MOBILITY MANAGEMENT IN NEXT GENERATION NETWORKS

XIE QUNYING

(B.Eng, Xi'an JiaoTong University, PRC)

A THESIS SUBMITTED
FOR THE DEGREE OF MASTER OF ENGINEERING
DEPARTMENT OF ELECTRICAL & COMPTER ENGINEERING
NATIONAL UNIVERSITY OF SINGAPORE
2004

ACKNOWLEDGEMENTS

First of all I would like to thank my supervisor Dr. Winston K. G. Seah for his enlightening advices and guidance during the elaboration of this work. What I have learned from him will provide me with lifetime benefits.

I would also like to thank Dr. Hoang M. Nguyen for his invaluable and patient guidance, encouragement and support accompanying me in every stage of my research.

Moreover, I wish to thank Mr. Paul Tan for many insightful discussions as well as the pleasant cooperation in the process of writing MWCN paper. Many thanks should be given to my senior Li Feng and Mr. He Dajiang for their great help in the simulation works. I am also grateful to Dr. Li Tonghong for his valuable suggestions on my thesis work. There remain so many thanks to lots of friends around. Although I can not list their names one by one, I should express my sincere appreciations for their friendship and I will not forget the precious time we spent together.

Finally, I wish to express my deep feeling to my parents and my sister. It is their full support and encouragement that makes me to pursue my goals all the way.

TABLE OF CONTENTS

SUMMARY

The next generation network is envisioned to evolve towards a convergence of wireless networks and the Internet, as well as towards convergence of voice and data into a common packet-switched network infrastructure. Among the existing packet technologies, the Internet Protocol (IP) has been adopted as a unifying network layer to support a multitude of link layer standards and technologies. The “All-IP” concept, which makes both strong economic and technical sense, extends IP solution to access networks and is promising in enabling terminal mobility across a range of wireless networks (e.g. wireless LAN and ad hoc networks). Mobility management is a significant aspect of mobile wireless networks for enabling mobile nodes to maintain communication sessions while moving.

In this thesis, we propose mobility management schemes in two scenarios:

1) The existing mobility management scheme for IP network is Mobile IP (v4 or v6) and other extended protocols, but considering the stringent requirement of real-time multimedia services, the packet loss and delay caused by the movement of users is not well addressed by Mobile IP. Multi-Protocol Label Switching (MPLS) is a technology which, when used in conjunction with IP, substitutes conventional IP address lookup and forwarding within a network with faster operations of label lookup and switching. Because of its added benefits, we adopt MPLS as the

layer below IP in an all-IP network model to realize a seamless handover scheme for an IP/MPLS based Hierarchical Mobile IPv6 network. By using Layer 2 (L2) trigger to reduce movement detection latency and taking advantage of Hierarchical Mobile IPv6 (HMIPv6) to reduce binding update delay, the handover performance can be enhanced. Our simulation results show that the handover delay and packet loss are greatly reduced.

2) With the observation that most existing research work on mobility management is done with the assumption that the mobile node must have link-layer connection with access point, we think it is worthwhile to study how to provide mobility management for those mobile nodes multi-hops away from the access point. We propose a mobility management scheme that aims to provide mobile nodes a continuous Internet connectivity in a hybrid network, which is a combination of the Internet and Mobile Ad hoc Networks (MANET). In this thesis, a multi-hop handover scheme is designed and through simulation we demonstrate that our scheme can reduce handover delay and packet loss.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AODV	Ad hoc On-demand Distance Vector
AP	Access Point
AR	Access Router
ARP	Address Resolution Protocol
BACK	Binding Acknowledge
BS	Base Station
BU	Binding Update
CBR	Constant Bit Rate
CN	Correspondent Node
CoA	Care of Address
FA	Foreign Agent
FEC	Forwarding Equivalence Class
FMIPv6	Fast Handover for Mobile IPv6
GW	Gateway
HA	Home Agent
HMIPv6	Hierarchical Mobile IPv6
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IP	Internet Protocol
LER	Label Edge Router
LSP	Label Switched Path
LSR	Label Switching Router
MAC	Medium Access Control
MANET	Mobile Ad hoc Network
MAP	Mobility Anchor Point
MIPv6	Mobile IPv6
MPLS	Multi-Protocol Label Switching
MN	Mobile Node
NAR	New Access Router
NS	Network Simulator
PAR	Previous Access Router
RA	Router Advertisement
UDP	User Datagram Protocol
VoIP	Voice over IP

CHAPTER 1 INTRODUCTION

1.1 Overview

The next generation networks will consist of multiple wireless IP access networks and wired IP networks. Most wireless IP nodes will be mobile and thus change their points of network attachments. Normally, there are two types of network attachment points: BS (base station) and AR (access router). The BS is a link layer device that provides connectivity between wireless hosts and the wired network. The AR is the edge router in the wireless IP access network that provides routing services for the wireless hosts. Therefore, a wireless IP node in motion may experience two types of handover: link-layer handover that is between two base stations and IP-layer handover that is between two ARs. With the increasing demands of mobile users for various services including voice, data and multimedia, next generation networks will evolve towards convergence of voice and data into a common packet-based network. An all-IP network is a promising solution, which uses IP technology from access network to core network [1][2][3]. The advantages of the all-IP network are cost reduction compared with traditional circuit-switched network and independent from radio access technology. In all-IP networks, the IP technology can be extended to traditional BS, namely, the function of AR is incorporated into BS. In this thesis, the AR that we refer to is located at the traditional BS's position and performs the functionalities of both traditional BS and AR's.

Mobility is one of the characteristics of wireless network, and thus mobility management is a key issue in all-IP networks. The task of mobility management is basically to enable network applications to continuously operate at the required quality of service throughout an IP-layer handover. While buffering and forwarding packets to the new base station from the old base station could be used to reduce packet loss due to handover, this procedure can introduce unacceptable delay into real-time media applications such as VoIP. Therefore, it is important to minimize the handover latency, which is defined as the period in which the mobile node is unable to receive application traffic during handover.

1.2 Contribution

The Mobile IP protocol provides fundamentally important functions for mobility management in the wireless IP network, but its functionality only realizes the very basic set of capabilities. A lot of research has been done to develop technologies that will enhance, or complement the basic Mobile IP in various areas. Our research presented in this dissertation is also in this direction.

The main contributions of this thesis are:

- Presented in [21], a seamless handover scheme in IP/MPLS based Hierarchical MIPv6 network is proposed. By using L2 trigger, the movement detection latency is reduced. Therefore, Layer 3 (L3) handover can be performed faster and the total handover latency as well as packet loss during handover is decreased. The use of bicasting can further reduce packet loss during handover.

- Presented in [22], a method to reduce L3 handover latency is proposed by extending the IEEE 802.11 management frame. This extension enables mobile nodes to discover neighboring candidate access routers more quickly and efficiently.
- In chapter 4, an efficient mobility management scheme providing continuous Internet connection for MANET nodes is presented. We propose a multi-hop handover scheme with approaches to reduce handover latency and consider load balancing in gateway selection algorithm. The impact of multi-hop handovers to the communication between MNs and CNs in the Internet is studied through simulation.

1.3 Organization

The remainder of the thesis is organized as follows. Chapter 2 reviews relevant background. Chapter 3 presents a seamless handover scheme in MPLS-based Hierarchical Mobile IPv6 networks. Chapter 4 presents a mobility management scheme that integrates Hierarchical Mobile IPv6 and AODV protocol to provide MANET nodes continuous connectivity with the Internet and discusses multi-hop handover in hybrid networks. Chapter 5 analyzes the scheme performance through simulation results. The conclusion and future works are given in Chapter 6.

CHAPTER 2 BACKGROUND

2.1 Mobility Management

2.1.1 Overview

There are three types of mobility [4] :

- Terminal mobility refers to the ability of the network to route calls or packets to a mobile node regardless of the type of network it is attached to. It allows the terminal to change location while maintaining all services, a familiar example of this is the SIM card mobility. With a SIM card plugged into a handphone, we can receive calls wherever in the whole country. The mobility management what we concerned in this thesis is terminal mobility.
- Personal mobility allows a user to access all services independently of terminals and networks, e.g., Virtual Home Environment (VHE) is the concept that a mobile user can get the same computing environment on the road as that in their home or corporate computing environment.
- Service mobility allows the service accessible through different network domains.

Mobility management contains two components:

- Location management:

Location management is a two-stage process: 1) Location registration (location update). In this stage, the mobile terminal periodically notifies the network of its new access point, allowing the network to authenticate the user and revise the user's location profile. 2) Call delivery. When a call comes, the network will query for the user's location profile, if the location profile just gives an approximate position of the terminal, the network will searches for the MN by sending messages to the cells close to the last reported location of the MN. When the called terminal receives the message, it will reply to network, and then the network will know its specific position. This process is called paging.

- Handover management:

Handover occurs only when the MN is transmitting or receiving data, the handover function can ensure users continuously get service while moving. Consequently, it is the most important part in mobility management. The three-stage process is: 1) Initiation: either the user or a network agent identifies the need for handover. 2) New connection generation: the network must find new resources for the handover connection and perform routing operations. 3) Execution phase: the data will be delivered from the old connection path to the new connection path.

Concerned with mobility management in the Internet, the famous Mobile IP protocol provides MNs mobility support that is transparent above the IP layer. There are different work groups in Internet Engineering Task Force (IETF), which study various aspects of mobility management. The previous Mobile IP Working group has been separated to three new working groups: MIPv4 Work Group (MIP4 WG), MIPv6 Work Group (MIP6 WG), and MIPv6 Signaling and Handoff Optimization (MIPSHOP). The basic Mobile Internet Protocol (MIP) is designed to provide IP mobility support for IPv4 nodes, which is specified in RFC3344. The MIP (v4) protocol support transparency above the IP layer and is currently deployed on a wide basis (e.g. in CDMA2000 networks). Later, Mobile IPv6 (MIPv6) [6] protocol (currently is studied under MIP6 WG) is proposed to support IP mobility for IPv6 hosts. MIPv6 outperforms MIPv4 on aspects such as built-in feature for route optimization and using IPv6 Neighbor Discovery Protocol (NDP) [7] instead of ARP so that it is decoupled from any particular link layer. To address the issues of signaling overhead, handover latency, and packet loss in MIP, Hierarchical Mobile IPv6 (HMIPv6) [10] and Fast Handover for Mobile IPv6 (FMIPv6) [11] have been developed. The two specifications are now being further studied by MIPSHOP WG. In the following sections, we will introduce the MIPv6 protocol, HMIPv6 protocol, and FMIPv6 protocol respectively.

2.1.2 Mobile IPv6 (MIPv6)

The main goal of Mobile IP (MIP) is that a mobile node is always addressable by its home address, whether it is currently attached to its home link or is away from home. MIP enables applications running on a mobile node to survive physical reconnection by inserting a few additional features at the network layer. These features allow the mobile node to always be addressable at its home address.

This mechanism is completely transparent for all layers above IP, e.g. for TCP, UDP and all applications.

In MIPv6 [6], three operation entities are defined: Mobile Node (MN), Correspondent Node (CN), and Home Agent (HA); four new IPv6 destination options are defined: Binding Update, Binding Acknowledgement, Binding Request and Home Address option; two ICMP messages are defined for "Dynamic Home Agent Address Discovery": ICMP home agent address discovery request message and ICMP home agent address discovery reply message; two new IPv6 options for "Neighbor Discovery": advertisement interval option and home agent information option.

MIPv6 is based on version 6 of the IP protocol. Therefore MIPv6 has a set of features present in IPv6. The main features are:

- Router advertisements (RA): RA is a message sent by routers on the networks they serve to inform hosts about their presence. An RA message contains the network prefix of the network and the address of the router that sends the advertisement.
- Neighbor discovery (ND): ND is a mechanism defined in IPv6 to let a host know the link-layer addresses of other nodes directly attached to the host. When a host connects to a network, it multicasts a neighbor solicitation message to other nodes at the network, which contains the link layer address of the host. Each node at the network replies to the host neighbor

advertisement message which contains the link layer address of the node and the source is the IP address of the node. MIPv6 exploits the ND feature to let a home agent intercept packets for a mobile node at home network and let a mobile node to locate routers when it attaches to foreign networks.

- Auto-configuration: auto-configuration is a mechanism that allows a host to automatically discover and register parameters needed to connect to the Internet. Two types of auto-configuration are provided by IPv6: 1) Stateless auto-configuration: a host generates its own IP address based on the network prefix and the IEEE 802 address of its network interface. It does not require consulting with server to form an IP address. 2) Stateful auto-configuration: a host multicasts a message to all Dynamic Host Configuration Protocol (DHCP) servers on the network, and DHCP servers reply with the parameters to the host to configure an IP address. In MIPv6, MNs use auto-configuration to construct the care-of-address (CoA) whenever they move to a foreign network.

IPv6 introduces header extensions to be inserted between the IPv6 header and the payload data. The feature of destination options is that they only need processing at the destination of the packet. Thus the intermediate nodes ignore destination options. The four new destination options provided by MIPv6 are:

- Binding Update (BU): BU option is used by an MN to inform its home agent or CN about its current care-of address.

- Binding Acknowledgement (BACK): BACK option is used to acknowledge the reception of a BU, if an acknowledgement is required.
- Binding Request (BR): The BR option can be used by any node to request an MN to send a BU.
- Home Address (Haddr): The Haddr option is used by a sender MN to inform the receiver about the sender's home address. It is used when an MN is attached to a foreign network and the routers perform ingress filtering.

All destination options can be piggy-bagged on a data packet, which can reduce the overhead of exchanging mobility information.

Three conceptual data structures are used in MIPv6:

- Binding cache: Binding cache is maintained by HAs and CNs. A binding cache is used to hold the binding for MNs. If a node receives a BU destined for it, it will add the binding <MN's CoA, MN's Haddr> to its binding cache. Before a node sends a packet, it checks the binding cache. If there is an entry for the destination of the packet, the packet is instead sent to the CoA mapped by the destination.
- Binding update list (BU list): BU list is maintained by an MN, which records the nodes that must receive BU. Each time an MN sends a BU, an entry in the BU list will be added or renewed.
- Home agent list (HA list): HA list is maintained by routers that serve as HAs for networks, which contains information of all HAs present at a

network and these HAs' individual preference. The information in a HA list is learned from RAs by MNs to perform dynamic HA discovery.

Location management:

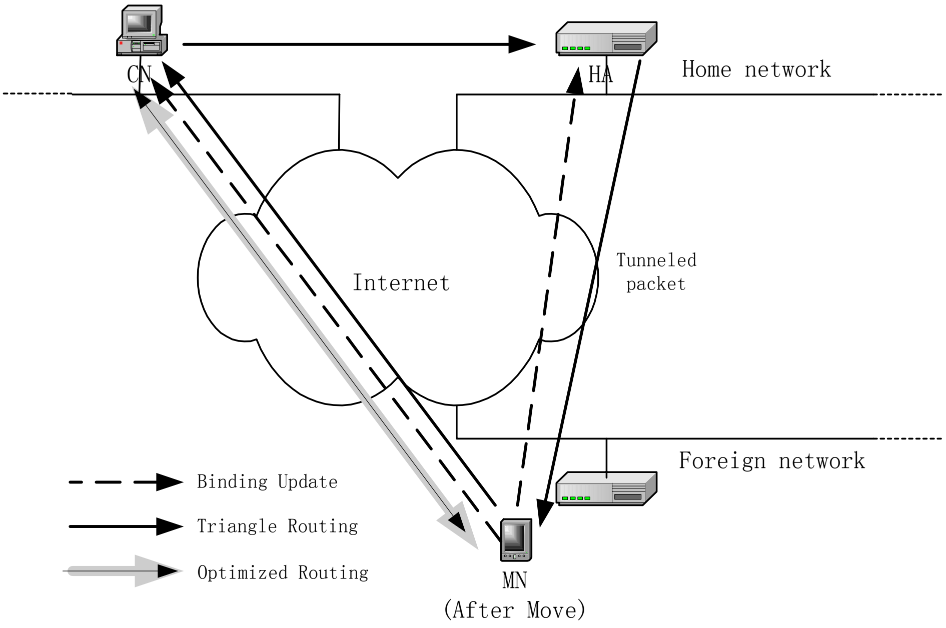


FIG 2.1 NETWORK TOPOLOGY (MIPV6)

- HA registration: Fig 2.1 shows the MIPv6 network topology. When an MN moves away from home, it selects one AR as its default router and uses the network prefix advertised by that AR as the network prefix of its primary care-of address. After a care-of address has been created using either stateless or stateful address auto-configuration, the MN creates a BU message containing the new care-of address and the MN's home address and sent to its HA. The HA registers the binding by adding or updating the binding in its binding cache and replies with a BACK message to the MN.
- Triangle routing: As illustrated in Fig 2.1, when an MN communicates with

a CN while being away from home, packets are routed from the CN to the HA and from the HA to the MN, while packets from the MN are routed directly to the CN. This phenomenon is called triangle routing. If an MN's point of attachment is far from the HA, triangle routing can cause a significant overhead compared to the direct route between a CN and an MN.

