Hindawi Publishing Corporation
EURASIP Journal on Wireless Communications and Networking
Volume 2009, Article ID 983594, 12 pages
doi:10.1155/2009/983594

Research Article

An ICMP-Based Mobility Management Approach Suitable for Protocol Deployment Limitation

Jeng-Yueng Chen (EURASIP Member),[1,2] Chun-Chuan Yang,[1] Wen-Shiung Chen,[3] Yi-Hung Huang,[4] and Heng-Te Chu[3]

[1] *Department of Computer Science and Information Engineering, National Chi Nan University, Nantou 54561, Taiwan*
[2] *Department of Information Networking Technology, Hsiuping Institute of Technology, Taichung 41280, Taiwan*
[3] *Department of Electrical Engineering, National Chi Nan University, Nantou 54561, Taiwan*
[4] *Department of Mathematics Education, National Taichung University, Taichung 40306, Taiwan*

Correspondence should be addressed to Jeng-Yueng Chen, s2321902@ncnu.edu.tw

Received 19 October 2008; Accepted 9 July 2009

Recommended by Wei Li

Mobility management is one of the important tasks on wireless networks. Many approaches have been proposed in the past, but none of them have been widely deployed so far. Mobile IP (MIP) and Route Optimization (ROMIP), respectively, suffer from triangular routing problem and binding cache supporting upon each node on the entire Internet. One step toward a solution is the Mobile Routing Table (MRT), which enables edge routers to take over address binding. However, this approach demands that all the edge routers on the Internet support MRT, resulting in protocol deployment difficulties. To address this problem and to offset the limitation of the original MRT approach, we propose two different schemes, an ICMP echo scheme and an ICMP destination-unreachable scheme. These two schemes work with the MRT to efficiently find MRT-enabled routers that greatly reduce the number of triangular routes. In this paper, we analyze and compare the standard MIP and the proposed approaches. Simulation results have shown that the proposed approaches reduce transmission delay, with only a few routers supporting MRT.

1. Introduction

Wireless networks are rapidly evolving from 2G cellular telephony networks to 3G and beyond. Following the blooming markets of cellular phone network and Internet services, mobile high-bandwidth data communication is becoming a new promising business niche. Multimedia communications such as voice over IP (VoIP) applications rely increasingly on IP-based techniques [1]. Mobile VoIP applications greatly attract users with their seamless handoff and roaming among different wireless networks while enjoying all of the multimedia services provided by the Internet. In all-IP wireless networks, IP is the key for end-to-end communications, from mobile end-user stations, via gateways, to the Internet, and vice versa. To satisfy users with greater mobility, an efficient protocol supporting mobility is needed for mobile wireless networks [2].

IP mobility works on the OSI network layer and tries to provide mobile hosts with continuous connectivity to the Internet while traveling from their home networks to foreign visiting networks [3]. Internet Engineer Task Force has drawn up a standard of mobility support for IPv4, called Mobile IP (MIP) [4]. The MIP technique is the most common solution for offering seamless handoff to mobile devices over the Internet. In MIP, when the home agent (HA) gets a packet from a corresponding node (CN), it transmits the packet to the mobile mode (MN) by tunneling. Although MIP can provide mobility management without protocol support in the CN, the MIP protocol suffers from problems such as triangular routing, needing home addresses, and temporary unfixed addresses, that is, Care-of-Addresses (CoAs), tunneling management, and so forth [5].

In the standard MIP mechanism, triangular routing increases transmission delay, packet loss, and additional signaling. Route Optimization (ROMIP) [6] uses a binding

update to inform a CN of the current IP address of the MN. The ROMIP can avoid the triangular routing problem and provide a smoother handoff. However, it is very difficult to implement a binding cache in every node of the entire Internet.

Instead of having a binding cache in every node, the work in [7] seeks to avoid the triangular routing problem by reducing the routing path via a mobility management scheme, called Mobile Routing Table (MRT), which operates in every edge router, including home agents and foreign agents. The MRT is actually an extension to the conventional address-mapping table of the MIP or ROMIP. The main idea of the MRT approach is to move the address mapping functionality from the CN to its MRT-enabled edge router. When a CN is going to send an upstream packet to an HA of the MN, the packet will pass through an MRT-enabled edge router. The MRT router first searches its table to check whether the destination IP address exists or not. If not, the router will send packets via an optimal path using the IP routing protocol. Otherwise, the router will send packets to the current CoA of the destination node, found in the associated record. Although only edge routers are needed to provide mobility, we cannot force all the edge routers on the Internet to support the MRT approach. In other words, some routers may support MRT while some may not. To make the MRT approach practical, it is necessary to find appropriate routers that support MRT.

We found appropriate MRT-enabled routers, which reduce the routing path as thoroughly as possible, by developing three different schemes that cooperate with the MRT [8]. After an HA forwards the first packet received from a CN to the MN by tunneling, the HA triggers an MRT router searching procedure. If the HA can find appropriate MRT routers which are located within the path between the CN and HA, then the HA can update those routers with MRT binding information so that the MRT routers are able to forward the follow-up packets to the MN directly. All three schemes proposed in [8] activate the MRT router discovery procedure at the HA. Taken together, all three can be categorized as a single HA-initiated scheme.

The main idea of this work is to introduce a novel approach that supports mobility management without further protocol support in the CN. The best way to achieve this objective is to use an existing protocol for mobility management. In other words, we propose that existing protocols run in conventional nodes for mobility management. In this paper, we have chosen ICMP to help the MRT router discovery procedure. ICMP, as defined by RFC 972 [9], is used for Internet error reporting and generating messages that require attention. Nodes running TCP/IP must contain the ICMP protocol. In such an approach, the CN actually does not need any new protocol installation to support mobility.

Figure 1 illustrates the operation of the HA-initiated scheme of the MRT approach, in which the main steps of the scheme are listed as follows.

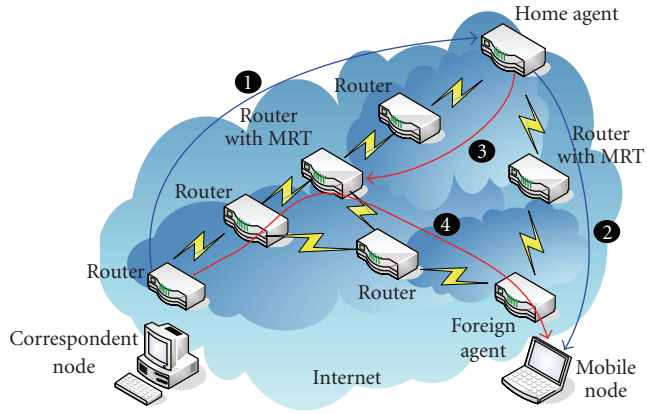


Figure 1: Illustration of traffic flow in the MRT approach.

(1) Since the CN only knows the home address of the MN, it will send all the packets destined for the MN to the HA.

(2) The HA checks its binding information and transmits those packets to the MN by tunneling.

