobility and Routing Management for Heterogeneous Multi-hop Wireless Networks," Proceeding of IEEE International Workshop on Heterogeneous Multi-Hop Wireless and Mobile Networks, Nov. 2005.

[77] E. De Win, and B. Preneel, "Elliptic Curve Public Key Cryptosystems – An Introduction," Springer Verlag, London, UK, 1997.

[78] IEEE P1363/D13, "Standard Specification for Public Key Cryptography,"1999.

[79] http://www.eskimo.com/~weidai/benchmarks.html.

APPENDIX

An integrated Internet and MANET with size of *n* nodes is considered. The implementation of the key creation, signature generation and signature verification here follows the ECDSA signature techniques as an example of a digital signature for the integrated Internet and MANET. In the integrated Internet and MANET, the FA is responsible for choosing system parameters, a prime **P**, elliptic curve **E**, base point **G** = (x,y), and order **r** of the point **G**. k**G** is a group-point obtained by multiplying the base-point **G** by the scalar K. The system parameters are published by FA and shared within the integrated Internet and MANET. If these parameters are not system-wide parameters, each MS selects its elliptic curve and base point. In this case, these parameters are included in the MS public key. Also let **Hash** (m) be the secure hash algorithm (SHA-1) and know to all in the system.

Key Generation: A node MS chooses an integer, marked as $\mathcal{K}^{1}{}_{\mathcal{MS}}$, as its private key, where $1 \leq \mathcal{K}^{1}{}_{\mathcal{MS}} \leq$ **r-1**. Then MS calculates $\mathcal{K}_{\mathcal{MS}} = \mathcal{K}^{1}{}_{\mathcal{MS}}$ **G** as the public key of the MS.

Signature Generation:

To sign a message (**m**), a MS generate a random number **u**, where $1 \leq$ **u** $\leq$ **r-1** and computes, (i): $\mathbf{V} = u\,\mathbf{G} = (\mathbf{x_V}, \mathbf{y_V})$ and $\mathbf{c} = \mathbf{x_V}$ mod **r,** where **c** is an integer $1 \leq \mathbf{c} \leq r\text{-}1$, if c = 0, the MS another random number u', and starts with the step (i) again.

(ii): $\mathbf{d} = \mathbf{u}^{-1}$**(Hash (m)** + $\mathcal{K}^{1}{}_{\mathcal{MS}}$ **c)** mod **r**. If **d** = 0, the MS another random number u', and starts with the step (i) again.

MS outputs (**c**, **d**) as the signature of message (**m**). MS submits **m**, **c**, and **d**.

Signature Verification:

To verify the signature of a message (**m**, **c**, **d**) from a sender, the receiver calculates,

(i): $\mathbf{h} = \mathbf{d}^{-1}$ **mod r**.

(ii): **h1** = $\mathcal{H}ash\,(m)$ **h** mod r and **h2** = **c** h mod r

(iii): **P** = $h1$ **G** + $h2\ \mathcal{K}_{\mathcal{MS}}$ = $(\mathbf{x_p}, \mathbf{y_p})$ and c' = $\mathbf{x_p}$ **mod r**

If c = c', then the signature is valid. Otherwise it is invalid.

CURRICULUM VITAE

NAME: Bin Xie

ADDRESS: 213 Huntington Park Drive
Louisville, KY 40213

DOB: HuNan, China - October 25, 1970

EUDUCATION: B.S., Electrical Engineering
Central South University
1991-95

M.S., Computer Science
University of Louisville
2001-03

AWARDS: University Fellowship
University of Louisville
2001-04

Student Travel Grant
Programming committee of conference IEEE MASS-2005
2005

E-day Presentation Certificate
Speed Scientific School, University of Louisville
2003-06

Excellent Student Scholarship
Central South University
1992-94

Excellent B.S. Student Award
Central South University
1995

PROFESSIONAL SOCIETIES: IEEE Student Member

REFEREED JOURNALS:

1. Bin Xie, and Anup Kumar, Dharma P. Agrawal, and S. Srinivasan "Securing Macro/micro mobility for Multi-hop Cellular IP," *Elsevier Special Issue of Pervasive and Mobile Computing (PMC) Journal on Security in Wireless Mobile Computing*, Volume 2, Issue 2, Pages 111-136, 2006.

2. Bin Xie, Anup Kumar, D. Cavalcanti, and Dharma P. Agrawal "Multi-hop Cellular IP: A New Approach to Heterogeneous Wireless Networks," *International Journal of Pervasive Computing and Communications (JPCC)*, March, 2006.

3. Bin Xie, and Anup Kumar, "A Protocol for Efficient Bi-directional Connectivity between Ad hoc networks and Internet," *Journal of Internet Technology (JIT), Special Issue on Wireless Ad hoc and Sensor Networks*, Volume 6, Page 101-109, No. 1, 2005.

REFEREED MAGAZINE

4. D. Cavalcanti, C. M. Cordeiro, D. P. Agrawal, B. Xie, and A. Kumar, "Issues in Integrating Cellular Networks, WLANs, and MANETs: A Futuristic Heterogeneous Wireless Network", *IEEE Wireless Communications Magazine, Special Issue on Toward Seamless Internetworking of Wireless LAN and Cellular Networks*, Volume 12, Issue 3, Pages 30-41, 2005.

BOOK CHAPTER

5. Bin Xie, Anup Kumar, and Dharma P. Agrawal "Security Issues in Integrated Cellular Network, WLANs, and MANETs," to appear in "*Wireless Ad Hoc Networking: Personal-Area, Local-Area, and Sensory-Area Networks*", to appear Auerbach Publications, 2006.

REFEREED CONFERENCES

6. Bin Xie, and Anup Kumar, Dharma P. Agrawal, and S. Srinivasan "GMSP: A Generalized Security Multi-hop Protocol for Heterogeneous Multi-hop Wireless Network," *IEEE WCNC*, 2006.

7. Bin Xie, Anup Kumar, D. Cavalcanti, and Dharma P. Agrawal "Mobility and Routing Management for Heterogeneous Multi-hop Wireless Networks," *IEEE international workshop on HMWMN (Heterogeneous Multi-hop Wireless and Mobile Network) in conjunction with MASS*, October 2005.

8. Bin Xie, and Anup Kumar, "A Framework for Internet and Ad hoc Network Security", *IEEE Symposium on Computers and Communications (ISCC)*, June 2004.

9. Bin Xie, and Anup Kumar, “Full Internet Connectivity for Ad hoc Networks”, *1st IEEE International Conference on Mobile Ad hoc and Sensor Systems (MASS)*, October, 2004.

10. Bin Xie, and Anup Kumar, "An Integrated Protocol for Internet and Ad hoc Network communication", *12th International Conference on Advanced Computing and Communication (ADCOM)*, December, 2004.