- Route optimization: To avoid triangle routing an MN can send BU to CN (as shown in Fig 2.1). Then CN can cache the MN's current care-of address and send packets directly to the MN. Any IPv6 node sending packets must first check its binding cache for the packet's destination address. If an entry is found, a routing header containing the MN's home address is added to the packet and the destination address is set to the MN's care-of address. When the MN receives packet, it will replace the destination address with the address in the routing header. Then the MN discovers that the destination now is its home address and passes the packet on to the transport layer. Using routing header instead of encapsulation can reduce overhead.

Handover management:

MIPv6 specifies that an MN can use any combination of mechanisms to detect its movement to another network. Two possibilities are Eager Cell Switching (ECS) handover initiation strategy and the Lazy Cell Switching (LCS) handover initiation strategy [15]. Using LCS, an MN will not change its current serving AR until it fails to receive another RA from its current AR within the specified lifetime. Using ECS, an MN switches immediately to a new AR upon receiving an RA from that AR. ECS assumes that mobile nodes follow steady trajectories while they move across a wireless network. Fig 2.2 shows the MIPv6 handover procedure.

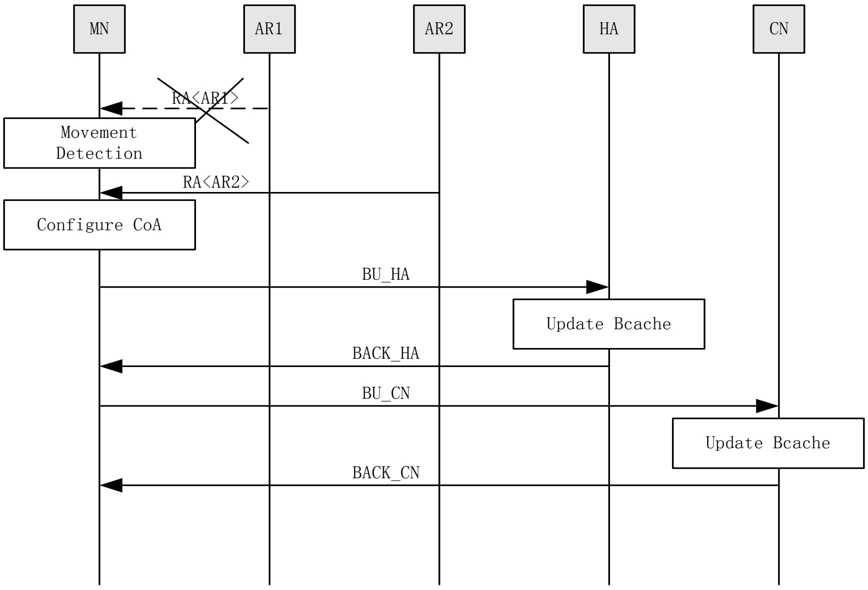


FIG 2.2 MIPV6 HANDOVER PROCEDURE

2.1.3 Hierarchical Mobile IPv6 (HMIPv6)

HMIPv6 is an extension of the basic MIPv6 presented in [10]. In HMIPv6, an MN has two CoAs:

- Regional CoA (RCoA): an address obtained by the MN from the visited domain
- Local CoA (LCoA): an on-link CoA configured on an MN’s interface based on the prefix advertised by ARs.

Location management

The two CoAs are used to handle global mobility and local mobility respectively. To manage local mobility, a new entity called Mobile Anchor Point (MAP) is introduced. The existence of a domain MAP is advertised by ARs as a new MAP option in the Router Advertisement (RA) message. The MAP option includes the distance vector, the MAP’s global IP address and the MAP’s subnet prefix. Upon reception of an RA message, an MN can configure its RCoA and LCoA by using MAP prefix and AR prefix. An MN registers its LCoA with the MAP and registers its RCoA with the HA and CNs. When an MN moves within a domain, it does not need to re-register its RCoA with its HA and CNs. Two modes of HMIPv6 are provided. One is basic mode: an MN forms its own unique RCoA on the MAP’s subnet. The other is extended mode: an MN is configured with an RCoA that is assigned to one of the MAP’s interfaces. The network topology of HMIPv6 is shown in Fig 2.3.

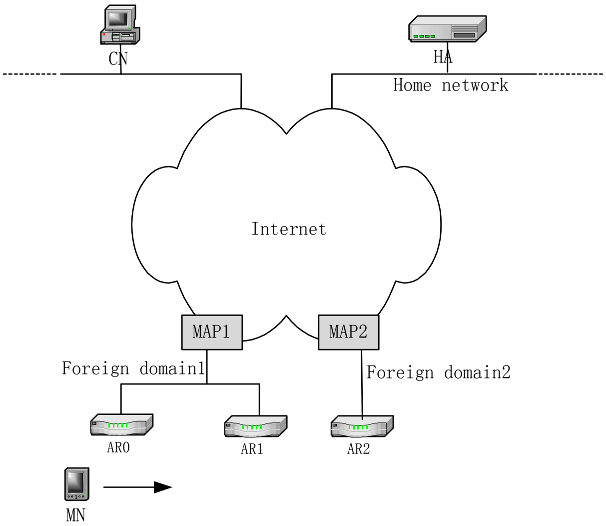


FIG 2.3 NETWORK TOPOLOGY (HMIPV6)

Handover management

The mobility of an MN can be classified into global mobility and local mobility.

- Global mobility: When an MN moves from one MAP domain to another MAP domain (E.g., the MN moves from AR1 to AR2 in Fig 2.3). The handover procedure is illustrated in Fig 2.4.

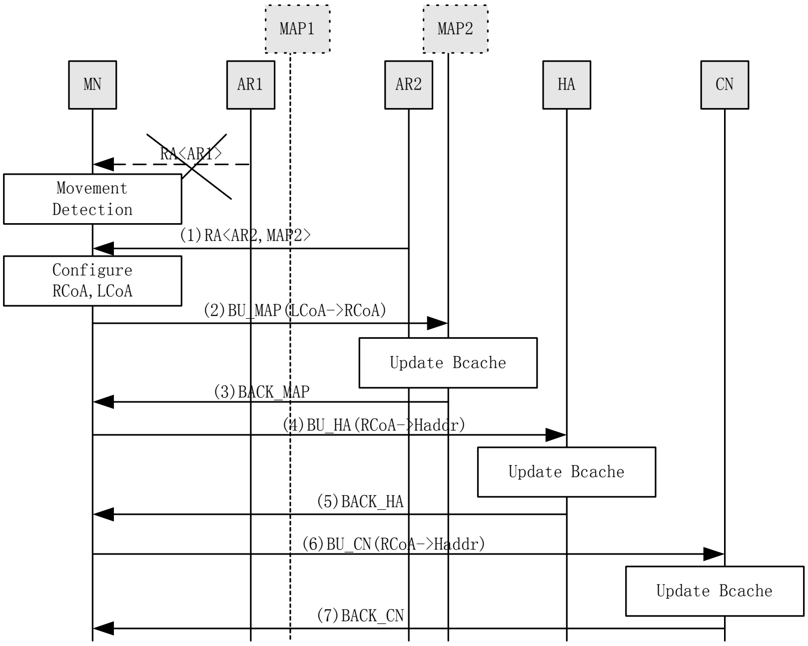


Fig 2.4 HMIPv6 Handover procedure (Global mobility)

(1) An MN detects its arrival to a new domain and receives RA from AR2. The MN configure its RCoA and LCoA

(2) The MN sends Binding Update (BU) which specify the binding between its RCoA and LCoA to the domain MAP

(3) Upon reception of BU, the MAP performs admission control. If the request is accepted, the MAP update its binding cache (Bcache) and sends Binding Acknowledgement (BACK_MAP) back to the MN

(4) The MN sends BU which specify the binding between its Home address (Haddr) and RCoA to its HA

(5) Upon reception of BU, the HA update its binding cache (Bcache) and sends acknowledgement (BACK_HA) back to the MN

(6)~(7) are similar with (4)~(5); the only difference is that in (6)~(7), the BU and BACK is exchanged between the MN and its CNs.

- Local mobility: When an MN moves from an old AR to a new AR within the same MAP domain (e.g., The MN moves from AR0 to AR1 in Fig 2.4). The handover procedure is illustrated in Fig 2.5.

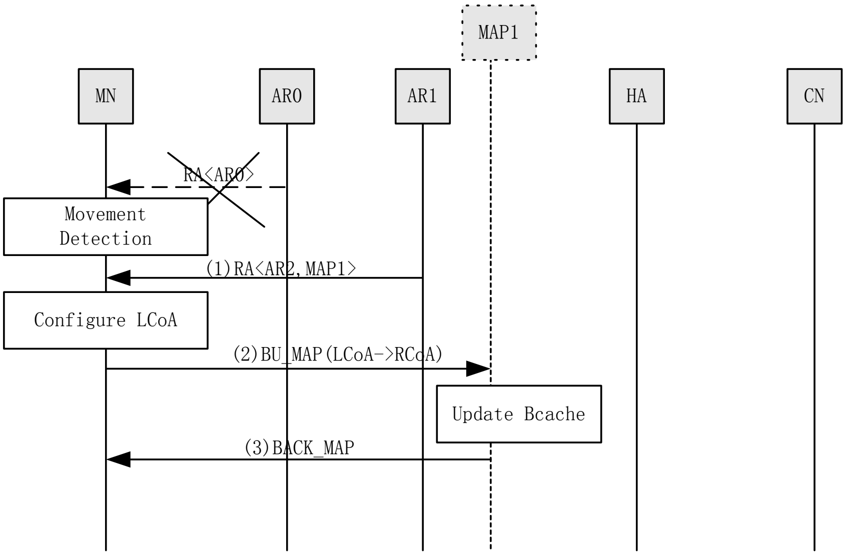


FIG 2.5 HMIPV6 HANDOVER PROCEDURE (LOCAL MOBILITY)

(1) An MN receives RA from AR2, and from the MAP option included in RA, the MN finds that it is still in the same MAP's domain; hence the RCoA is not changed. The MN configure its LCoA

(2) The MN sends BU message that specify the binding between its RCoA and its new LCoA to MAP

(3) Upon reception of BU, the MAP updates its binding cache (Bcache) and sends acknowledgement (BACK_MAP) back to the MN

In the case that an MN is moving in a foreign domain which is far away from its

HA and CNs, HMIPv6 can significantly reduce signaling overhead and also reduce handover latency because signaling messages travel only up to the MAP for local handover.

2.1.4 Fast Handover for Mobile IPv6 (FMIPv6)

FMIPv6 [11] reduces packet loss by providing fast IP connectivity as soon as a new link is established. It achieves this by setting up the routing during link configuration and binding update, so that packets delivered to the old CoA are forwarded to the new subnet while the MN is still attached to the old subnet. This reduces the amount of preconfiguration time in the new subnet. Fig 2.6 shows the network topology of FMIPv6.

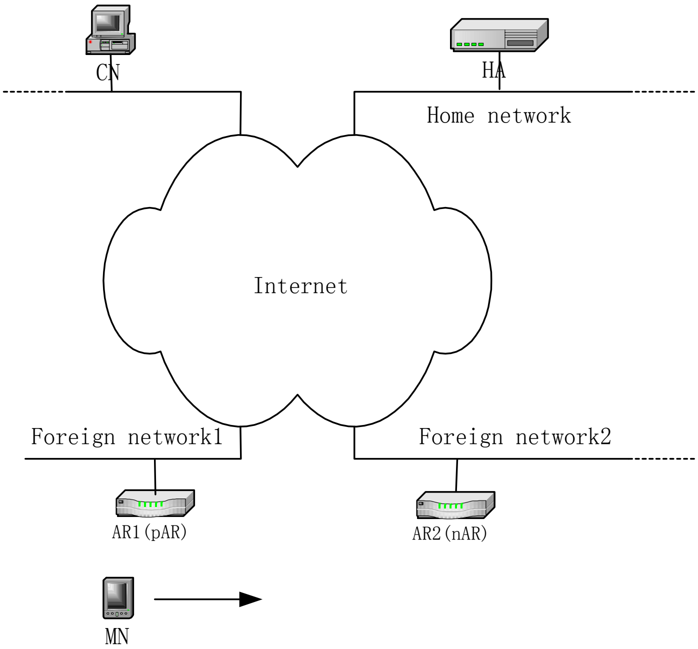


FIG 2.6 NETWORK TOPOLOGY (FMIPV6)

FMIPv6 messages:

(1) Router Solicitation for Proxy Advertisement (RtSolPr): a message from the MN to the previous AR (PAR) to request information for a potential

handover.

(2) Proxy Router Advertisement (PrRtAdv): a message from the PAR to the MN that aids in movement detection.

(3) Fast Binding Update (FBU): a message from the MN instructing its PAR to redirect its traffic towards the new AR (NAR).

(4) Handover Initiate (HI): a message from the PAR to the NAR to initiate handover.

(5) Handover Acknowledgement (Hack): a message from the NAR to the PAR as a response to HI.

(6) Fast Binding Acknowledgement (FBack): a message from the PAR in response to FBU.

(7) Fast Neighbor Advertisement (FNA): a message from the MN to the NAR to announce attachment and to confirm use of NCoA if the MN has not received FBack from PAR's link.

FMIPv6 operation:

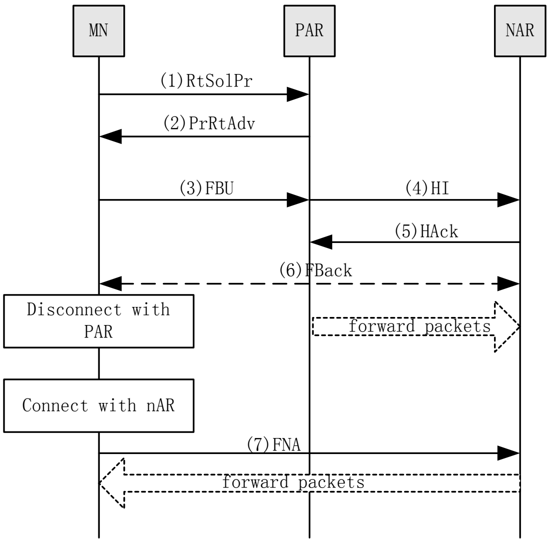


FIG 2.7 PREDICTIVE MODE (FBU IS SENT FROM PAR'S LINK)

The protocol discussion is under the assumption that an MN is moving to a different subnet. FMIPv6 protocol begins when an MN sends RtSolPr which contains NAR's link layer address to its current AR (PAR) to resolve NAR's information. In response, PAR sends a PrRtAdv message which contains the NAR's IP address. The MN configures a new Care-of-address and sends a FBU to PAR, which makes PAR bind the previous Care-of-address (PCoA) to the new Care-of-address (NCoA), so that subsequent packets arriving at PAR can be tunneled to NAR. The FBU may be sent from PAR's link (as in Fig 2.7) or from a NAR's link (as illustrated in Fig 2.8). The former case is called "predictive mode" and the latter case is called "reactive mode". In predictive mode, PAR will communicate with NAR by HI/Hack exchange to validate the NCoA, and sends an FBack to the MN. If the MN fails to receive FBack on the previous link, the circumstances may be that the MN has not sent FBU or the MN has left the link after sending the FBU. In any case, the MN should send an FBU as soon as it attaches to NAR (as illustrated in Fig 2.8).