(3) After the HA transmits the first packet to the MN by tunneling, the HA uses one of the proposed schemes to find the appropriate MRT-enabled routers. The HA then uses an MRT update message to insert the binding record to the MRT tables of the MRT routers.

(4) The MRT router can then forward the follow-up packets to the MN directly.

Although the HA-initiated schemes in [8] can completely overcome the protocol deployment limitation, those three schemes may not work well because of security concerns. Moreover, those schemes need a large amount of signaling, which results in unacceptable overhead.

The HA-initiated schemes begin the MRT router discovery procedure when they receive the first packet from the CN. If the procedure can be triggered by the MN when the MN leaves its home network and handoffs to a new foreign network, then the discovery procedure can be started earlier than the HA-initiated scheme. To achieve this earlier triggering, this paper proposes two new MN-initiated schemes, in which the process of MRT router discovery is triggered by the MN. The MN-initiated schemes can find appropriate MRT routers via one or two round trip control messages. Therefore, the signaling overhead can be significantly reduced, an idea our approach inherits from the MRT. Moreover, it also borrows from the enhanced MRT architecture, proposed in [8], thus easily achieving the desired mobility support without a great deal of protocol deployment in the Internet.

The remainder of this paper is organized as follows. Section 2 briefly surveys related mobility management approaches. Section 3 describes the proposed schemes. Section 4 shows the security considerations about our approach. Section 5 evaluates their performances. Finally, we conclude the work in Section 6.

2. Classification of Mobility Management Protocols

Mobility management is an important research issue in enabling a ubiquitous wireless IP network. There have been many related works on mobility management. According to where the binding information is maintained, the mobility management protocols can be divided into the following classes: binding information maintained only at the HA; binding information maintained at the CN; binding information maintained near the MN; binding information maintained near the CN; and miscellaneous approaches.

2.1. Binding Information Maintained Only at the HA. The standard MIP proposed in [4] is a typical method in this category of mobility management protocols. MIP works at the IP layer to support mobility and can benefit upper layers [10]. In MIP, an HA is required to maintain the address mapping and packet forwarding for an MN. The MN sends its binding information to the HA when its current CoA changes. The HA forwards any packets destined for the MN through an IP tunnel to the MN. In this way, the ongoing communications are maintained. However, MIP has several drawbacks [10]. A triangle route that occurs between the MN and the CN causes extra transmission delay and may exacerbate the jitter in real-time applications. The IP-in-IP encapsulation also increases additional system overhead. Moreover, the HA becomes a traffic bottleneck and may also result in a single point of failure problem. In the proposed approach, the maintenance of binding is distributed to the MRT routers, thus eliminating the possibility of a bottleneck or a failure.

2.2. Binding Information Maintained at the CN. Both ROMIP [6] and MIPv6 [11] maintain the binding information at the HAs and the CNs. Packets addressed to an MN home address are transparently routed to its CoA. However, ROMIP and MIPv6 suffer from serious service disruption problems due to long binding delays. Although fast handover for MIPv6 (FMIPv6) [12] can reduce the service disruption time, it deals with localized handoff only. In our approach, every CN does not have to be modified.

2.3. Binding Information Maintained Near the MN. Mobility management protocols that support micromobility maintain the binding information in network agents near the MN. Cellular IP (CIP) [13], HAWAII [14], Telecommunication-enhanced Mobile IP (TeleMIP) [15], Hierarchical Mobile IP (HMIP), [16] and Regional Registration [17] all support micromobility. They utilize the hierarchical structure of the network to localize address binding via a special agent node in each administrative domain which accommodates local handoff within the administrative domain without contacting the HA of the MN. Micromobility protocols usually cooperate with MIP for macromobility support. However, the triangular route problem may still occur when interdomain handoff occurs.

Mailbox [18] and MIP with home agent handover (HH-MIP) [19] propose a special agent, which is located somewhere close to the MN, and supports functionality of the HA. During each handoff, a choice can be made whether to report this handoff to the HA or simply to the special agent. When an MN updates its new location only with the special agent which is close to it, the registration delay can be reduced. These two approaches can reduce both registration and transmission delays, but every CN is required to support the protocols in order to maintain different locations of the MN. In other words, every node in the Internet has to support mobility functionality in these approaches. The effort of protocol deployment is hardly affordable.

2.4. Binding Information Maintained Near the CN. Routing-aware Mobile IP (R-MIP) [20] proposes a router which is near the crossover point between the new and old routing paths; this router can forward packets to the MN directly. By the help of forwarding router discovery and proactive handover procedures, R-MIP enhances the handoff performance and also minimizes packet misordering and bandwidth consumption problems. However, R-MIP can only be used in IPv6 environments. The CN also has to support related procedures if run under IPv4.

By using peer-to-peer technology, the End-system-based Mobile IPv6 (EMIPv6) [21] does not need an HA to manage address mapping. It achieves application transparency by implementing a binding cache in every node. Thus, every CN also has to support the EMIPv6. Moreover, the EMIPv6 can only be used in an IPv6 environment. The EMIPv6 retrieves related CoA information by the help of the Peer Name Resolution Protocol (PNRP) overlay network and distributed subscription and notification services which are the core functions in the EMIPv6. The address maintenance cost is higher than traditional approaches because the MN has to update multiple binding caches maintained by different nodes.

The Virtual Mobility Control Domain (VMCD) [22] is a distributed system that activates multiple anchor points and manages binding information like an HA. Under VMCD, an MN first reaches the CN via its HA with a bidirectional tunnel. The first packet sent by the CN is routed to an MN via the closet anchor point to the CN. After receiving tunneled packets from the anchor point, the MN starts using the anchor point instead of the HA. Therefore, the transmission path can be optimized. The CN does not have to support any new protocol in the VMCD approach. However, the binding cache in an anchor point needs to be synchronized with other anchor points via explicit or implicit signaling. Thus, the VMCD system maintenance cost is very high.

The MRT is a special approach that aims to reduce transmission delay without protocol support in the CN. Although this approach releases the limitation of protocol support in the CN, it demands every edge router support the MRT. The work in [8] loosens the limitation but the protocol overhead is still too high. We will explain the reason in Section 3.1.

2.5. Miscellaneous Approaches. Two application layer protocols, Session Initiation Protocol (SIP) [23] and Mobile Internet Telephony Protocol (MITP) [24], can be extended to support terminal mobility. In the SIP approach, when an MN moves and changes IP address, it just resends a new INVITE message to the CN to reestablish a session. In the MITP approach, when an MN moves and changes IP address, it sends the join and departure messages to MITP servers in order to reestablish a communication session. Although application approaches need not change the underlying protocol stack, they must reestablish connections after an IP address change [25]. Hence, the handoff latency may be large and may not be suitable for real-time multimedia applications.