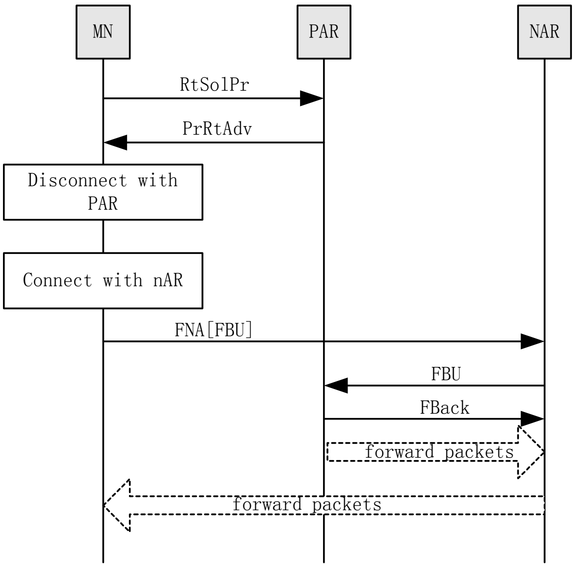


FIG 2.8 REACTIVE MODE (FBU IS SENT FROM NAR'S LINK)

In order to verify NCoA and to enable NAR forward packets, the MN encapsulates FBU in FNA. After processing FNA, the NAR deliver the FBU to PAR, and then a tunnel from PAR and NAR is constructed. At this point, the MN has accomplished the IP connection with the new access point and can resume communicating with CN through the tunnel between PAR and NAR. To make CN send packet directly to the NAR, the MN should perform the normal MIPv6 process of sending BUs to CN. The trick of FMIPv6 is that since a bidirectional tunnel has been constructed to forward packets, the BU relay latency will not disrupt the communication.

2.1.5 Macro/Micro-mobility management in the Internet

The concept of Macro/Micro-mobility management emerges due to the drawback of Mobile IP that every movement of an MN to a new point of attachment requires the registration with its HA. When the HA is remote from the MN’s foreign network, it will introduce much signaling overhead as well as large handover delay. Consequently, micro-mobility protocols are proposed to address the movement in a relative smaller area.

Existing proposals for micro-mobility management can be broadly classified into two types: routing-based and tunnel-based schemes.

- Routing-based schemes: A distributed mobile host location database is created and maintained by all the mobility agents within the network domain. There is a domain root router to handle all inbound and outbound mobile traffic. These schemes are exemplified by the Cellular IP [12] and HAWAII [13] protocols, which differ from each other in the functionality of the nodes and the construction methods of the lookup tables.

- Tunnel-based schemes: In hierarchical tunneling approaches the location database is maintained in a distributed form by a set of Foreign Agents (FA) constructed in a tree structure in the access network, e.g., Regional Registration [14], HMIPv6 [10]. In Regional Registration, encapsulated traffic from the home agent is delivered to the Gateway Foreign Agent (GFA). Each FA on the tree decapsulates and then re-encapsulates data packets as they are forwarded down the tree of FAs towards the mobile host's point of attachment. When a mobile host moves between different ARs, location updates are made at the optimal point on the tree.

Both routing-based and tunnel-based schemes have their advantages and disadvantages. The routing-based schemes can avoid tunneling overhead, but they may suffer from difficulty in scaling because each registered MN will have an entry recorded at each router on the uplink path from the AR to the root router. Furthermore, the root router in the domain has the vulnerability of a single point of failure. On the contrary, although tunnel-based schemes may introduce tunneling overhead, they are possible to designate multiple GFAs or MAPs within the micro-mobility domain, thus achieving higher robustness. Combined with label switching technology (e.g., MPLS), the tunneling overhead can be greatly reduced and thus the tunneling-based scheme seems to be a preferred solution for supporting micro-mobility in wireless networks [16].

2.2 MPLS

Multi-protocol Label Switching (MPLS) [9] is an Internet Engineering Task Force (IETF) specified framework that provides for the efficient designation, routing, forwarding, and switching of traffic flows through the network. In MPLS, data transmission occurs on Label Switched Paths (LSP). LSP is a sequence of labels at each node along the path from the source to the destination. LSPs are established either prior to data transmission (control-driven) or upon detection of a certain flow of data (data-driven). The labels are distributed using Label Distribution Protocol (LDP) or piggybacked on routing protocols like Border Gateway Protocol (BGP) and Open Shortest Path First (OSPF). Each data packet encapsulates and carries the labels during their journey from source to destination. High-speed switching of data is possible because the fixed-length labels are inserted at the header of packets and can be used by hardware to switch packets quickly between links. The devices that participate in MPLS can be classified into Label Edge Router (LER) and Label Switching Router (LSR). An LSR is a device in the core of an MPLS network that participates in the establishment of LSPs using the appropriate label signaling protocol and high speed switching of the data traffic based on the established paths. An LER is a device that operates at the edge of the access network and MPLS network. LERs supports multiple ports connected to dissimilar networks (such as ATM, Ethernet, and frame relay) and forwards this traffic on to the MPLS network after establishing LSPs, using the label signaling protocol at the ingress and distributing the traffic back to the access networks at the egress. The LER plays a very important role in the assignment and removal of labels. The Forward Equivalence Class (FEC) is a representation of a group of packets that share the same requirements for their transport. In MPLS, the assignment of a particular packet to a particular FEC is done just once, as the packet enters the network. FECs are based on service requirements for a given set of packets or simply for an address prefix. Each LSR builds a table to specify how a packet must be forwarded. This table, called a Label Information Base (LIB), is comprised of FEC-label bindings. A unique feature of MPLS is that it can control the entire path of a packet without explicitly specifying the intermediate routers. It does this by creating tunnels through the intermediary routers that can span multiple segments.

2.3 MANET

2.3.1 Overview

Mobile Ad hoc network (MANET) is a type of mobile wireless networks. In contrast to an infrastructure wireless network, a MANET is an infrastructure-less network. In a MANET, there is no fixed router and each MN can serve as a router that discovers and maintains routes to other nodes. The MANET concept applies to situations such as emergency rescue operations and data sharing in a conference. To support the routing in the networks, many protocols have been proposed in recent years. The MANET routing protocols can generally be categorized as table-driven routing protocols and on-demand routing protocols [31]. In the following subsections, we review some popular MANET routing protocols in both categories.

2.3.2 Table-driven Routing protocols

Table-driven routing protocols build routes in a proactive way between nodes in a MANET. Routing information is periodically disseminated among all the nodes in the network; therefore, every node has the up-to-date information for all possible routes. As an example of table-driven routing protocol, we introduce one famous routing protocol: DSDV.

Destination-Sequenced Distance-Vector (DSDV)

Destination-Sequenced Distance-Vector (DSDV) routing is based on the classical Bellman-Ford routing scheme. DSDV, unlike traditional distance vector protocols, guarantees loop-freedom by tagging each route table entry with a sequence number to order the routing information. Each node maintains a routing table with all available destinations along with information like next hop, the number of hops to reach the destination, sequence number of the destination, etc. DSDV uses both periodic and triggered routing updates to maintain table consistency. Triggered routing updates are used when network topology changes are detected, so that routing information is propagated as quickly as possible. Mobile nodes cause broken links when they move from place to place. When a link to the next hop is broken, any route through that next hop is immediately assigned infinity metric and an updated sequence number. This is the only situation when any mobile node other than the destination node assigns the sequence number. Sequence numbers assigned by the origination nodes are even numbers, and sequence numbers assigned to indicate infinity metrics are odd numbers. When a node receives infinity metric, and it has an equal or later sequence number with a finite metric, it triggers a route update broadcast, and the route with infinity metric will be quickly replaced by the new route. When a mobile node receives a new route update packet, it compares it to the information already available in the table and the table is updated based on the flowing criteria:

- If the received sequence number is greater, then the information in the table is replaced with the information in the update packet.
- Otherwise, the table is updated if the sequence numbers are the same and the metric in the update packet is better.

DSDV requires nodes to periodically transmit routing update packets. These update packets are broadcast throughout the network. When the number of nodes in the network grows, the size of the routing tables and the bandwidth required to update them also grows, which could cause excessive communication overhead. This overhead is nearly constant with respect to mobility rate.

2.3.3 On-demand Routing protocols

On-demand routing protocols discover routes only as needed. When a node wishes to communicate with another node, it checks with its existing information for a valid route to the destination. If one exists, the node uses that route for a valid route to the destination. If one exists, the node uses that route for communication with the destination node. If not, the source node initiates a route request procedure, to which either the destination node or one of the intermediate nodes sends a reply back to the source node with a valid route. A soft state is maintained for each of these routes- if the routes are not used for some period of time, the routes are considered to be no longer needed and are removed from the routing table; if a route is used before it expires, and then the lifetime of the route is extended. Compared with table-driven routing protocols, on-demand routing protocols may have lower computation costs and lower packet overhead since they do not need to exchange routing information periodically and maintain route tables. However, the on-demand feature results in longer packet transfer delay. In the following, we introduce a well-known on-demand routing protocol Ad hoc On-demand Distance Vector (AODV) [30]. One of the reasons to why AODV has been used in this study is that it is one of the most developed routing protocols for MANET.

Ad-hoc On-Demand Distance Vector Routing (AODV)

AODV is essentially a combination of both DSR and DSDV. It borrows the conception of sequence numbers from DSDV, plus the use of the on-demand mechanism of route discovery and route maintenance from DSR. When a source node needs to send a packet to a destination node for which it has no routing information in its table, the Route Discovery process is initiated. The source node broadcasts a route request (RREQ) to its neighbors. Each node that forwards the RREQ packet creates a reverse route for itself back to source node. Every node maintains two separate counters: a node sequence number and a broadcast id. Broadcast id is incremented when the source issues a new RREQ. Together with the source's address, it uniquely identifies a RREQ. In addition to the source node's IP address, current sequence number and broadcast id, the RREQ also contains the most recent sequence number for the destination which the source node is aware of. A node receiving RREQ may unicast a route reply (RREP) to the source if it is either the destination or it has a fresh enough route to the destination, namely, it has a route to the destination with corresponding sequence number greater than or equal to that contained in the RREQ. Otherwise, it re-broadcasts the RREQ. Each node that participates in forwarding a RREP packet back to the source of RREQ creates a forwarding route to the source node. As the RREP packet back to the source, nodes set up forward pointers to the destination. Once the source node receives the RREP, it may begin to forward data packets to the destination. At any time a node receives a RREP (for any existing destination in its routing table) containing a greater sequence number or the same sequence number with a smaller hop count, it may update its routing information for that destination and begin using the better route. Routes are maintained as follows: If an upstream node in an active route senses a break in the active route, it can reinitiate the route discovery procedure to establish a new route to the destination (local route repair) or it can propagate an unsolicited RERR with a fresh sequence number and infinity hop count to all active downstream neighbors. Those nodes subsequently relay that message to their active neighbors. This process continues until all active source nodes are notified. Upon receiving notification of a broken link, source nodes can restart the discovery process if they still require the destination. Link failure can be detected by using HELLO messages or by using link-layer acknowledgements.

There are a couple of important distinctions between DSR and AODV. The most notable distinction is that the AODV is a kind of hop-by-hop routing protocol in contrast to the source routing in DSR. During the process of forwarding the RREQ, intermediate nodes record in their route tables the address of the neighbor from which the first copy of the RREQ is received, thereby building a reverse route. If an intermediate node knows a fresh route to the destination, it unicasts a RREP to the neighbor from which it receives the RREQ. While the RREP is routed back along the reverse route, each node on the route builds a forward route entry to the destination according to the source address contained in the RREP. The different routing type makes the overhead of AODV smaller than that of DSR since each DSR packet contains full route information, whereas in AODV packets only contain the destination address. Also, the RREP in

AODV is smaller than the route reply message in DSR since the RREP only needs to carry the destination address and sequence number. AODV is capable of both unicast and multicast routing. It maintains these routes as long as they are needed by the sources. Additionally, AODV forms trees that connect multicast group members. The trees are composed of the group members and the nodes needed to connect the members. The major drawback of AODV is that it requires bidirectional links between nodes since the RREP is forwarded along the path established by the RREQ.

2.4 Summary

In this chapter we introduced mobility management, MPLS, MANET, and described Mobile IPv6 as well as its two extension protocols in detail. These concepts and protocols are the important components of the mobility management in wireless networks that we will study in the following chapters. In the next chapter, we will present a mobility management scheme in MPLS-based Hierarchical Mobile IPv6 network.

CHAPTER 3　MOBILITY MANAGEMENT IN IP/MPLS BASED HMIPv6 NETWORKS

This chapter presents a mobility management scheme in MPLS-based Hierarchical Mobile IPv6 (HMIPv6) network. The proposed scheme takes advantage of HMIPv6 [10] to localize registration in one domain, Multiprotocol Label Switching (MPLS) [9] under IP layer to provide fast packet forwarding, and uses Layer 2 (L2) information to anticipate handover to reduce handover latency. This scheme gives a fast and smooth handover to support real-time applications. To further reduce packet loss during handover, we also consider using Bicasting, which will be introduced later in this chapter. The simulation results and analysis are presented in Chapter 5.

3.1 Introduction

The next generation networks are expected to provide global mobility support to potentially a large number of mobile nodes (MNs) and to accommodate various kinds of services including voice, data, as well as real-time traffic with stringent performance bounds. With the “all-IP network” trend and QoS requirements, the combination of Mobile IPv6 (MIPv6) [6] and Multiprotocol Label Switching (MPLS) [9] is seen as a promising solution for the next generation networks.

As described in Chapter 2, Hierarchical MIPv6 (HMIPv6) [10] and Fast Handover for Mobile IPv6 (FMIPv6) [11] are two proposals to enhance the performance of MIPv6. The HMIPv6 minimizes the amount of signaling to the HA and correspondent nodes by allowing MNs to locally register in an administrative domain. FMIPv6 protocol provides anticipation by using Layer 2 (L2) trigger to initiate handover operation and thus MNs can recover traffic immediately upon arriving at the new AR. According to the tests performed in [16], the L2 handover could take a long time, especially if there are several active MNs. The traditional handover, including Layer 3 (L3) handover that begins after the completion of the L2 handover, will take even more time that is unacceptable to real-time applications. A natural idea is combining the advantages of HMIPv6, MPLS, and FMIPv6 to obtain a better performance of handover.

The rest of this chapter is organized as follows. Section 3.2 presents an overview of related works. Section 3.3 illustrates the detail of registration procedure and intra-MAP handover mechanism. The extensions to Network Simulator 2 (NS2) and the simulation model, followed by the performance analysis based on the simulation results are presented in Chapter 5.

3.2 Related Works

For mobility management in IP/MPLS network, there have been some works done in [17][18][19]. However, these works are all based on MIPv4. Taking into consideration the presence of IPv6 in future networks and the advantages of MIPv6 over MIPv4, a scheme based on MIPv6 is worth studying. Moreover, the existing works do not take advantage of using L2 or link layer information.

Conventional MPLS does not support mobility. By incorporating Mobile IP with MPLS, a scheme to support mobility in MPLS networks is given in [17]. A Label Switch Path (LSP) from a HA to a FA is established during the registration process, which uses an MN's CoA as the Forwarding Equivalence Class (FEC). However, the integration of Mobile IP and MPLS suffers the same inefficiency as in pure Mobile IP that lacks micro-mobility support.