2.6. Comparison. The comparison between different classes of mobility management protocols is summarized in Table 1.

3. Proposed Enhanced MRT Schemes

3.1. Previous Works. As mentioned earlier, we cannot claim that each CN does or must implement a binding cache. Thus, the MRT approach proposed in [8] to remove such limitations demands that only edge routers must support the MRT. However, even if we ask all the edge routers in the Internet to support the MRT, the approach still results in protocol deployment limitation. In order to reduce the limitation, for example, some routers will support and some will not, there must be a method which is able to find the appropriate routers supporting MRT that can be used to enhance transmission performance.

To overcome those limitations, we have introduced three HA-initiated schemes that cooperate with the MRT to optimize routing paths [8]. When the HA is going to forward a packet received from the CN to the MN by tunneling, the HA will try to find the appropriate MRT-enabled routers along the path between the HA and the CN. If found, the HA will send the MRT binding update messages with corresponding binding information to MRT-enabled routers such that the MRT-enabled routers can forward the successive packets to the MN directly. Hence, the triangular routing problem can be avoided. All the HA-initiated schemes begin the MRT router discovery procedure at the HA and have similar signaling overhead and performance representation. Therefore, we describe a representative scheme, a backward tracking scheme, in detail in what follows. More information on the other two schemes can be found in [8].

The backward tracking scheme uses the ICMP router discovery and SNMP query messages to find the routers that may support MRT. The ICMP messages include "router advertisements" and "router solicitations." Each router periodically broadcasts such advertisements via each of its interfaces to announce the IP address(es) of that interface. After the HA forwards the first packet received from the CN to the MN by tunneling, the HA broadcasts an ICMP router solicitation message to the network where the CN is located. After receiving the ICMP router advertisement message, the HA can find the default gateway of the CN.

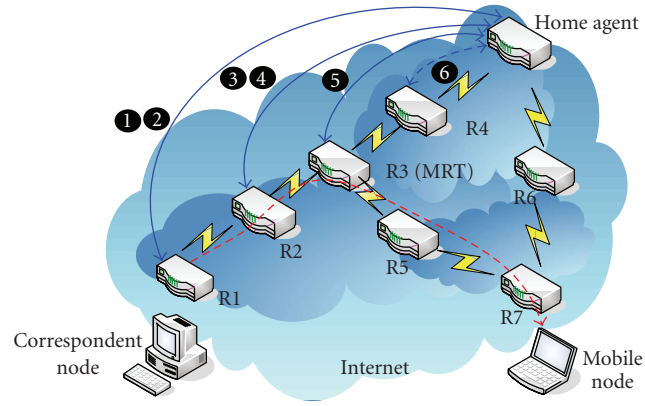


Figure 2: Illustration of backward tracking scheme.

The HA then sends an MRT binding message to the found router. If the HA receives an MRT acknowledgement message, the gateway is an MRT-enabled router. If no MRT acknowledgement is received within the time limit which is set by network administrator, that router does not support MRT. Then the HA uses SNMP query messages and combines with a Reverse Path Forwarding (RPF) [26] mechanism to find a router with one hop closer to the HA. The HA then sends an MRT binding message to the possible router. The HA iterates sending SNMP messages and MRT binding messages to the found routers. The iteration will be stopped when the HA receives an MRT response or tracks back to the HA. This approach can find the possible routers that support MRT in the path from CN to HA. If there is no MRT router located on this path, the process may fail, yet the packets are still transmitted by tunneling at the HA.

Figure 2 illustrates the operation of the backward tracking scheme of the MRT approach, in which the main steps of the scheme are listed as follows.

(1) The HA acquires the address of R1 by an ICMP router solicitation message. The HA then sends an MRT binding message to R1, but no MRT ACK replies.

(2) The HA obtains the address of the next hop router closer to the HA, R2, via SNMP query.

(3) The HA sends an MRT binding message to R2, but still no MRT ACK replies.

(4) The HA obtains the address of the next hop router closer to the HA, R3, via SNMP query.

(5) The HA sends an MRT binding message to R3 and receives an MRT ACK message. The appropriate MRT router is found and can forward the follow-up packets to the MN directly.

(6) If R3 does not support the MRT approach, the backward tracking continues until it reaches the HA.

3.2. Proposed MN-Initiated Schemes. The HA-initiated schemes can support discovering the MRT routers. Unfortunately, those schemes may generate too many control

Table 1: Comparison between different classes.

	Binding information maintained at				
	HA	MN	CN	*Near CN	Misc
Triangular route	High	High	Low	Low	Low
Supports only micromobility	No	Yes	No	No	No
Modification at CN	No	No	Yes	No	Yes
Modification at MN	Yes	Yes	Yes	Yes	Yes
Modification at router	FA	FA	FA	Few	No
Handoff delay	High	Low	High	Medium	High
Packet loss without fast handoff	High	Low	High	Medium	High
Fast handoff support	Yes	Yes	Yes	Yes	No
Signaling cost	High	Low	High	Medium	High

* MRT approach belongs to this class.

messages in order to find a useful MRT-enabled router. Furthermore, the backward tracking scheme starts the searching procedure with ICMP router solicitation, which is sent to the CN by directed broadcast. Many routers may block directed broadcast messages to avoid Denial of Service (DoS) or Distributed DoS (DDoS) attacks by default [27]. The other two HA-initiated schemes also have similar security concerns. Thus, the HA-initiated schemes may not work as expected.

In a client/server model, the MN, for the most part, may act as a client and the CN acts as a server. Therefore, the communications will be initiated at the MN. If the discovery procedure can be triggered when the MN leaves its home network and visits a foreign network, then the execution of the MRT router discovery procedure can be started earlier than the MN-initiated schemes. In other words, if the MN handoffs to a new foreign network and activates the discovery procedure before it communicates with a CN, then the first packet sent from the CN can be redirected to the MN by the discovered MRT router without the tunneling process at the HA.

We propose two new MN-initiated schemes, in which the discovery procedure is activated by the MN. These two schemes can find appropriate MRT routers within fewer control messages than HA-initiated schemes. Therefore, the signaling overhead can be significantly reduced. These two MN-initiated schemes can provide a more efficient discovery procedure while avoiding the security problems that affect other schemes. However, if the communication model is peer-to-peer, the communication may be started with the CN. The MN can activate the discovery procedure when the MN receives the first tunneled packet sent by the CN.

These two new MN-initiated schemes, the ICMP echo scheme and the ICMP destination-unreachable scheme, are presented in the following sections.

3.2.1. ICMP Echo Scheme. In the traditional ICMP echo mechanism, the sender can issue a request packet which can carry any information in the payload. The receiver just sends back a reply packet with the same payload it received. Therefore, we may put the MRT binding information into the payload of an echo request packet. In this scheme, the MN just issues an ICMP echo request packet, in which the MRT binding information is inserted into its payload, when the MN leaves its home network and visits a foreign network or receives a tunneled packet from the CN. The CN will just send back an ICMP echo reply message as usual. The intermediate MRT-enabled routers will identify the payload of the ICMP echo reply message and begin updating the HA of MN with new binding information. Therefore, the MRT-enabled router can forward packets destined to the MN directly.