The Hierarchical Mobile MPLS (H-MPLS) [18] is proposed to improve the Mobile MPLS [17] that is able to handle movement of MNs locally. This is achieved by introducing Foreign Domain Agent (FDA) into each MPLS domain. Thus, no location update messages need to be sent to the remote HA when an MN moves within the same MPLS domain. The drawback of H-MPLS is its rigid hierarchy of mobile agents. The flexible hierarchy structure of HMIPv6 can be a solution to address this problem.

In [19], path rerouting during handover is proposed. The crossover mobility agent in the foreign agent hierarchy is an optimal point to perform a rerouting upon handover, which can reduce the registration latency. However, the paper does not explain how to identify the crossover agent and lacks simulation results.

3.3 Scheme Overview

A simplified network topology is shown in Fig 3.1. The MAP (Mobile Anchor Point) and the ARs at the edge of the MPLS network are called Label Edge Routers (LERs). Several ARs are connected to intermediate LSRs, which in turn are interconnected to MAP. MAP provides access to outer networks, and communicates with CNs and HAs. To ensure that the registration request and other handover related messages are transmitted efficiently, the uplink Label Switching Paths (LSPs) from every AR to MAP are pre-established.

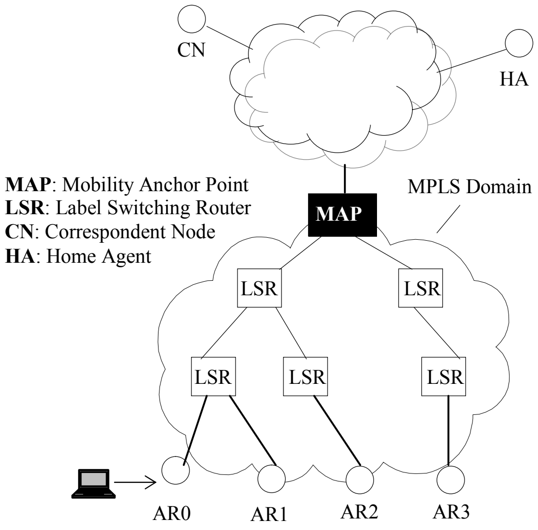


FIG 3.1 NETWORK TOPOLOGY (IP/MPLS BASED HMIPV6 NETWORK)

3.3.1 Registration

The registration process illustrated in Fig 3.2 is similar with that in HMIPv6 as described in Section 2.1.3.

(1) Upon completion of the link layer attachment, an MN receives an RA message from an AR and then auto-configures RCoA and LCoA based on the information contained in the RA message.

(2) The MN sends registration message “BU_MAP” to the MAP through the selected AR. The AR then forwards BU_MAP message to the MAP via the pre-established LSP.

(3) When the MAP receives the registration message from the lower-level LSR,

it adds its record about that MN. After that, MAP sends mobility binding message “BU_HA” to the MN’s HA with the MN’s home IP address and RCoA.

(4) When the HA receives BU_HA message, it updates the MN’s entry in its binding cache and sends binding acknowledge message “BACK_HA” to the MAP.

(5) When the BACK_HA message is received at the MAP, this MAP sends registration reply message “BACK_MAP” back to the MN.

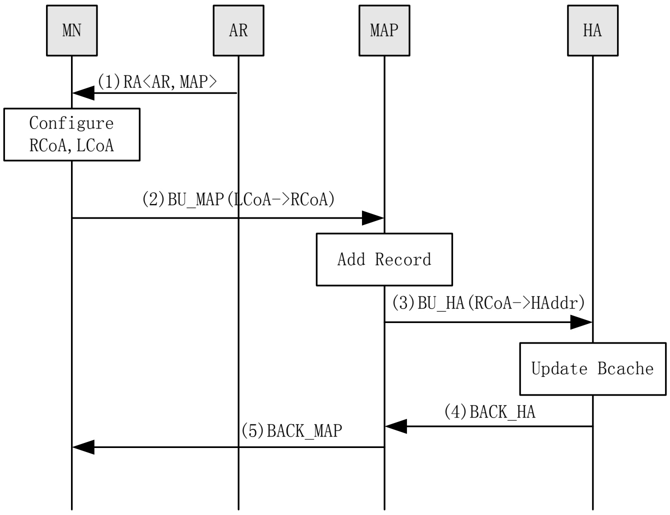


FIG 3.2 REGISTRATION PROCESS IN IP/MPLS BASED HMIPV6 NETWORK

When data traffic is initiated from or to the MN, new LSP tunnels are set up with a bandwidth reservation between a MAP and an AR. Considering downlink traffic and the use of LDP (Label Distribution Protocol), the MAP initiates the setup by sending LDP Request message downlink to the AR; and then AR sends LDP Mapping message back to MAP.

3.3.2 Intra-MAP handover mechanism

In this section, we discuss intra-MAP handover, where both new AR and old

AR are under the same MAP. Considering micro-mobility, this kind of handover is more frequent than inter-MAP handover. The intra-MAP handover procedure in our scheme is shown in Fig 3.3. The L2 message is assumed to contain enough information for the MN to create new CoAs and trigger the MN to make a handover decision.

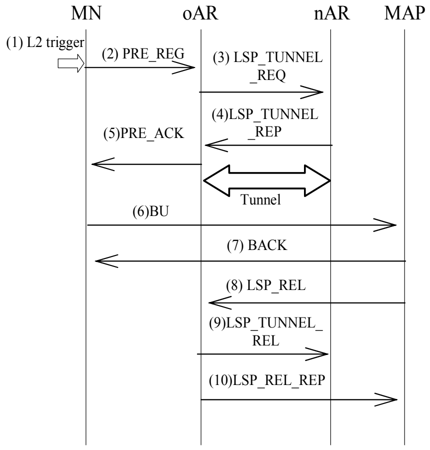


FIG 3.3 INTRA_MAP HANDOVER

The handover mechanism is described as follows:

(1) Before an MN detaches its current link, the MN receives a L2 message, which provides the information of the new subnet prefix. The MN may use stateless address auto-configuration to form a new CoA.

(2) After the MN forms its new CoA, it sends a PRE_REG (pre-registration) message, which includes the IP address of the new AR, the MN’s new CoA, and the MN’s old CoA to its old AR.

(3) Upon receiving the PRE_REG message, the old AR sends an LSP_TUNNEL_REQ (LSP tunnel request) message to the new AR.

(4) When the new AR receives LSP_TUNNEL_REQ message, it adds an entry for the MN that binds the MN’s old CoA with the new CoA and sends back the LSP_TUNNEL_REP (LSP tunnel reply) message to the old AR.

(5) When the old AR receives the LSP_TUNNEL_REP message, a bi-direction LSP tunnel between old AR and new AR has been established. The old AR uses bicasting for a limited period during which it keeps sending packets both to the MN's old CoA and through the LSP tunnel to the new CoA. Then, it sends a PRE_ACK (pre-registration acknowledge) message to the MN.

(6) When arriving at the new AR, the MN sends a HMIPv6 BU (binding update) message to the MAP.

(7) When the MAP receives the BU from the MN, it sends a BACK (binding acknowledgement) message to the MN. By this time, the L3 handover is completed.

(8) The MAP sends a LSP_REL (LSP release) message to the old AR.

(9) The old AR sends a LSP_TUNNEL_REL (LSP tunnel release) message to the new AR.

(10) The old AR sends a LSP_REL_REP (LSP release reply) message to the MAP.

3.3.3 Approaches to achieve seamless handover

In last section, we presented the intra-MAP handover procedure. In this section, we discuss how our scheme can achieve seamless handover that reduces handover latency and packet loss.

In general, IP handover latency can be divided into three parts: movement detection, Care-of-address (CoA) configuration, and binding update propagation. In MIPv6 [6], the movement detection algorithm relies on the periodic Router Advertisements (RA) from ARs to enable MNs determining their current locations.

Consequently, to achieve optimum detection performance, RAs can be broadcast at a faster rate, which however results in overhead on wireless links. In our intra-MAP handover scheme described in Section 3.3.2, by using L2 message that provides network layer information, the movement to a new access point can trigger the initiation of L3 handover. Therefore MNs don’t need to wait for RA messages or send solicitations to discover movement, and thus the movement detection delay is reduced. The L2 message can be implemented using the method proposed in [22], in which we show how to extend the IEEE 802.11 management frames to carry extensible application specific information elements. The extended messages allow access points to advertise the capabilities information of its associated network and to improve movement detection.

Since the CoA configuration is an orthogonal issue with the work in this thesis, we neglect this part of delay. Binding update propagation latency is reduced by using HMIPv6 that localizes the mobility update in a MAP’s domain.

To reduce handover latency and packet loss, FMIPv6 [11] proposes a bi-directional tunnel between old AR and new AR before the start of L3 handover. Using MPLS, a tunnel can be built easily by implementing an explicit LSP. Although the FMIPv6 protocol may greatly reduce the handover latency, it requires a synchronization of the redirection of the packets and the actual movement of the MN; otherwise, some packets that are in transit may be lost during handover process. Specifically, if the packets are redirected too early, they arrive at the new link but the MN is not there yet; if the redirection is too late, the

MN may arrive at the new link but packets are still being routed to the old link. To address this problem, we consider using the bicasting mechanism [16] as an extension to the bidirectional tunnel between old AR and new AR. The original Bicasting allows an MN to simultaneously register with several ARs; all the packet intended for the MN are duplicated from HAs and forward to several potential locations. However, the bicasting done by the HA is not scalable and generates lots of traffic on both the wired and wireless links. In a hierarchical network, the bicasting can be localized in a domain. We propose to perform bicasting at the old AR during an MN's movement. The old AR will continuously receive packets destined to the MN's old care-of address (CoA) until the L3 handover is completed. When the old AR has constructed the tunnel between itself and the new AR, it bicasts all the packets received from the MAP to the MN's old CoA and through the tunnel to the new AR during a limited period. Hence, the MN can receive packets from the old AR or the new AR depending on with which it is attached. The bicasting mechanism can also be a good solution to the ping-pong phenomenon.

3.4 Summary

In this chapter, we discussed mobility management in IP/MPLS based HMIPv6 network. The registration and data LSP construction process was discussed in section 3.3.1. In section 3.3.2, we proposed a seamless intra-MAP handover scheme to support real-time applications for MNs. A L2 message contains network layer information of the new subnet is used to trigger an MN to initiate L3 handover; therefore movement detection is performed faster. To reduce packet loss, temporary bicasting is used at the old AR. We will examine the effect of L2 trigger and bicasting on handover performance, taking into consideration the different overlaps between ARs and different Router Advertisement intervals in Chapter 5.

CHAPTER 4 MOBILITY MANAGEMENT IN HYBRID NETWORKS

This chapter proposes a mobility management scheme to integrate MANET (Mobile Ad-hoc Network) and the Internet. Hierarchical Mobile IPv6 (HMIPv6) [10] and Ad hoc On Demand Distance Vector (AODV) routing protocol [30] are chosen to be the two fundamental protocols in our scheme. We will show how the two protocols can be integrated efficiently to extend mobility management to mobile nodes that are multiple hops away from the Gateways (GW). We define and discuss comprehensively the handover issue in hybrid network. The highlight of our scheme is that it can provide smooth multi-hop handover without incurring too much signaling overhead. We will show the performance of our handover schemes in chapter 5 (Section 5.3).

4.1 Introduction

As stated in Chapter 1, in future all-IP networks, the Internet will be accessible from different kinds of wireless networks [1][2][3]. MANET is an important class of mobile wireless networks that is infrastructure-less with the advantage of auto-configurability, which makes it very promising to be widely used in the future. The integration of a MANET and the Internet is referred to as a hybrid network, as show in Fig 4.1. A GW between the Internet and the MANET is required. This GW functions as Access Router (AR) for MNs and understands protocols in both networks.

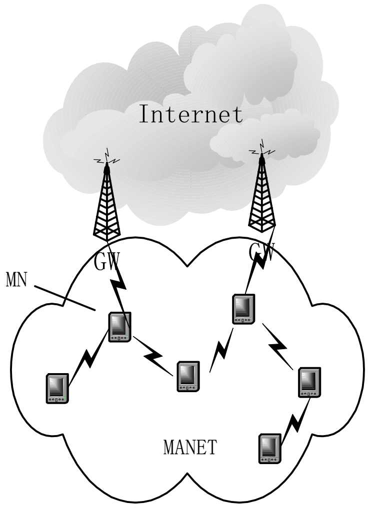


FIG 4.1 NETWORK TOPOLOGY (HYBRID NETWORK)

Since mobility management is the key mechanism to ensure communication between MNs and wired networks, it is worthwhile to study mobility management in the hybrid network. The most well-known mobility management scheme in the Internet is Mobile IP (MIP); however, the standard MIP only concerns those mobile nodes within GWs' transmission range. In hybrid networks, some MNs are multiple hops away from GWs. A promising idea is to extend MIP into MANET. However, there exist some challenges:

1) Challenges in Location management

As described in Section 2.1, location management aims to inform related mobility agents (e.g., HA, MAP) in network about the current locations of MNs and then the packets can be delivered to the desired MN. Therefore, the first step is to make MNs know the existence of GWs in a hybrid network.

1a) How to discover GWs?

To communicate with hosts in the Internet, an MN must find a GW and all the traffic needs to be relayed through the GW to reach the Internet host. In MIP, ARs periodically broadcast Router Advertisements (RA) to inform MNs of their existence. To make MNs that are multiple hops away know the existence of GWs, a method is to flood the RAs in the MANET (proactive approach). However, any flooding in a wireless network is undesirable since it consumes much bandwidth, which is already scarce. Another way is for the GW to unicast an RA to the requested MN (reactive approach). Although the reactive approach seems to reduce overhead, it introduces more delay because an MN must wait for the GW's reply; and it may incur more overhead than in proactive approach when the number of MNs requesting GWs increases. To combine the advantages of both the proactive and reactive approaches, a possible approach is to limit the RA flooding range. Hence, MNs inside of that range can receive periodically broadcasted RAs while MNs out of that range solicit for RAs (hybrid approach). Since the RA flooding range is an important parameter to tradeoff between signaling overhead in wireless network and MNs' connectivity with the Internet, the subsequent question is: how to set the flooding range? A possible way is to set the flooding range dynamically according to the current conditions of the network, e.g., node density distribution in the wireless network, the number of MNs that require Internet connectivity, mobility pattern of MNs, etc.

1b) How to efficiently integrate Mobile IP and MANET routing protocol?

Mobile IP is designed for infrastructure-based networks in which the AR is a special node in the network that helps MNs to detect movement and maintain Internet connectivity. However, a MANET is an infrastructure-less network in which every node is equal. In hybrid networks, a GW performs both the AR's function in Mobile IP and also acts as a member of the MANET. Simply adding Mobile IP on top of a MANET routing protocol is not efficient because: (1) Considering on-demand routing protocols, to discover a GW, an MN either waits for a flooded RA message or sends solicitation. After receiving an RA message, the MN has to initiate route discovery to find the route to the GW before it starts communication with the Internet. If the GW discovery phase is integrated with the GW route discovery, control overhead as well as route discovery delay can be reduced. One possible solution is giving RA and solicitation messages the ability to construct route, so that when an MN receives an RA message, it also gets the route to that GW. (2) In normal Mobile IP, the MN will solicit for RAs when the current GW entry expires. In a hybrid network, it is possible that the route to a GW breaks while the GW entry in Mobile IP is still valid, which delays the MN from finding a new GW. This problem can be addressed by making the routing protocol inform Mobile IP upon route failure to the current GW.

In summary, to efficiently address the issues in location management, the GW discovery process and route discovery process should be integrated to reduce control overhead and route discovery delay. To make an MN detect the loss of its current GW faster, the MANET routing protocol should inform Mobile IP upon route failure to the current GW, which can be achieved by distinguish routes for

GWs in MANET routing table. The above analysis concerns more for on-demand routing protocols. Considering proactive routing protocols (e.g. DSDV), each MN maintains the routes to all other MNs by periodically exchanging routing information with neighboring MNs. Since in hybrid networks a GW is a member of the MANET, the proactive routing protocol can also provide MNs with updated routes to all GWs. When the route to one GW breaks, an MN can easily use the route to another GW as long as it can distinguish the routes to GWs from the routes to other MNs. Since there exists more challenges for using on-demand routing protocol in hybrid network, the discussion later will focus on on-demand routing protocols.

2) Challenges in handover management

Handover management aims to ensure a continuous session for an MN moving across service domains. Considering multiple GWs located in hybrid networks, making a wise decision on selecting a GW will be beneficial to an MN's quality of service.

2a) How does an MN select GWs?

There are various possible criteria for an MN to choose a GW, e.g., distance from the MN to the GW, the Round Trip Time (RTT) between the MN and the GW, traffic load at the GW, etc. A simple criterion is the hop count from an MN to a GW since it is provided in the MN's routing table. The number of hops to a GW is closely related to the traffic delay as well as the throughput experienced by the

MN; hence, to choose a GW that is fewer hops away is reasonable. However, using only hop count to choose GWs is not enough when the traffic load is very high in network, e.g., if one GW is chosen by many MNs, its performance will degrade; in this case, load balancing between multiple GWs is important. One possible solution is to provide load information at GWs in RA messages, so that an MN can make a decision that prevents overloading one GW. Another possible solution is to make multiple GWs exchange load information with each other and guide an MN to choose the appropriate GW.

2b) When should an MN handover to another GW?

When the route to a GW breaks, the MN can choose to recover route to the lost GW, use an existing valid route to another GW, or discover a new route to other possible GWs. Ideally, the decision on whether to change the serving GW should be based on the movement direction of an MN in relation to the locations of GWs. If the MN is moving towards another GW, it is better to change the serving GW; if the MN is moving back and forth at the edge of the multi-hop coverage range of two GWs, it is better to stay with the current GW. The difficulty is that an MN does not know the exact pattern of its movement. A possible solution is to define a distance range for MNs that if an MN is inside the distance range to its current GW before the route breaks, it will try to recover the route to the same GW; if the MN is out of this range, it will try to find another GW upon route failure. Sometimes, inter-GW handover can also be performed even if the route to the current GW is still valid, e.g., the route to a new GW is constantly shorter, which can reduce packet delay.

2c) How to reduce handover latency?

Low delay handover is very important to real-time applications. As discussed in Section 3.3.3, the traditional handover latency is composed of movement detection latency, care-of-address configuration latency, and binding update propagation latency. In the hybrid network, an additional latency is introduced by route discovery/recovery. After an MN detects the route break to the current GW, it has to recover the route or discovery a new route to another GW. This delay may take up a substantial part in handover latency. Therefore, to reduce the route discovery/recovery latency is very important.

4.2 Related Works

MIPMANET [24] is one of the earliest papers that discuss integrating Mobile IP and AODV to connect MANET with the Internet. MIPMANET proposes the gateway discovery schemes and Candidate Access Router (CAR) selection algorithm named “MMCS (MIPMANET Cell Switching)”. The gateway discovery schemes include periodically flooding RA messages in the MANET and periodically unicasting RA messages to each registered MN. The MMCS requires an MN to perform handover to another GW if that GW is at least two hops closer to the MN’s current GW for two consecutive RAs. The limitations in MIPMANET are: (1) Mobile IP and AODV are transparent to each other, therefore the flooded RA is not used to construct the route to the GW and thus the MN still needs to perform route discovery for the GW later, which is not efficient. (2) It does not address micro-mobility issue. (3) The MMCS is only suitable for the case when the route to an MN's current GW is still valid. If the route is already broken, to make an MN wait for two consecutive RAs to decide changing the serving GW is not wise because it may cause long handover latency.

The scheme proposed in [25] is mostly based on [24] with minor modification to MMCS as CAR selection algorithm. An improvement in [25] is that the flooded RA message is used to construct a route to the GW, i.e., upon receiving an RA message, an MN gets a route for the GW and adds the route in AODV routing table. But this scheme incurs high overhead in network by flooding RA messages.

In [26], a scheme integrating MIPv4 and DSDV to extend traditional access point's coverage area to MANET is investigated. The CAR selection algorithm is similar with MMCS. However, it has the same limitations with those in [24].

In [27], the author discusses the issues on GW discovery, IPv6 address auto-configuration of MNs, and the routing procedure when the MANET and the Internet connect with each other. It proposes to use hybrid gateway discovery to balance the control overhead and connectivity performance. However, there are no simulation or implementation results presented.

Global6 [28] is an Internet draft which describes the issues of how to obtain a globally routable address and the Internet gateway operation. The connectivity method is not dependent on a particular MANET routing protocol. It also discusses the use of MIPv6 for global connectivity. However, it does not discuss the handover issue.

In [29], the gateway forwarding strategies in MANET using Mobile IP is studied. The paper compares two gateway forwarding strategies, namely traditional *default routes* and *tunnels*. Through analysis and simulation results, the paper concludes that using tunnel forwarding is a more suitable forwarding strategy in a multi-hop environment with multiple gateways.

In summary, the existing proposals [24, 25, 26, 27, 28] aim at providing Internet connectivity for MANET nodes. Reference [24] enlightens other later proposals to study in this area, hence the ideas are more or less similar, although they have their own contributions on discussing different aspects including the global address configuration [27][28], gateway discovery [24~28], communication in different scenarios [26][28], and gateway forwarding strategies [29]. However, with respect to GW selection, those proposals only use hop count as criterion without considering load balancing. Moreover, the multi-hop handover management is lack of studying although it is an important issue to provide continuous Internet connection for MANET. Especially when an on-demand routing protocol is used, the multi-hop nature makes the movement detection more difficult and may introduce longer communication disruption.

4.3 Scheme Overview

Fig 4.2 shows the hybrid network topology in our proposal. Since more challenges exist when using on-demand routing protocols in a hybrid network, we choose the AODV routing protocol as the representative on-demand routing protocol to study mobility management in hybrid networks. HMIPv6 is chosen to localize registration in each administrative domain. The MNs and GWs understand both HMIPv6 and AODV as illustrated in Fig 4.3.

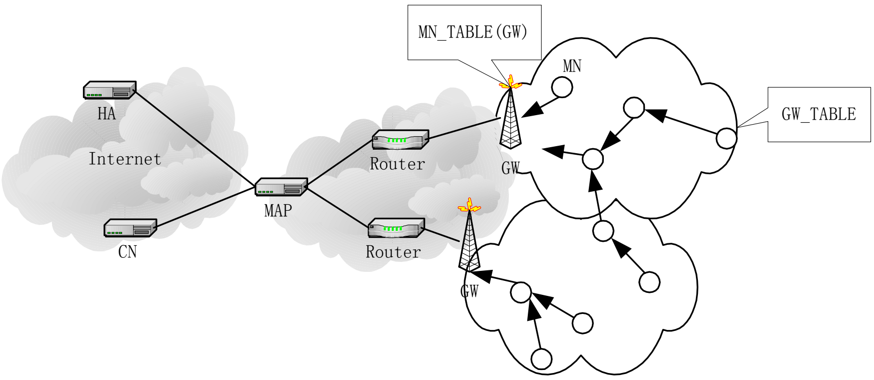


FIG 4.2 PROPOSAL NETWORK TOPOLOGY

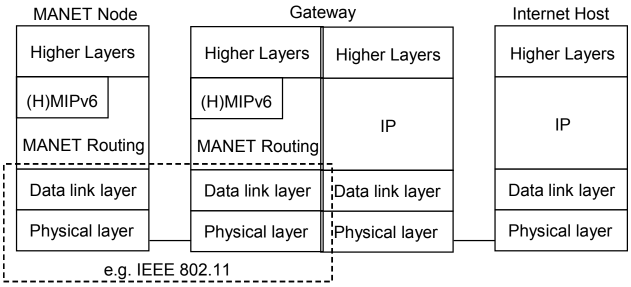


FIG 4.3 PROPOSAL ARCHITECTUR

The objective of our proposal is to integrate HMIPv6 and AODV efficiently and provide smooth handover for MNs connecting with the Internet. We assume each MN has a unique IP address for identification in wireless and wired networks; we call it the *home address* conforming to Mobile IP. The process of address configuration, which has been proposed in [27][28] is out of scope of this thesis. We also assume that an MN has the knowledge of whether the CN is located in the

Internet or the MANET, and focus on the former case in which the MN communicates with CNs in the Internet.

4.3.1 Gateway Discovery

To connect with the Internet, the MNs must find a GW. There are basically two approaches to realize GW discovery:

- Proactive approach: A GW periodically floods RAs through the network.
- Reactive approach: A GW will not send out RAs until it receives a solicitation from an MN, and then the GW unicasts an RA to that MN.

Both approaches can be implemented by using modified ad hoc routing protocols or modified Neighbor Discovery Protocol (NDP) [7]. Since the normal Router Advertisement (RA) and Router Solicitation (RS) messages in NDP are sent to on-link neighbors, the NDP should be changed to allow those messages to propagate through intermediate nodes. To integrate GW discovery and route discovery, the hop count information can be included in the option fields of the RA message. When an MN receives an RA message, it adds or updates the route to the GW using the hop count information and increments the hop count value before rebroadcasting the RA message.

To balance control overhead and MNs' connectivity with the Internet, we use hybrid GW discovery scheme by setting flooding range of RA. The flooding range means the number of hops that an RA is allowed to be propagated, and MNs out of this range need to send solicitations for RAs. It can be controlled by setting the initial value (N) of TTL field in the IP header of RAs. When N is equal to 1, the RA message is the same as in normal Mobile IP, which is only broadcasted to MNs in the transmission range of GWs. To prevent MNs from processing duplicate RAs, the RA message format is modified to contain a broadcast ID, so that <broadcast ID, GW's IP address> can uniquely identify an RA. The GW discovery process is different for the MNs inside or outside of RA flooding range.

For the MNs in the N-hop RA flooding range:

Upon receiving an RA message, an MN auto-configures its Local Care-of-address (LCoA) and Regional Care-of-address (RCoA), and records the GW/MAP information in its GW_TABLE, as shown in Table 4.1.

TABLE 4.1 GW_TABLE AT AN MN

MAP's IP address	GW's IP address	RCoA	LCoA	Metric	Other information
MAP1	GW1	RCoA1	LCoA1	Metrics1	...
MAP1	GW2	RCoA1	LCoA2	Metrics2	...
MAP2	GW3	RCoA2	LCoA3	Metrics3	...
...	...	...	...	...	...

The metrics such as hop count, signaling level and load information included in the options of the RA message provide the criteria for the MN to make GW/MAP selection. At the same time, the MN also adds an entry in its routing table since the modified RA message contains hop count information. If the MN moves into a new MAP's domain, it sends Binding Updates (BU) to the MAP, HA, and CNs. If the TTL is larger than 0, the MN increments the value of hop count in an RA message and rebroadcasts it.

<u>For the MNs out of the N-hop RA flooding range:</u>

. The MN sends solicitation for GWs. The solicitation is different from normal Router Solicitation in NDP [7] because the MN not only requires a GW's information, but also needs to know the route. On the reverse direction, the GW also needs to know the route to the MN. To realize these requirements, we propose two modified RREQ and RREP messages by using the reserved bit in normal RREQ/RREP messages to perform route discovery for GWs. The two new messages are named RREQ_GW and RREP_GW, which can construct bidirectional route as they propagate through the network, like normal RREQ/RREP messages. The format of the RREQ_GW/RREP_GW messages are shown in Fig 4.4 and Fig 4.5, respectively.

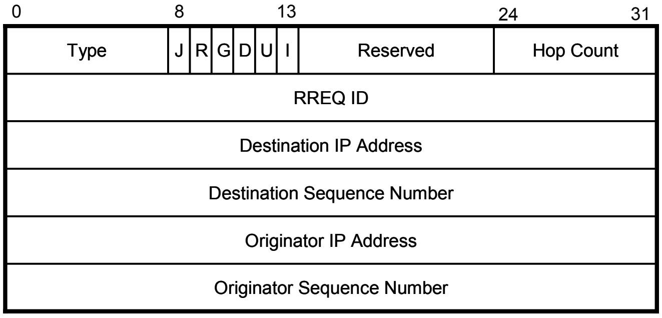


FIG 4.4 THE FORMAT OF RREQ_GW

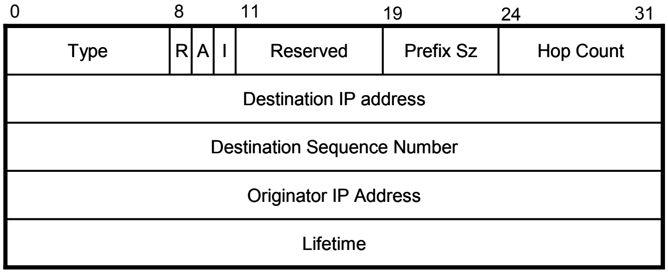


FIG 4.5 THE FORMAT OF RREP_GW

The "I-flag" added in normal RREQ/RREP message is called "Internet GW Discovery Flag". The RREQ_GW functions both as the RREQ to discover route and as the Router Solicitation to request GW information. Upon receiving RREQ_GW, the GW piggybacks an RA to RREP_GW and sends it back to the requesting MN. When the MN receives the RREP_GW, it performs the same procedure as when it is in the N-hop RA flooding range and receives an RA message, which has been discussed previously.

4.3.2 Registration & Packet Delivery

Registration

During initialization in a network, an MN will perform registration through a GW upon receiving the GW's information. After an MN receives an RA message that contains the GW's network prefix, MAP's network prefix and also MAP options, it adds an entry in its GW_TABLE and choose one of <MAP, GW> to register with. The choice can be based on criteria such as distance, cost or other information contained in RA message. We assume MNs can get enough information from the RA message to auto-configure a unique RCoA and LCoA which will be included in Binding Update (BU) messages.

When a GW receives a BU from an MN, it records the MN at its MN_TABLE, which is shown in Table 4.2.

TABLE 4.2 MN_TABLE AT A GW

MN's LCoA	MN's Home address	Other information
LCoA1	MN1	…
LCoA2	MN2	…
LCoA3	MN3	…
…	…	…

This table is used to keep track of registered MNs for routing decision, which will be explained later when discussing packet delivery. Moreover, the GW can record an MN's distance or location in the "other information" fields in MN_TABLE, which can be used to adjust the RA flooding range and to facilitate load balancing between GWs.

When MAP receives a BU from an MN, it adds an entry in its binding cache which maps the MN's RCoA to its LCoA; similarly, the HA updates the MN's entry to bind the MN's home address to its RCoA.

Packet delivery

After acquiring the route to a GW, the MN can forward packets using different approaches. As proposed in [28], one approach is called *next hop routing* or *default routing*. In this method, the MN sets the destination address to CN's IP address and sends out to the next hop towards its current GW; the packet is subsequently relayed depending on the next hop routing at other intermediate nodes. Another approach is *tunneling* in which the packet is encapsulated as it is sent to the gateway. Although the next hop routing has smaller packet header size, it may cause the problem of incorrect routes because the intermediate node's serving GW may not be the same as that of the source node. According to [29], tunnel forwarding outperforms default routing in a multi-hop environment with multiple gateways. In our scheme, we make use of IPv6 routing header to tunnel packets to GWs.

Fig 4.6 and Fig 4.7 show the uplink and downlink traffic delivery procedures. When the traffic from CN arrives at the GW under which the MN is located, the GW checks its MN_TABLE and changes the destination address to the MN's Home address (Haddr), and searches the MN's route entry in its routing table. The last part of the packets' journey is managed by the MANET routing protocol. The purpose of changing destination address to the MN's Haddr is to simplify the processing in MANET. Since MANET nodes identify each other using Haddr at the initiation time, it is better to keep the identity; otherwise they may not be able to identify each other and the routing table will be difficult to maintain.