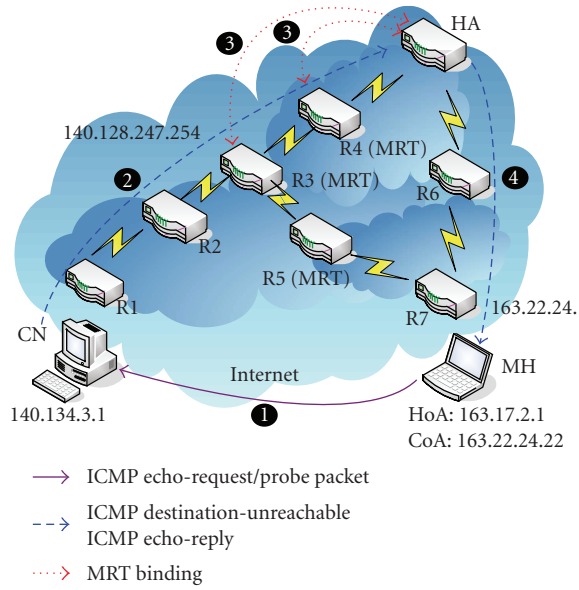
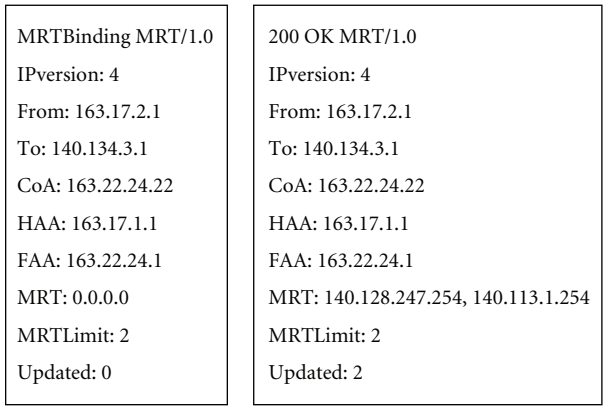


Figure 3: Illustration of MN-initiated scheme.

Figure 3 illustrates the operation of the ICMP echo scheme of the MRT approach, in which the main steps of the scheme are listed as follows.

(1) The MN encapsulates an ICMP echo request packet which contains its binding information as illustrated in Figure 4(a). The packet is sent to the CN by conventional routing.



(a) Echo request (b) Echo reply

Figure 4: An example of payload of echo request/reply messages.

(2) The CN sends back an ICMP echo reply packet with the same payload it received. The packet is sent to the HA.

(3) The intermediate MRT routers, R3 and R4, inspect every ICMP echo reply message. If there is an MRT binding message in the payload, R3 and R4 will send binding requests to the HA listed in the ICMP payload. Thus, R3 and R4 can update its MRT table. During R3 and R4 are sending binding request messages to the HA, they modify the content of the payload as illustrated in Figure 4(b). Then, the modified ICMP echo reply packet is sent to the HA.

(4) After ICMP reply message reaches the HA, HA forwards the packet to the MN by tunneling as usual.

In this scheme, the signaling for mobility management is triggered by the exchanges of ICMP echo messages. The payload of the ICMP echo request and reply messages are specified by means of text-based messages similar to SIP [23]. Figure 4(a) shows an example of an MRT binding message within a payload of an ICMP echo request, and Figure 4(b) shows an example of an MRT binding message within a payload of an ICMP echo reply message.

The MRT binding information within the payload starts with "MRTBinding" used for the MRT-enabled router to quickly identify the packet. The protocol is named MRT in the current version of 1.0. The second line in the example shows that the address binding is an IP version 4 address. The "From" and "To" fields show the addresses of the MN and CN, respectively. The "CoA" field keeps the current CoA of the MN. The "HAA" and "FAA" fields keep the addresses of the HA and FA, respectively. The most important field is "MRT;" the MRT-enabled routers will insert their IP addresses into this field delimited with commas if more than one MRT-enabled router found. The "MRTLimit" filed is used to limit the number of found MRT-enabled routers that will begin its binding update procedure. The "Updated" field is used to store the number of MRT-enabled routers found. Each MRT-enabled router will increase its value by 1 before forwarding. Figure 4(b) shows an example of the payload of ICMP echo reply. It starts with "200 OK," and the "MRT" and "Updated" fields have been modified by the MRT-enabled routers.

3.2.2. ICMP Destination-Unreachable Scheme. As defined in the RFC 792 [9], when a host or router cannot deliver a datagram, the datagram is discarded, and the host or router sends an ICMP destination-unreachable message back to the source host. The code field for this message specifies the reason for discarding the datagram. In this scheme, we use a special transport port number which is unused in well-known Internet services and is used only for the MRT discovery scheme. When the MN leaves its home network and visits a foreign network or receives a tunneled packet from the CN, the MN first sends a probe packet to the CN with a predefined and unused destination port number, for example, 10101. The probe packet is a general UDP message used to trigger CN to reply an ICMP destination-unreachable error report. Thus, The CN will issue an ICMP destination-unreachable message back to the MN with the code filed of ICMP message equal to 3 since the CN does not listen to that port number. The ICMP error message will be transferred to the HA. When the intermediate MRT-enabled routers receive an ICMP destination-unreachable message with the code equal to 3 and the destination port number of the original transport header equals the predefined number, they will issue an MRT binding request to the destination address of the ICMP message and try to update their binding tables. In other words, the ICMP destination-unreachable message is used to trigger the MRT-enabled routers to update their binding tables.

Figure 3 also illustrates the operation of the ICMP destination-unreachable scheme of the MRT approach, in which the main steps of the scheme are listed as follows.

(1) The MN sends a probe message to the CN.

(2) The CN issues an ICMP destination-unreachable error message destined to the home address of the MN.

(3) The intermediate MRT routers, for example, R3 and R4, inspect every ICMP destination-unreachable message. If the error reason is "destination port unreachable" (code equals to 3) and the port number matches, each MRT router sends an MRT binding request to the HA after forwarding the ICMP error message to the next hop.

(4) The HA receives the ICMP destination-unreachable message and tunnels to the MN. The HA also has to inform the MN of the addresses of related MRT routers.

3.3. Handoff Operation. In the original MRT approach [7], the Last Elapsed Time (LET) timer specifies how long an MRT router should wait in the absence of MRT binding update messages about an entry in the binding cache before it removes that entry. In order to keep the accuracy of the MRT binding entries, when the MN moves and changes its CoA, the MRT router has to be updated with a binding message. To

trigger the MRT routers to modify their binding information, we have to store the addresses of related MRT routers. Two nodes may be used to keep those addresses: the HA and the MN. Due to concerns about size of the address table, using the MN to maintain the address table is better than using the HA. Thus, we add an address table in the MN to store the addresses of related MRT routers.

In the ICMP echo scheme, when the CN sends an echo reply message back to the MN, each intermediate MRT router appends its own address to the "MRT" field and forwards the echo packet to the MN via HA. The MN can thus obtain IP addresses of the MRT routers that have been found. In the ICMP destination-unreachable scheme, the intermediate MRT-enabled routers issue the MRT binding request messages to the HA. After replying to the MRT binding response message, the HA also uses a table to keep addresses of MRT routers and informs the MN.

Once the MN moves and changes its CoA, the MN sends a binding update message to its HA and binding warning messages to all the related MRT routers. Those messages are encrypted by the session key described in Section 4.1. The binding warning messages trigger the MRT routers to begin their binding update procedures with HA. Therefore, the MRT routers can quickly forward the following packets to the new CoA accurately.

The number of MRT binding update messages may be too high and affects the system performance if many MRT routers are found by the MN-initiated schemes. However, this can be reduced by limiting the number of MRT-enabled routers to learn. This can be done by reducing the value of "MRTLimit" field in the ICMP echo scheme.

3.4. Signaling Overhead. As mentioned above, the HA-initiated schemes may generate too many control messages. By contrast, the ICMP echo scheme triggered by the MN issues only one single ICMP echo request packet. The signaling overhead of the ICMP echo scheme is very low. The ICMP destination-unreachable scheme also needs few control messages. Moreover, the macrodomain handoffs happen infrequently. It seems that these control messages are unlikely to impose serious overhead to the involved domains.

In most cases, the CN is an Internet server with a fixed location. Furthermore, the HA and CN are not moved during the communications. So, the MRT router discovery process needs to be executed only once for each CN no matter how the MN moves.

3.5. Implementation Cost. Although the MRT approach benefits protocol deployment, it has two deficits: binding cache size and maintenance. First, the MRT routers have to store the binding information of each MN. The MNs also have to store the IP addresses of the MRT routers they used. When the MNs increase and each MN communicates with large numbers of CNs, the cache size may become a serious problem, since most routers only get installed with relatively small amounts of memory. In such a case, the MRT router may bypass the binding information or override the oldest binding record. If the binding cache is large enough to support many MNs, then the cache maintenance will result in large overhead. Therefore, we should choose an appropriate cache size that balances between these competing needs.

In the proposed MN-initiated schemes, each MRT-enabled router has to inspect the ICMP echo reply and destination-unreachable messages. The loading of routers increases as the number of ICMP messages increases.

3.6. Impact of Dynamic Routing. Routing operations in the Internet are dynamic. Packets may be sent through different paths, which mean that packets issued from the CN can be forwarded through the path without MRT routers. In such a case, the HA will send those packets to the MN by tunneling. If the HA-initiated scheme is used, the HA will begin the HA-initiated scheme when a packet reaches the HA. Although this may help the discovery of the potential MRT-enabled routers on different routes, it also increases signaling cost.

If the MN-initiated scheme is used, it can find more potential MRT-enabled routers on the path from the CN to HA. This can reduce the likelihood that packets take a detour on which no MRT-enabled router can be found and the packets reach the HA. However, even if some packets bypass MRT-enabled routers and reach the HA, the HA will send those packets to the MN by tunneling. When the MN receives the HA-tunneled packets, it can do nothing but suffer from longer delays, or the MN can trigger a new MN-initiated scheme to find potential MRT-enabled routers on the current route, which results in increasing signaling overhead.

No matter which strategy is used for discovering new potential MRT-enabled routers, the signaling cost increases. Thus, we prefer that the HA or MN should not trigger any additional MRT discovery scheme to reduce signaling overhead. Furthermore, packets issued from the CN can be forwarded through the path without MRT routers under dynamic routing. In such a case, the MN will receive out-of-order packets, which is normal in dynamic routing.

The most popular routing protocol running between different autonomous systems is BGP in current Internet. BGP is a policy-based routing protocol that routes traffic via predefined policies. Thus, the multiple routing paths are not happened usually under BGP. We think that the proposed MRT approach can still work in most case.

3.7. Comparison between MN-Initiated Schemes. The ICMP echo scheme only needs fewer messages to discover MRT-enabled routers compared with other schemes. The MRT-enabled router informs the MN of its address within the ICMP echo reply message directly, and no additional control message is needed. The overhead is very light compared to the other schemes and we prefer using this scheme in most cases. However, some enterprises or departments, including National Chi Nan University and Hsiuping Institute of Technology, both in Taiwan, may block the ICMP echo packets because of security policies. Thus, the ICMP echo scheme may not work well in all situations.

The ICMP destination-unreachable scheme is a feasible alternative. It needs three round trip messages: the first

round-trip message triggers the MRT routers to start the binding update procedure, which is completed by the second round-trip message. The third round-trip message informs the MN of the addresses of MRT-enabled routers. Although the number of control messages is slightly higher than the ICMP echo scheme, ICMP destination-unreachable messages are generally not filtered out by routers based on security concerns. The ICMP destination-unreachable scheme should, therefore, work well. However, we suggest that the MN use the ICMP destination-unreachable scheme only if the ICMP echo scheme cannot work well, because the ICMP echo scheme generates minimum signaling overhead compared with other schemes.

4. Security Concern

In recent years, with the explosion in web-based commerce and information systems, Internet is now becoming a critical resource whose disruption has financial implications or even dire consequences on human safety. An increasing number of critical services are using the Internet for daily operation. However, not all users on the Internet are good guys. Thus, security becomes more important nowadays. Furthermore, wireless links are more subject to various attacks. There are some security considerations that may affect our approach. First of all, malicious users may forge signaling messages to modify binding information. This may cause a serious security problem. We discuss the authentication problem in Section 4.1. Secondly, ingress filtering has been proposed to filter unauthorized source IP addresses to be transmitted out to Internet. Packets issued by the MN may be filtered out if its source IP address is HoA in the MIP and proposed MRT approach. We discuss the ingress filtering strategy in Section 4.2. Finally, we propose using ICMP protocol to assist discovery of MRT-enabled routers. However, ICMP protocol itself has many security problems such as DoS or DDoS attacks. We discuss the ICMP attack problem in Section 4.3.

4.1. Binding Message Authentication. The MIP, defined in RFC 3344 [4], specifies methods that the MN and HA can authenticate registration requests and replies. The MN performs this authentication by calculating a signature, called an authenticator, and including the signature within authentication extension to the registration request. Relatively, the HA also uses an authentication extension to authenticate its registration reply sent to the MN. In addition to RFC 3344, there are also some other approaches proposed to authenticate the registration messages [28–31].

In the MRT approach, only the authenticated packets can create or change the binding information. When an MN creates a session with a CN, a session key (K_{session}) is assigned by the HA to authenticate subsequent registration messages including binding update messages to its HA and binding warning messages to the MRT routers. The session key is calculated by secure hash function such as HMAC-MD5 [32] or SHA-1 [33]. For example,

$$K_{\text{session}} = \text{MD5}(\text{IP}_{\text{HoA}}, \text{Random}, K_{\text{share}}), \quad (1)$$

where IP_{HoA} is the MN's home address, Random is a random value generated by HA, and the K_{share} is a preshared secret key known by all trusted MRT routers and HAs. Because key distribution can be very complicated, we use shared key mechanism to assist authentication process. After an appropriate MRT router found by proposed schemes, the MRT router begins its binding update procedure with the HA. The MRT router first sends a binding request to the MN's HA. The binding request message is encrypted by the preshared key (K_{share}) such that the HA cannot be easily compromised by malicious routers. After receiving a binding request from the MRT router, the HA replies an encrypted binding message to the MRT routers to create a binding entry for the MN. The session key (K_{session}) for this communication session is also delivered to the MRT routers via the encrypted binding messages. Therefore, the MRT routers can validate further binding warning messages from the MN.