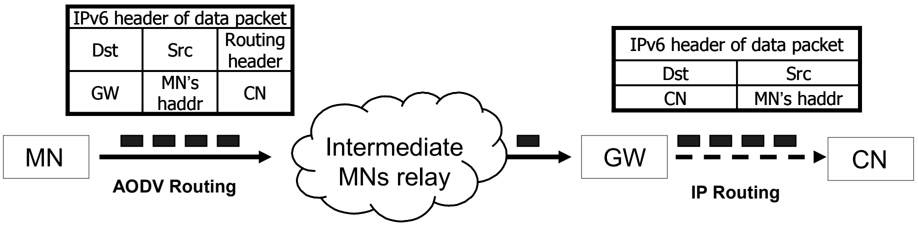


FIG 4.6 TRAFFIC DELIVERY FROM AN MN TO A CN

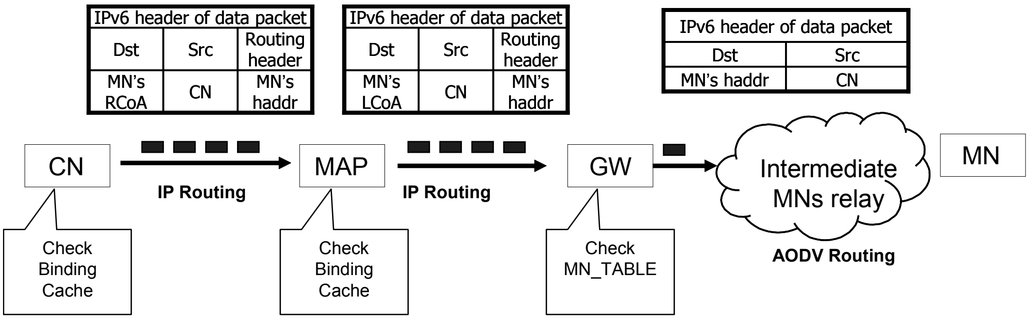


FIG 4.7 TRAFFIC DELIVERY FROM A CN TO AN MN

4.3.3 Multi-hop Handover

Definition and classification

The concept of handover is similar to that in Mobile IP; however, it needs to be redefined in hybrid networks because of the multi-hop nature. During handover in the traditional one-hop scenario, the ongoing traffic is redirected through another AR; consequently, in multi-hop scenario, the redirection of ongoing traffic through another route can be considered as handover. **Therefore, we define the multi-hop handover in hybrid network as a route change from an MN to GWs during communication.** The multi-hop handover may occur when an MN itself or any of the intermediate MNs moves and breaks the active route during the MN's communication with a CN in the Internet.

In normal Mobile IP, the handover process includes Layer 2 handover and Layer 3 handover. In a hybrid network, since some MNs do not have link layer connections with GWs, the handover is only performed at Layer 3.

According to different situations, multi-hop handover can be categorized as follows:

- Intra-GW handover & Inter-GW handover

  Since each route change from an MN to its serving GW is considered as a handover in our definition, the route change may happen with the MN connected to the same GW (intra-GW handover) or to different GWs (inter-GW handover).

- Compulsory handover & Optimized handover

  Compulsory handover happens when an MN detects a break in the route to its current GW. Optimized handover happens when the MN changes to a better GW while the route to its current GW is still active.

Combining the above two categories of handover, we classify multi-hop handover into four types:

1. Compulsory Intra-GW handover (Comp_intraGW_HO): an MN recovers a route to the same GW upon the detection of a route breakage to it's serving GW.
2. Compulsory Inter-GW handover (Comp_interGW_HO): an MN finds and uses a route to another GW upon the detection of a route breakage to it's serving GW.
3. Optimized Intra-GW handover (Opt_intraGW_HO): an MN uses a shorter path to its current GW without route breakage, which is managed by normal AODV route maintenance mechanism.
4. Optimized Inter-GW handover (Opt_interGW_HO): an MN uses a shorter path to another GW without route breakage.

The impact of multi-hop handover to communication

Considering the four types of handover, the compulsory handovers (Comp_intraGW_HO & Comp_interGW_HO) are more stringent issues to address because the route to the current GW is not available any more. Before the route to a GW (whether the same as or different from current the GW) is found, the communication between an MN and its CN will be disrupted. As a result, to quickly find a route is critical to the performance of communication. Optimized Inter-GW handover (Opt_interGW_HO) is also interesting because it may affect the communication performance too. In multi-hop scenarios, using a shorter path can reduce end-to-end delay of packets; moreover, a shorter path to a new GW may indicate the movement of the MN towards another GW. Therefore, using Opt_interGW_HO can prevent a potential Comp_interGW_HO and thus prevent a future route breakage. Optimized intra-AR handover (Opt_intraGW_HO) happens when an MN changes to a shorter path to its current GW; since it can be achieved by the normal AODV mechanism, it will not be discussed later.

In the hybrid network, it is possible for an MN's route to a GW to break while the GW entry is still valid in GW list. It is because an entry in GW list (which is managed by HMIPv6) has a life time 3 times that of RA interval and AODV does not distinguish the routes to GWs from routes to other MNs. As a result, when the route to the current GW breaks, the HMIPv6 module of the MN is not aware of the loss of its current GW and considers that GW valid until the GW entry expires. In our scheme, this problem is addressed by setting a flag for the route to the GW to help an MN identify its GW route.

Multi-hop handover mechanism

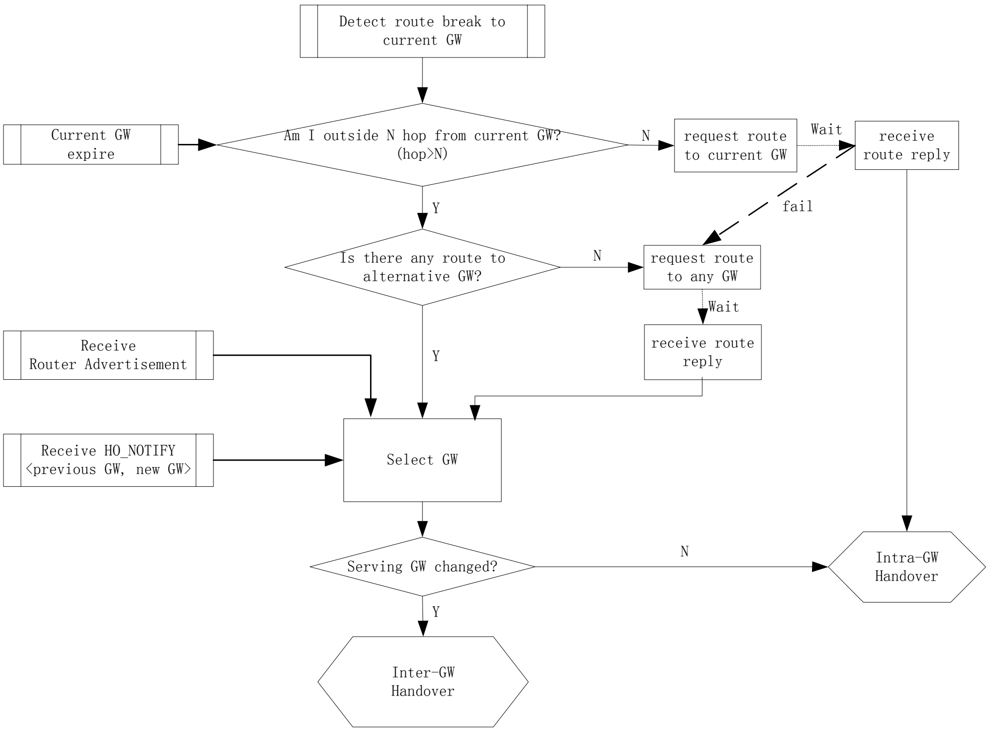


FIG 4.8 MULTI-HOP HANDOVER MECHANISM

Our multi-hop handover mechanism is shown in the Fig 4.8; the intra-GW and inter-GW handover procedures are illustrated in Fig 4.9 and Fig 4.10.

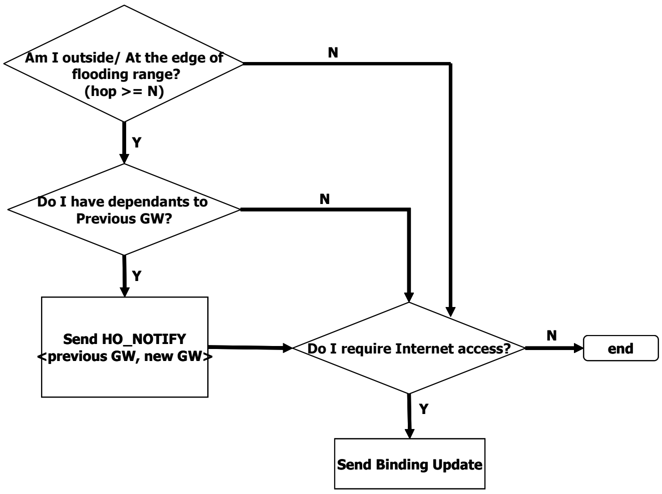


FIG 4.9 INTER-GW HANDOVER MECHANISM

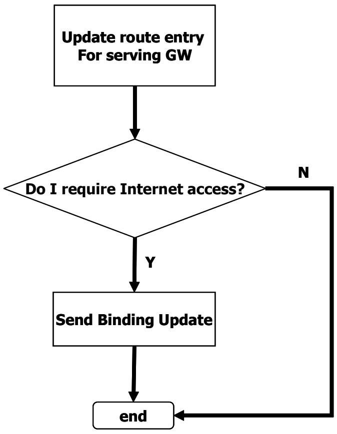


FIG 4.10 INTRA-GW HANDOVER MECHANISM

As illustrated in Fig 4.8, the N-hop RA flooding range not only balances the tradeoff between signaling overhead and MNs' connectivity with the Internet, it also helps an MN to make Inter-GW handover decisions. Inside the flooding range, an MN can periodically receive RAs and update the route to that GW. Hence, upon route failure it will attempt to recover the route to the previous GW. Outside the range, the MN uses RREQ_GW to request route for any possible GW.

The metric for GW selection

The existing proposals in hybrid networks use MMCS [24] or similar schemes to select GWs. In MMCS, an MN uses only hop count as metric to select a GW, and performs inter-GW handover if a new GW is at least two hops closer than the current one for two consecutive RAs. In our GW selection algorithm, the workload at a GW is also considered as one criterion besides the hop count. The reason is that there is a certain overhead involved with providing mobility services for an MN at a GW, such as forwarding the MN's registration messages, decapsulating and forwarding packets tunneled by the CNs, etc. These tasks represent an operational overhead that may decrease the quality of service by a

GW that is serving several MNs simultaneously. Moreover, it is possible that there exist hot spots in the network where the density of MNs requiring Internet connectivity is high. If those MNs happen to select the same GW, some wireless links will be congested and collisions will increase. We measure the workload at a GW by the number of MNs served at the GW. When a GW receives a registration or BU message from an MN, the GW adds or updates MN_TABLE for the MN and updates the number of served MNs, which is included in its RA messages.

Considering the multiple selection criteria, the combined selection metric C is defined as: $C = \sum_{i=0}^{1} w_i * C_i$, where C_0 denotes the hop count to a GW, C_1 denotes the number of served MNs at a GW, and w_i (i=0 or 1) denotes the weight factor of the two criteria. Preference is given to the GW with the smallest value of C. The value of w_i (i= 0 or 1) should be obtained from a knowledge base that is derived from empirical observations. Given an arbitrary scenario, the selection of w_i to obtain maximum performance is an interesting issue. Since the current focus of our work is studying methods of reducing handover latency, we set w_i to be 1 or 0 for simplicity. The two sets of values for w_0 and w_1 that we use throughout simulations are: 1) w_0=1, w_1=0, which considers only hop count. 2) w_0=1, w_1=1, which considers both hop count and workload at GWs.

<u>Approaches to reduce multi-hop handover latency</u>

The main objective of our handover scheme is to achieve low handover latency and small packet loss. As discussed in Section 4.1, the handover latency consists of the time spent on 1) movement detection, 2) CoA configuration and 3) BU propagation. In hybrid networks, an additional latency is introduced into multi-hop handover by route discovery/recovery. To reduce movement detection, we make AODV inform HMIP upon route breakage to the current GW; therefore it reduces the time delay before HMIP initiates its handover mechanism. By using HMIPv6, the mobility is localized in a MAP's domain and the BU propagation time is reduced. Since CoA configuration time is out of scope in this thesis, the remaining issue is how to reduce the route recovery latency.

We propose three approaches to reduce route recovery latency:

- Use handover notification (HO_NOTIFY):

When an MN performs a compulsory handover, it informs its dependants with the HO_NOTIFY message which contains <previous GW, new GW> routing information. When its dependants receive HO_NOTIFY, the dependants can construct route to the new GW, and thus they do not need to perform route discovery again. Hence, using HO_NOTIFY not only helps to reduce handover latency, but also reduce control overhead. To show the benefits of using HO_NOTIFY, we use a simple example here. The network topology is shown in Fig 4.11. At the beginning, MN(1) is in the transmission range of GW(0) only; MN(0) and MN(2) are out of transmission range of both GWs. MN(2) has CBR traffic to send (0.004Mbps) and depends on route MN(2)-MN(0)-MN(1) to connect to GW(0). MN(0) and MN(2) are neighbors and keep static during simulation. MN(1) will move towards GW(1) and there will be a period in which MN(1) loses its connection with either GW. The distance between MN(1) and MN(0) is set to ensure that they are neighbors during MN(1)'s movement. After MN(1) constructs a connection with GW(1), MN(2) will recover a route to GW(1).

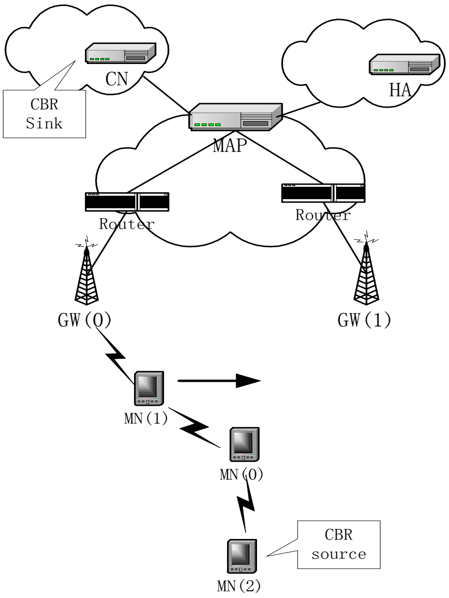


FIG 4.11 SIMPLE EXAMPLE SCENARIO

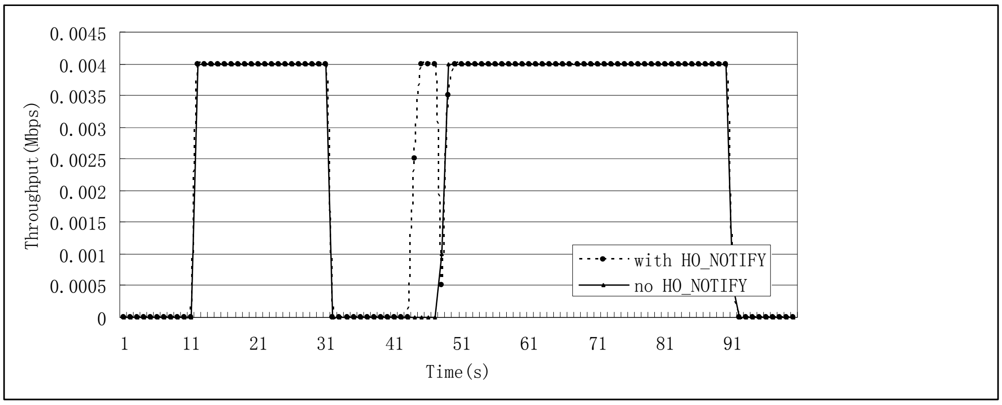


FIG 4.12 THROUGHPUT COMPARISON

As shown in Fig 4.12, there is a traffic disruption period during which MN(1) is moving from GW(0) to GW(1) and is out of transmission range of both GWs. The data traffic is recovered earlier in the scheme with HO_NOTIFY because when MN(1) handover to GW(1), it informs its dependants about the new GW's information; consequently, MN(2) does not need to perform route recovery.