4.2. Ingress Filtering Problem. As mentioned earlier, MIP allows the CN to communicate with the MN without knowing the instantaneous whereabouts of the MN. When MIP is implemented, the CN simply sends its packets toward the MN's HoA. When the MN is away from its home network, these packets are intercepted by the HA, HA then tunnels those packets toward the MN's CoA. In the reverse direction, MIP is not needed for ordinary routing purposes, since IP routing mechanisms make routing decisions only with destination address in the IP packet. Theoretically, the MN could simply use its own HoA as source address in packets it transmits to the CN. In other words, in the reverse direction, no tunneling needs to be implemented and the triangular routing through the HA that exists in the forward direction can be avoided.

However, with the increasing security threats to the enterprise networks, the security designers began to realize that conventional firewalls can only filter inbound traffic would not protect the networks from internal threats. Ingress filtering [34] is now an important security component that prevents attacks from malicious nodes that physically reside inside the network boundaries. When ingress filtering is implemented on an edge router's interface, the router will not forward IP packets received on that interface unless the packet's source address matches the interface network prefix.

To avoid the ingress filtering problem, the MN must use its topologically correct CoA as the source address for any packets it is sending from its current location. However, the packets sent to the CN must include the MN's HoA, in order to hide the routing and mobility complexity from the applications run at the CN. In order to comply with these two conflicting requirements, the MN will tunnel those packets in the reverse direction, using its CoA as the source address in the outer packet to lead the packet through the ingress filtering, while using its HoA as the source address of the inner packet directed to the CN. This mechanism called reverse or bidirectional tunneling [35].

With such a security requirement, we use the same idea as bidirectional tunneling and use the MRT-enabled router to replace the HA. In other words, the MN first sends its

tunneled packets using its CoA as the source address and the address of found MRT-enabled router as destination address in the outer packet. The MRT router will decapsulate the outer packet and then forwards the packet to the CN if the MRT router is not influenced by ingress filtering policy in its network domain. If the MRT router is protected by ingress filtering, then the decapsulated packet should be further encapsulated with a new outer header using the MRT router's address as the source address and then tunneled to the CN.

4.3. ICMP Attack Problem. The work in [36] addresses some security issues in the MIP approach. There are also some security problems that may occur in proposed MRT approach. In order to support mobility management without any new protocol installation, we propose using ICMP mechanisms to assist mobility management. However, ICMP has some serious security problems that we may encounter. For example, hackers may begin DoS or DDoS attacks by flooding with ICMP messages. In the proposed approach, the MRT-enabled routers must inspect ICMP traffic and perform related procedures when special pattern of ICMP messages matched. This may not only increase router's loading as ICMP traffic increases but also be more subject to ICMP attack. In fact, MRT-enabled routers are more easily attacked in such a condition. If the flooded ICMP messages do not match the special pattern defined in proposed schemes, they will not trigger MRT routers to begin their binding procedures. The attack can be dealt with conventional security policy. If the pattern matches, then the MRT routers may be compromised. No matter what condition occurs, the ICMP attack cannot be completely avoided. It can only be reduced to some extent by limiting the number of ICMP messages to be processed within a predefined period.

In our previous research, we proposed that the MRT-enabled routers can update their binding tables after receiving ICMP reply messages sent by the CN. The MRT-enabled routers can update their tables directly because payload in the ICMP echo reply message has already contained binding information. Unfortunately, the binding information can also be modified by a malicious ICMP message. This may cause severe consequence, regarding security problem. Thus, the ICMP message can only be used for triggering the MRT-enabled routers to update their binding information with the HA. Moreover, the binding update should be further protected by the secret key mechanism as explained in Section 4.1.

5. Performance Evaluation

5.1. Simulation Environment and Performance Criteria. In this section, we explore the transmission performance of MIP, ROMIP, and the MRT-based schemes. The topology of simulation contains $100 * 100$ macrodomains. Each macrodomain stands for a business enterprise or an autonomous system that may contain different networking facilities. We randomly select three macrodomains where the MN, CN, and HA are located. We set the probability that routers support the MRT to be the value between 1% and 90% and randomly select nodes that support the MRT (uniform distribution). The MN stays in one domain for 10 seconds and then may have the chance to switch into a neighbor domain based on the moving probability, which is set to 0.5. Simulation executes 10000 times. Each time takes 10000 simulated seconds.

Criteria for performance evaluation and comparison include (1) MRT-enabled routers discovery cost (number of hops used by a signaling message), (2) transmission distance (number of hops used in data delivery sent by the CN), and (3) handoff cost (number of hops used by a handoff signaling message).

5.2. Performance Comparison. In the proposed MRT mechanism, the MN or HA must search the MRT-enabled routers that can be used to reduce transmission distance. Therefore, we first compare the MRT-enabled router discovery cost among the different methods including the HA-initiated and MN-initiated schemes. The discovery cost is measured by calculating transmission hops that the signaling message will pass.

Figure 5 shows the signaling overhead generated by the proposed schemes when the discovery procedures are activated. The HA-initiated scheme: backward tracing needs a lot of control messages, including the router solicitation/advertisement, the SNMP query messages, and the MRT binding update messages. On the other hand, the figure also shows that the MN-initiated scheme seems do not impose serious overhead in the network. The MN-initiated scheme needs few control messages if only a few routers support the MRT. However, we note that the average signaling cost of this scheme increases as the probability of routers supporting MRT increases. In other words, the more routers that support the MRT approach, the more signaling messages are activated. The major reason for this is that the ICMP echo and destination-unreachable messages are only used to trigger the MRT-enabled routers for binding update. The more MRT-enabled routers there are, the more binding related messages will be generated. The signaling overhead can be further reduced by decreasing the value of "MRTLimit" field in the ICMP echo scheme.

The HA-initiated scheme stops the discovery procedure when the HA receives an MRT response or tracks back to the HA. If more routers support the MRT approach, the discovery procedure may finish more quickly. Therefore, the signaling cost decreases as the probability of routers supporting MRT increases.

Within the 10000 simulated seconds, the MN moves randomly among the $100 * 100$ domains based on a moving probability of 0.5. We calculate the average transmission hop count used by the binding update messages generated by the MIP, ROMIP, and MRT approaches. We also calculate the average transmission hop count used by the data packets sent by the CN destined to the MN. Figure 6 shows the average transmission distance used by the CN when the CN sent data to the MN.