- Make intermediate nodes reply to route requests for GWs if they have active routes to GWs.

When an MN is outside the flooding range of RAs and wants to connect to the Internet, it sends out RREQ_GW with destination to a GW multicast address. Receiving RREQ_GW, a GW piggybacks an RA in the RREP_GW to the MN. In existing proposals, the intermediate nodes will not process this kind of route request, and just rebroadcast it until it arrives at a GW. To let intermediate nodes reply RREQ_GW with the information of the active GWs in their GW list can reduce the route recovery latency.

- Make the reply from a GW heard by intermediate nodes.

When a GW receives RREQ_GW and sends back RREP_GW including RA, the intermediate nodes processing the reply can update their GW table as well.

We name the handover scheme with the three approaches as “Enhanced HMIPAODV (E-HMIPAODV)”, and the scheme without the three approaches as “Plain HMIPAODV (P-HMIPAODV)”. We studied the performance of both E-HMIAODV and P-HMIPAODV through simulations, which will be presented later in chapter 5.

4.4 Summary

In this chapter, we discussed providing mobility management for mobile nodes multiple hops away from Internet gateways. HMIPv6 functionalities are extended into MANET by efficiently integrating with AODV routing protocol. We defined multi-hop handover in hybrid network and proposed a handover scheme that aims to reduce handover latency and packet loss. In the next chapter, we will show the simulation work to study the performance of handover schemes we have proposed in this chapter and in Chapter 3.

CHAPTER 5 SIMULATION RESULTS

This chapter presents simulation results of handover in IP/MPLS based HMIPv6 networks (chapter 3) and multi-hop handover in hybrid networks (chapter 4). In Section 5.2, two handover schemes (with or without L2 trigger) are compared in terms of handover latency and packet loss. The effect of bicasting is also examined. In Section 5.3, the performance of two multi-hop handover schemes, namely E-HMIPAODV and P-HMIPAODV, are studied and compared.

5.1 Simulation Tools

The simulator we have used is the Network Simulator 2 (NS2). To simulate the mobile wireless environment, we have used a mobility extension to NS2 that is developed by the CMU Monarch project at Carnegie Mellon University. NS2 is a discrete event simulator targeted at networking research, which is written in C++ and a script language called OTcl. NS2 uses an OTcl interpreter for which the user writes an OTcl script to define the network, traffic and protocols. This script is then used by NS2 during simulations. The result of the simulation is an output trace file which is then processed to obtain performance data, such as delay and throughput. Network Animator (NAM) is a program to visualize the simulation.

MobiWan is developed by Motorola Labs Paris in collaboration with INRIA PLANETE Team, as a patch for NS2 to simulate MIPv6 under large area networks.

5.2 Simulation of Handover in IP/MPLS Based HMIPv6 Networks

Two handover schemes, namely scheme 1(using HMIPv6 + MPLS) and scheme 2 (using HMIPv6 + MPLS + L2 trigger), are compared in terms of handover latency and packet loss. The goal of the simulations is to examine the effects of using L2 trigger on handover performance.

5.1.1 Simulation Model

The NS2 extension

We extend the MPLS module in NS2 to support hierarchical address format, which is necessary for Mobile IP based simulation. Our simulation uses IEEE 802.11 as MAC layer protocol. The L2 trigger is implemented as proposed in [20][22], where we propose to extend the IEEE 802.11 beacon frame that is advertised by the access points to carry extensible Information Element such as the network prefix. When the MN hears an extended beacon from a new access point while moving within the overlapping area, it can detect its movement into a new subnet and initiates the handover process (pre-registration). This extended beacon frame is expected to assist MNs to achieve faster movement detection.

Simulation Scenario

We use the same network topology for the two schemes as shown in Fig 5.1. The simulation parameters are summarized in Table 5.1.

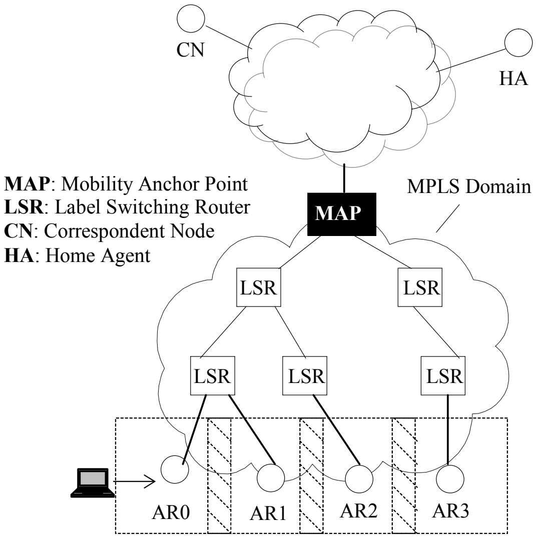


FIG 5.1 SIMULATION SCENARIO

TABLE 5.1 SIMULATION PARAMETERS (A)

Simulation time	600s
Wired link bandwidth	100Mbps
Wired link delay	5ms
Overlap of AR	[0, 5, 10, 15, 20]m
Number of MN	1
Speed of MN	10m/s
packet Size	500byte
Traffic interval	10ms
RA interval	[1, 3, 5, 7]s
IEEE802.11 Beacon interval	[0.2, 0.5, 0.7]s

UDP CBR traffic is directed from CN to an MN at a sending rate of 400kbps. The ARs are positioned 200 meters apart, and there is an overlap area between each pair of neighboring ARs as represented by the shadowed area in Fig 5.1. We change the overlapping area by adjusting the transmitting power of ARs. The MN moves linearly from one AR to another at a constant speed of 10m/s, and it moves back and forth between the two edge ARs. When the MN crosses the cell boundary, it performs the handover process after receiving the first MIPv6 Router Advertisement or an IEEE802.11 beacon (which is extended to carry network-prefix information as proposed in [20]). The MN is able to send/receive data only via the AR that corresponds to its current CoA.

5.1.2 Simulation Results

1) Handover Latency

In the simulations, the handover latency is calculated as the time that elapses between the last packet received by an MN via the old AR and the arrival of the first packet from the new AR after a handover. We take three critical parameters into consideration: sending rate of Router Advertisement (RA) message at AR, sending rate of beacon frame at AR, and overlap between ARs. Fig 5.2 shows the average handover latency experienced by the MN when the overlap of AR varies from 0 to 20m. Handover latency in scheme 2 which has the L2 trigger is much smaller than scheme 1; the difference is 1.4s when overlap is 0m and about 0.4s when overlap is 20m. When the overlap increases, the MN is more likely to receive the RA message sent from the new subnet before it loses contact with the old AR. This is why the handover latency experienced by scheme 1 is reduced largely with increased overlap. We also noted that the performance of scheme 2 using different beacon intervals of 0.2s, 0.5s, and 0.7s would have little difference when the overlap is larger than 10m. A possible reason is that, the probability of an MN losing an RA message or beacon frame decreases as the overlapping region increases. Consequently, with increased overlapping region, the binding list at the MN is updated more promptly. Therefore the increased sending rate of beacon will just renew the binding list although it is not about to expire yet. Fig 5.3 shows the average handover latency at various RA intervals. When the RA interval is set to 1s, the two schemes have little difference. But as the interval increases, the handover latency for scheme 1 increases greatly. This means that the performance improvement will be more apparent for our scheme when the RA interval is large.

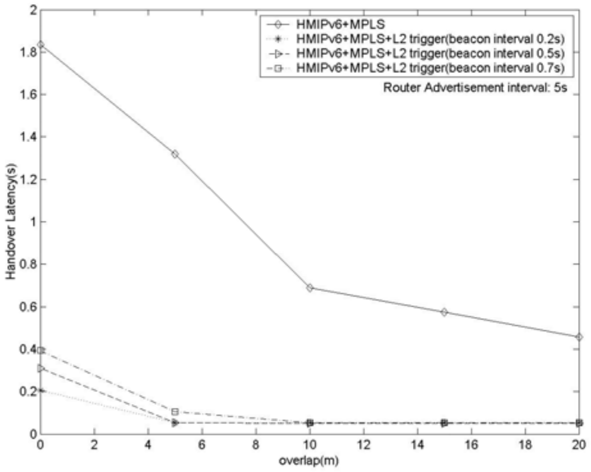


FIG 5.2 HANDOVER LATENCY VS. OVERLAP

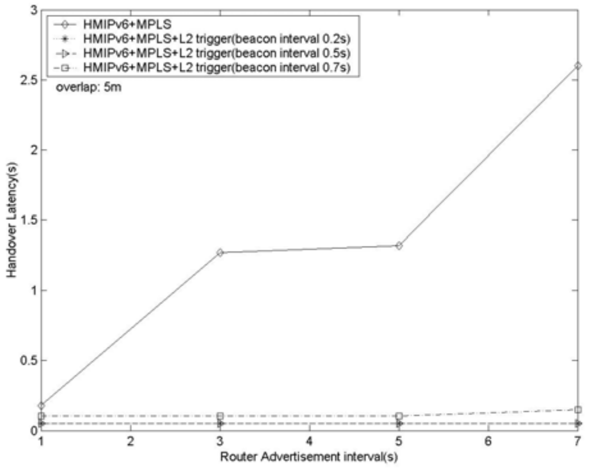


FIG 5.3 HANDOVER LATENCY VS. ROUTER ADVERTISEMENT INTERVAL

2) Packet Loss

Fig 5.4 and Fig 5.5 show the packet loss ratios when the overlap between ARs and RA interval is changed respectively. The scheme with L2 trigger decreases the packet loss during handover. The two graphs look similar to Fig 5.2 and Fig 5.3 respectively. It can be explained that as we only consider one MN in the simulations, the packet loss experienced by the MN is mainly due to the handover latency during which the traffic is disrupted. If the number of MNs is increased, the packet loss will be additionally caused by the contention between the MNs.

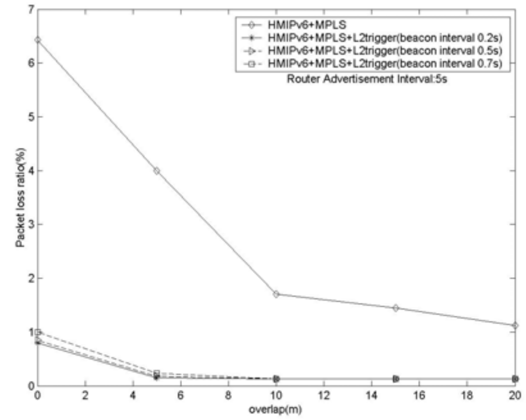


FIG 5.4 PACKET LOSS RATIO VS. OVERLAP

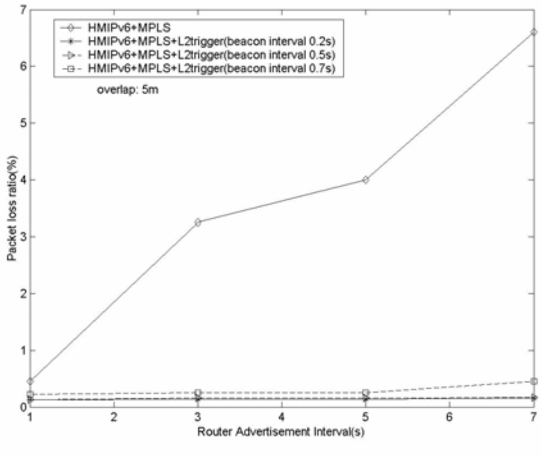


FIG 5.5 PACKET LOSS RATIO VS. ROUTER ADVERTISEMENT INTERVAL

3) Effect of bicasting

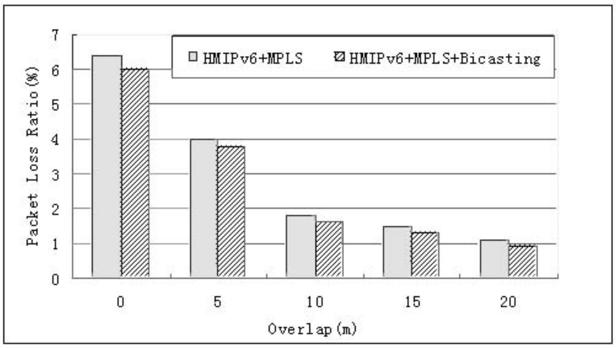


Fig 5.6 Packet Loss vs. overlap (effect of Bicasting)

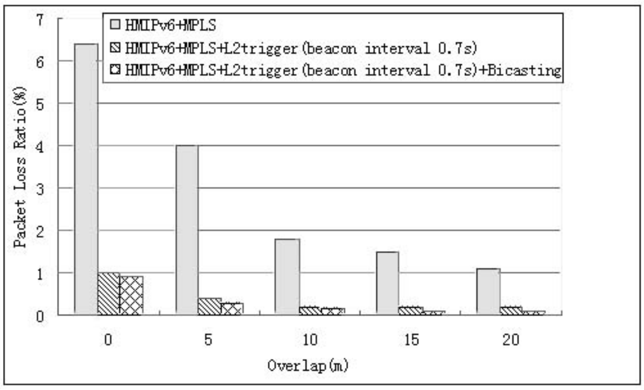


Fig 5.7 Packet Loss vs. overlap (effect of L2trigger and Bicasting)

Fig 5.6 and Fig 5.7 show the packet loss ratio with the changing of overlap area of ARs and compared with different combination of approaches. Fig 5.6 shows the effect of bicasting. Comparing with Fig 5.4, the performance enhanced by only using bicasting is not so significant, as the L2 message is sent from access point. Therefore, when an MN receives the message and initiates handover, it is in the transmission range of the new AR. However, the bicasting happens when the MN receives the L2 message. The MN informs the previous AR, and then the previous AR bicasts packets both directly to the MN and through the tunnel to the new AR. Hence, the packets that were sent directly from the old AR do not help much for the packet loss. Theoretically, the enhancement of traffic performance by bicasting is greater in the case when an MN anticipates a handover before losing connection with its current AR. Fig 5.7 shows the performance of the three schemes. With L2 trigger, performance can be greatly enhanced and it is further enhanced when combined with bicasting.

Results analysis

From the simulation results, we can conclude that in terms of handover latency and packet loss, the scheme with L2 trigger shows a better performance. But as a tradeoff, L2 trigger may introduce some overhead at wireless channel. Indeed, in our simulation, the new information added in the beacon frame enlarges the original frame size, and because of the relatively high sending rate of beacon, the L2 overhead will be increased largely. However, with the assistant of L2 trigger, the L3 signaling such as RA message can be reduced (e.g. we can set a higher RA interval). Further, to eliminate the L2 overhead as much as possible, ARs can adjust the sending rate of beacon based on the movement information collected from all the mobile nodes under its service area.

When the MNs' mobility increases, the beacon rate should be increased to assist low latency handover; and when the MNs become more stable, the beacon rate can be reduced to prevent the unnecessary signaling overhead. To find a good balance between the signaling overhead and handover performance, an optimal combination of beacon sending rate, RA sending rate, and the lifetime of entries of AR list or Binding update list at an MN is an interesting issue worth further study, as well as the mechanism to collect the information of MNs' movements.

5.3 Simulation of Multi-hop Handover in Hybrid Networks

The main aims of the simulations in this section are to study the multi-hop handover in hybrid networks and to examine the effect of our approaches to reduce handover latency (as discussed in Section 4.3.3) by comparing the performance of E-HMIPAODV and P-HMIPAODV.