We first observe a significant reduction of hop counts in the MRT-based approach. Both MIP and ROMIP approaches

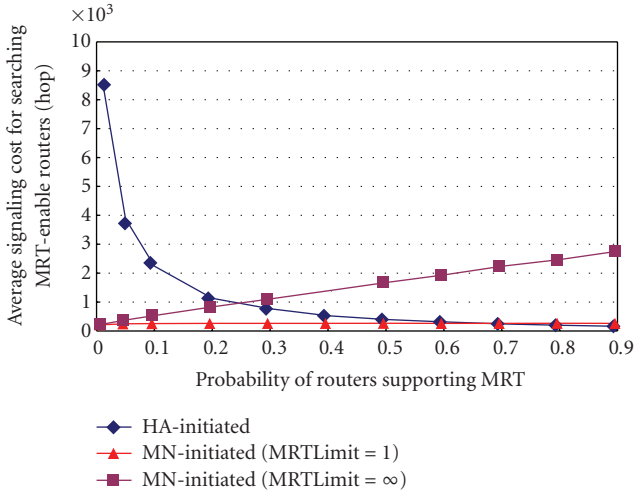


Figure 5: MRT-enabled routers discovery cost.

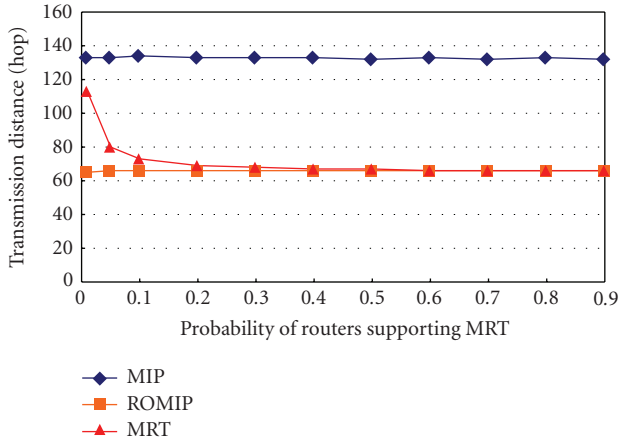


Figure 6: Transmission distance versus MRT probability.

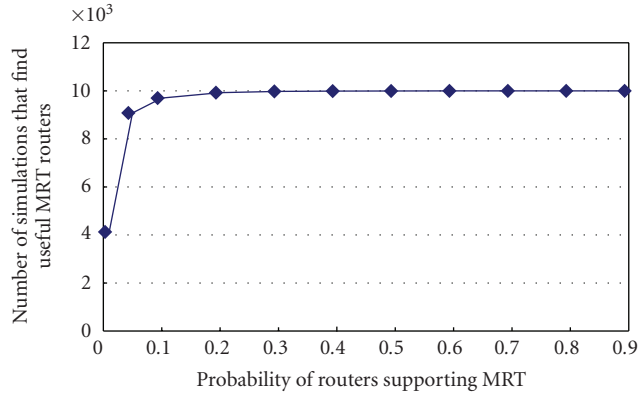


Figure 7: Number of simulations that find an MRT router.

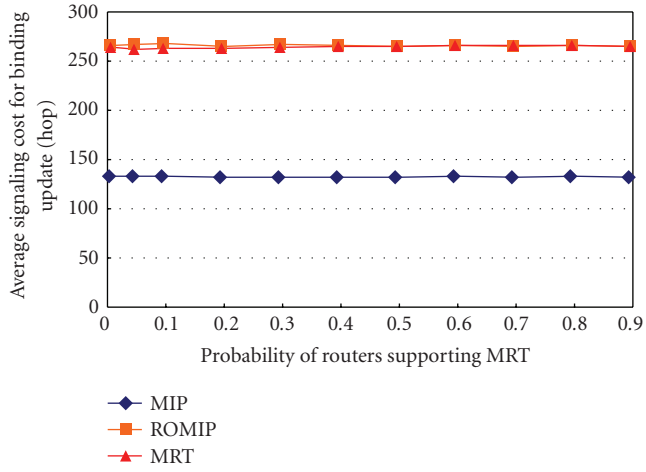


Figure 8: The handoff cost versus the probability of routers supporting MRT.

are not affected by the distribution of the MRT-enabled routers. When the probability that routers support the MRT is only 0.01, the average hop count (113) of the MRT-based approach is closer to the MIP (133). In the 10,000 repeated simulations, an MRT-enabled router located in the path between the HA and CN was found 4,125 times, as shown in Figure 7. The average distances between the CN and HA is approximately 60 hops. So, there may be more chances that routers will support the MRT.

When the probability that routers support the MRT is 0.1, the average hop counts (73) of the MRT-based approaches are closer to the ROMIP (66). We can see that the average hop count of the MRT approach is quickly converging to the ROMIP approach when the probability that routers support the MRT increases. On the average, the transmission hop for the MRT approach will be less than that of the MIP and greater than that of the ROMIP. However, the ROMIP approach needs to implement a binding cache at every CN. On the contrary, the MRT approach simply burdens routers. Furthermore, it is not necessary to implement the MRT scheme at every router. For the worst case of the MRT approach, even if the MN or HA cannot discover any MRT-enabled router, the transmission hop count degrades to the original MIP. For the best case, once the MN or HA discovers the MRT router that is next to the CN, the transmission hop count for the MRT approach is close to the ROMIP approach.

Figure 8 shows the comparison of the average binding maintenance cost of different approaches. The average binding maintenance cost is measured by calculating transmission hops that the binding messages will pass. In the MIP approach, the signaling cost contains the binding updates and acknowledgements exchanged between the MN and HA. In the ROMIP approach, the signaling cost contains the binding update messages sent by the MN destined to the HA and previous FA, the binding acknowledgements replied by the HA and previous FA, the binding warning messages sent by the previous FA destined to the CN, and the binding information exchanged between the CN and HA. In the MRT approach mechanism, the signaling cost is similar to the ROMIP except CN changed to MRT router.

It is clear that the MIP approach has the minimum handoff signaling cost and the ROMIP approach has the

maximum signaling cost. Both MIP and ROMIP approaches are not affected by the distribution of the MRT-enabled routers. Although the signaling cost of the MRT approach is quite close to the ROMIP approach, the MRT approach gains the protocol deployment benefit.

6. Conclusion

The MRT approach is an efficient mechanism to avoid the triangular routing problem by removing the need for a binding cache at each CN. This paper illustrated enhanced MRT schemes suitable for mobile wireless environments. Two efficient schemes were proposed to cooperate with the MRT table for removing all the limitations of the original MRT approach. Under the proposed architecture, the CN is not required to support any mobility management protocol. The proposed methods can still work well even with only a few routers supporting the MRT. Besides, security problems may affect the effectiveness of proposed approach. Some security consideration and possible countermeasures are also discussed in this paper. Simulations have shown that the proposed schemes equal, in the best case, the performance of ROMIP approach, and are only slightly weaker than the MIP approach.

Acknowledgment

This work was supported in part by the National Science Council, Taiwan, under Grants NSC93-2219-E-260-004 and NSC95-2219-E-260-005.

References

[1] R. Ramjee, T. F. La Porta, L. Salgarelli, S. Thuel, and K. Varadhan, "IP-based access network infrastructure for next-generation wireless data networks," *IEEE Personal Communications*, vol. 7, no. 4, pp. 34–41, 2000.

[2] Y.-H. Huang, J.-Y. Chen, W.-S. Chen, C.-C. Yang, and H.-T. Chu, "A comparison between SIP and network layer mobility management protocols in IP-based wireless networks," in *Proceedings of the 5th IEE International Conference on 3G Mobile Communication Technologies (3G '04)*, no. 503, pp. 317–321, 2004.

[3] J. D. Solomon, *Mobile IP: The Internet Unplugged*, Prentice-Hall, Upper Saddle River, NJ, USA, 1998.

[4] C. Perkins, "IP mobility support for IPv4," *RFC 3344*, 2002.

[5] A. T. Campbell, J. Gomez, S. Kim, Z. R. Turányi, A. G. Valkó, and C.-Y. Wan, "Internet micromobility," *Journal of High Speed Networks*, vol. 11, no. 3-4, pp. 177–198, 2002.

[6] D. Johnson and C. Perkins, "Route optimization in mobile IP," *draft-ietf-mobileip-optim-11.txt*, 2001.

[7] I.-W. Wu, W.-S. Chen, H.-E. Liao, and F. F. Young, "A seamless handoff approach of mobile IP protocol for mobile wireless data networks," *IEEE Transactions on Consumer Electronics*, vol. 48, no. 2, pp. 335–344, 2002.

[8] J.-Y. Chen, W.-S. Chen, Y.-H. Huang, C.-C. Yang, and H.-T. Chu, "An enhanced mobile IP protocol for mobile wireless networks," in *Proceedings of the International Conference on Software, Telecommunications, and Computer Networks (SoftCOM '04)*, pp. 419–423, 2004.

[9] J. Postel, "Internet control message protocol," *RFC 792*, 1981.

[10] W. M. Eddy, "At what layer does mobility belong?" *IEEE Communications Magazine*, vol. 42, no. 10, pp. 155–159, 2004.

[11] D. B. Johnson, C. E. Perkins, and J. Arkko, "Mobility support in IPv6," *RFC 3775*, June 2004.

[12] R. Koodli, "Fast handovers for mobile IPv6," *RFC 4068*, July 2005.

[13] A. Campbell, J. Gomez, C-Y. Wan, S. Kim, Z. R. Turányi, and A. G. Valkó, "Cellular IP," *draft-ietf-mobileip-cellularip-00.txt*, January 2000.

[14] R. Ramjee, T. La Porta, S. Thuel, K. Varadhan, and S. Y. Wang, "HAWAII: a domain-based approach for supporting mobility in wide-area wireless networks," in *Proceedings of the 7th International Conference on Network Protocols (ICNP '99)*, pp. 283–292, Toronto, Canada, October 1999.

[15] S. Das, A. Misra, P. Agrawal, and S. K. Das, "TeleMIP: telecommunications-enhanced mobile IP architecture for fast intradomain mobility," *IEEE Personal Communications*, vol. 7, no. 4, pp. 50–58, 2000.

[16] H. Soliman, C. Castelluccia, K. El-Malki, and L. Bellier, "Hierarchical mobile IPv6 mobility management (HMIPv6)," *RFC 4140*, 2005.

[17] E. Gustaffson, A. Jonsson, and C. Perkins, "Mobile IP v4 regional registration," *draft-ietf-mobileip-reg-tunnel-09.txt*, July 2004.

[18] J. Cao, L. Zhang, H. Chan, and S. K. Das, "Design and performance evaluation of an improved mobile IP protocol," in *Proceedings of the Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '04)*, vol. 1, pp. 319–329, 2004.

[19] L.-S. Yu and C.-C. Yang, "An enhancement of mobile IP by home agent handover," in *Proceedings of 62nd IEEE Semiannual Vehicular Technology Conference (VTC '05)*, September 2005.

[20] T. Takahashi, K. Asatani, J. Harju, and H. Tominaga, "Proactive handover scheme based on forwarding router discovery for mobile IP networks," *IEICE Transactions on Communications*, vol. E88-B, no. 7, pp. 2718–2725, 2005.

[21] C. Guo, H. Wu, K. Tan, et al., "End-system-based mobility support in IPv6," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 11, pp. 2104–2116, 2005.

[22] R. Wakikawa, Y. Ohara, and J. Murai, "Virtual mobility control domain for enhancements of mobility protocols," in *Proceedings of the Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '05)*, vol. 4, pp. 2792–2797, 2005.

[23] J. Rosenberg, H. Schulzrinne, G. Camarillo, et al., "SIP: session initiation protocol," *RFC 3261*, June 2002.

[24] W. Liao, "Mobile internet telephony protocol: an application layer protocol for mobile Internet telephony services," in *Proceedings of the IEEE International Conference on Communications (ICC '99)*, vol. 1, pp. 339–343, 1999.

[25] T. T. Kwon, M. Gerla, S. Das, and S. Das, "Mobility management for VoIP service: mobile IP vs. SIP," *IEEE Wireless Communications*, vol. 9, no. 5, pp. 66–75, 2002.

[26] R. Ogier, F. Templin, and M. Lewis, "Topology dissemination based on reverse-path forwarding," *RFC 3684*, 2004.

[27] Cisco Systems, "Improving security on Cisco routers," http://www.cisco.com/warp/public/707/21.pdf.

[28] J. Zao, S. Kent, J. Gahm, et al., "A public-key based secure mobile IP," *Wireless Networks*, vol. 5, no. 5, pp. 373–390, 1999.

[29] C. C. Yang, M. S. Hwang, J. W. Li, and T. Y. Chang, "A solution to mobile IP registration for AAA," in *Mobile Communications*,

vol. 2524 of *Lecture Notes in Computer Science*, pp. 329–337, Springer, Berlin, Germany, 2003.

[30] H.-S. Kang and C.-S. Park, "A key management scheme for secure Mobile IP registration based on AAA protocol," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E89-A, no. 6, pp. 1842–1846, 2006.

[31] C. Perkins, "AAA registration keys for mobile IP," *RFC 3957*, 2005.

[32] H. Krawczyk, M. Bellare, and R. Canetti, "HMAC: keyed-hashing for message authentication," *RFC 2104*, February 1997.

[33] D. Eastlake and P. Jones, "US secure hash algorithm 1 (SHA1)," *RFC 3174*, September 2001.

[34] P. Ferguson and D. Senie, "Network ingress filtering: defeating denial of service attacks which employ IP source address spoofing," *RFC 2267*, January 1998.

[35] G. Montenegro, "Reverse tunneling for mobile IP," *RFC 3024*, January 2001.

[36] W. Haitao and Z. Shaoren, "The security issues and countermeasures in mobile IP," in *Proceedings of the International Conference on Info-Tech and Info-Net (ICII '01)*, November 2001.