We study handover latency, packet loss ratio, and control overhead under various mobility levels (with varying pause times), and other related network parameters (e.g. RA interval, RA flooding range). The control overhead includes AODV control overhead (RREQ/RREP, RREQ_GW/RREP_GW, RERR) and HMIPv6 control overhead (BU/BACK, RA), which is measured by the number of transmissions of messages.

5.3.1 Simulation Model

The NS2 extenstion

We modify the AODV routing protocol and MIPv6 module to implement our multi-hop handover scheme because the existing MIPv6 module in NS2, which is named MobiWan, cannot work with MANET routing protocols.

Simulation Scenario

Fig 5.8 shows our simulation scenario, and the simulation parameters are summarized in Table 5.2.

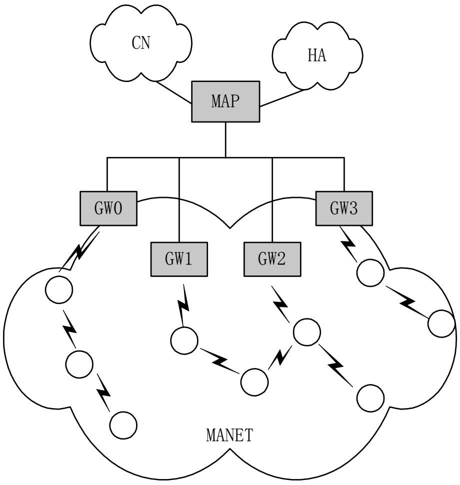


FIG 5.8 SIMULATION SCENARIO

TABLE 5.2 SIMULATION PARAMETERS (B)

Simulation time	600s
Simulation area	1000mx1000m
Number of GW	4
Number of MN	50
Speed of MN	Uniform [0, 10]m/s
Pause time	[5, 10, 20, 100, 200, 300, 400]s
packet Size	50byte
Traffic interval	100ms
RA interval	[5, 10, 20, 30, 40]s
RA flooding range	[1, 2, 3, 4, 5]hop

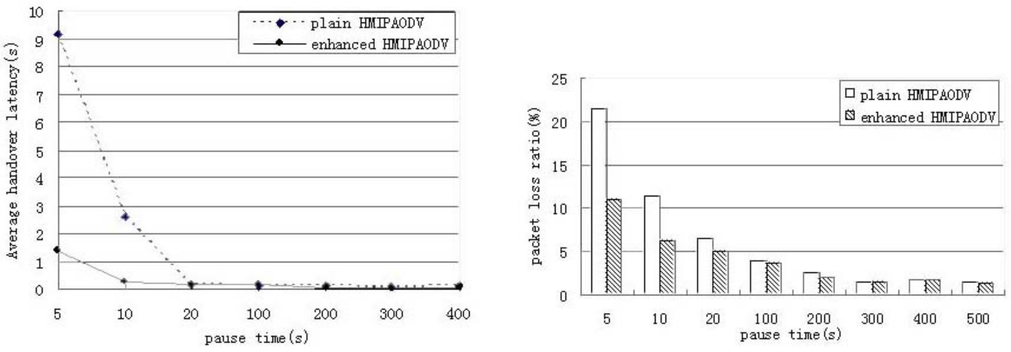
The wired network consists of a cloud of five CNs (CN0 to CN4), one HA, one MAP, and four GWs. In the wireless network, we study the network with 50MNs over a terrain size of 1000x1000m. To simplify the simulation, we use one MAP to serve the whole wireless network, thus there is no handover between MAPs. Out of the 50 MNs, 5 are CBR sources, and the 5 CNs are CBR sinks. Each source node sends constant bit rate (CBR) traffic at a rate of 10packet/s with each packet size as 50bytes. We use random way point mobility model to simulate the movement scenario. The mobility model that we have used is the Random Waypoint with a maximum speed of 10m/s. In addition, we also use varying pause times to simulate different levels of mobility.

5.3.2 Simulation Results

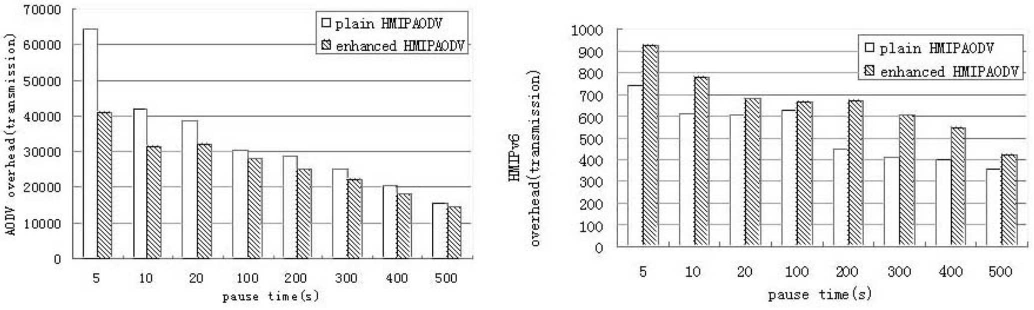
1) The effect of mobility

The purpose of this set of simulations is to study network performance under different mobility levels. We compare the performance of “Enhanced HMIPAODV” (E-HMIPAODV) and “Plain HMIPAODV” (P-HMIPAODV) as specified in Chapter 4.3.3.

Each mobile node moves randomly with speeds uniformly distributed between 0m/s and 10m/s. The pause time is set to [5, 10, 20, 100, 200, 300, 400]s in each simulation. RA flooding range is set to 1 and RA interval is set to 10s for all simulations in this set.



(A) Average Handover Latency (B) Packet Loss Ratio



(C) AODV Overhead (D) HMIPv6 Overhead

FIG 5.9 THE EFFECT OF MOBILITY

Fig 5.9(A) shows that E-HMIPAODV has less average handover latency than that of P-HMIPAODV. For both schemes, the handover latency decreases with increased pause time. As for traffic performance, from Fig 5.9(B) we can see that E-HMIPAODV has less packet loss than P-HMIPAODV under different mobility levels. We also note that the enhancement is greater with smaller pause times. We record the occurrence of different kinds of handover within the entire simulation time. Table 5.3 shows the total number of occurrences experienced by the 5 MNs which generate CBR traffic. The occurrence of handovers decreases with increasing pause times. E-HMIPAODV generally experiences a higher frequency of handovers than that of P-HMIPAODV. It can be explained that E-HMIPAODV is more sensitive to route breakages and can recover routes more quickly, and thus it can finish the handover process faster. Therefore, in a given time period that a GW route breaks for both schemes, the E-HMIPAODV may have recovered the route and experienced another route breakage while P-HMIPAODV has not recovered the GW route.

TABLE 5.3 HANDOVER RECORD

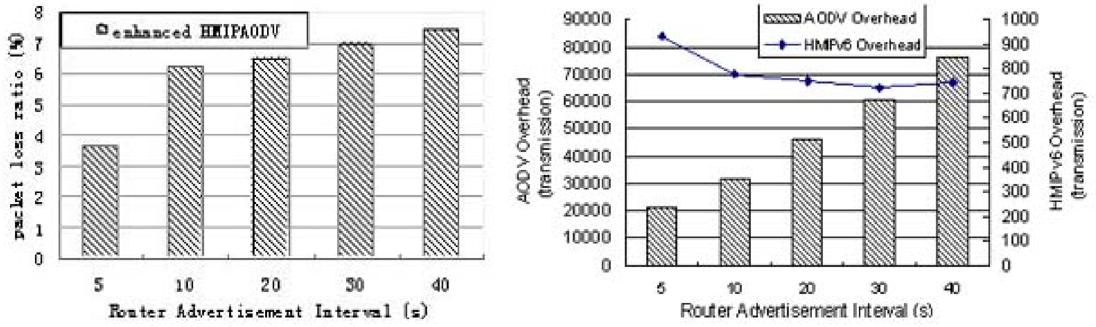
Pause time (s)	Number of intraGW_HO		Number of interGW_HO		Number of Opt_HO	
	P-HMIPAODV	E-HMIPAODV	P-HMIPAODV	E-HMIPAODV	P-HMIPAODV	E-HMIPAODV
5	12	21	18	22	9	18
10	12	18	11	21	7	11
20	10	16	6	10	5	8
100	8	13	5	9	3	6
200	6	10	3	6	4	8
300	4	6	4	8	3	5
400	3	8	3	6	2	6

Fig 5.9(C) and Fig 5.9(D) show control overhead which is measured as the

total number of packet transmissions during the simulation time. E-HMIPAODV reduces AODV control overhead by using HO_NOTIFY and allowing intermediate nodes to reply to RREQ_GW. However, E-HMIPAODV introduces more HMIPv6 control overhead because each time the MN performs a handover (intra-GW or inter-GW), it sends a BU to its MAP. The purpose of sending BU upon intra-GW handover is to update the downlink route from a GW to an MN. Since E-HMIPAODV performs more handovers than P-HMIPADOV, it sends out more BUs into the wireless network.

2) The effect of Router Advertisement interval

This set of simulations is done with RA flooding range set to 1 and pause time set to 10s. We focus on studying E-HMIPAODV to examine the effect of RA interval.



(A) Packet Loss Ratio (B) Control Overhead

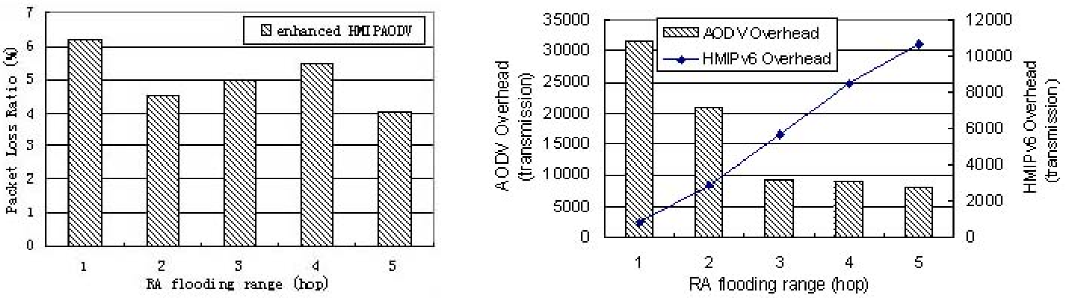
FIG 5.10 THE EFFECT OF ROUTER ADVERTISEMENT INTERVAL

As shown in Fig 5.10(A), the packet loss ratio increases when the RA interval increases, because the modified RA message is also used to update the route to a GW. The increased RA interval causes the MNs in the RA flooding range to send requests to discover routes to GWs, which increases the handover latency and

AODV control overhead (as shown in Fig 5.10(B)) while reducing HMIPv6 control overhead (as shown in Fig 5.10(B)).

Theoretically, when RA messages are flooded in the wireless network with sending rate faster than the time required to detect link breakages in AODV routing protocol, there will be no handover delay and thus no communication disruption for MNs. AODV uses periodic "HELLO" messages with a default broadcast interval of 1s to maintain connectivity with neighbors and the default permitted loss number of HELLO messages is 3. Accordingly, the maximum time required to detect link break is 3s. If RA messages are flooded in the whole network with sending intervals of less than 3s, there will be no communication disruptions. However, this will incur excessive overhead in the network, which can be a great waste of limited bandwidth especially when there are few MNs that require Internet connectivity. Moreover, the frequently flooded RA messages may affect the data traffic.

3) The effect of Router Advertisement flooding range



(A) Packet Loss Ratio (B) Control Overhead

FIG 5.11 THE EFFECT OF ROUTER ADVERTISEMENT FLOODING RANGE

This set of simulations is done with RA interval set to 10s and we focus on studying E-HMIPAODV to examine the effect of RA flooding range. From Fig 5.11(B), the AODV overhead is reduced when RA flooding range is increased, but the decrease is not significant when the flooding range N is larger than 3. As shown in Fig 5.11(B), the HMIPv6 overhead increases with N because the main component of HMIPv6 overhead is the propagations of RA messages. Fig 5.11(A) shows the traffic performance, which does not show much improvement when N is larger than 3. From the observations in Fig 5.11(A) and Fig 5.11(B), we can deduce that most MNs move within 3 hops from GWs. Accordingly, when N is larger than 3, the flooded RAs will not benefit much for MNs and also increase overhead. Under this set of network parameter, the RA flooding range of 3 is optimal considering both traffic performance and control overhead.

4) The effect of considering load balancing in the GW selection

In the simulation works presented above, we only use hop count as GW selection metric for simplicity by setting w_0=1 and w_1=0 (as shown in the GW selection metric proposed in Section 4.3.3). To examine the effect of considering load balancing in GW selection, we compare the traffic performance between w_1=0 and w_1=1: when w_1=0, the GW selection metric considers only hop count; when w_1=1, the workload at a GW is also considered. We increase the number of source nodes in the 50 MNs from 5 to 30 to simulate increased traffic load in the network, and calculate the average packet loss ratio.

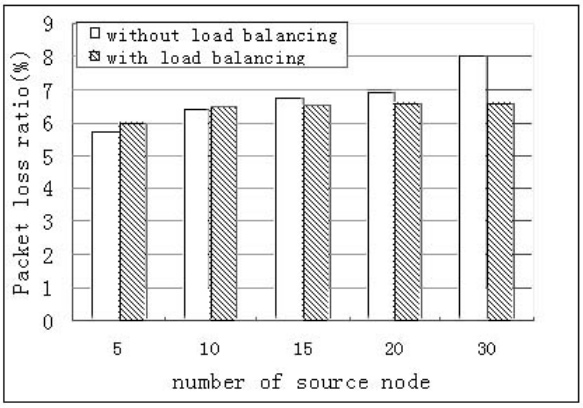


FIG 5.12 PACKET LOSS RATIO VS. NUMBER OF SOURCE NODE

Fig 5.12 shows the average packet loss ratio when the number of source nodes increases. Without load balancing, the packet loss ratio increases when the traffic load is increased in the network. This is because some mobile nodes choose the same GW, which causes the formation of “hot spot” areas in the MANET. Consequently, more packets are dropped due to increased collisions. This phenomenon verifies our analysis in Section 4.3.3 that it is reasonable to consider load balancing in GW selection. When load information is included in the GW selection criteria by MNs, the probability of collisions can be reduced and the traffic performance is enhanced as shown in Fig 5.12.

5.4 Summary

In this chapter, we conducted simulations to evaluate the performance of the handover schemes proposed in Chapter 3 and Chapter 4. For handover in IP/MPLS based HMIPv6 networks, simulation results show that using L2 trigger can greatly reduce handover latency by reducing the delay of movement detection. We also studied the multi-hop handover in a hybrid network under different network parameters. The effect of the approaches to reduce multi-hop handover latency is examined. Simulation results show that the three approaches effectively reduce the delay of route discovery for GWs and thus achieve smoother handovers. The importance of load balancing in a hybrid network is also studied.

CHAPTER 6 CONCLUSIONS AND FUTURE WORK

6.1 Conclusions

The increasing popularity of real-time Internet applications and the rapid growth of mobile systems indicate that the future network architecture will have to support Internet connectivity to various mobile networks. The main contribution of this thesis work involves studying mobility management in two kinds of wireless networks.

A framework of mobility management scheme in IP/MPLS based HMIPv6 networks is presented. By combining the advantages of HMIPv6 and MPLS, the signaling overhead and binding update latency is reduced in the event of local handover. By using L2 trigger to perform faster movement detection, handover latency is greatly reduced. We implemented the L2 trigger in IEEE802.11 in Network Simulator (NS). Through simulations, the effects of using L2 trigger and bicasting are studied. The simulation results show improved performance of using L2 trigger in terms of reduced handover latency and decreased packet loss.

A mobility management scheme in hybrid networks is proposed. By efficiently integrating HMIPv6 and AODV, the MNs that are multiple hops away from GWs can continuously connect to the Internet. To provide MNs with smoother communication during movement, we have defined multi-hop handover and proposed approaches to reduce handover latency. The key to provide a smooth handover for MNs in hybrid networks is to reduce the latency of route recovery to GWs upon the occurrence of route breakages. We have proposed the use of handover notification and also some modifications to AODV when processing the route maintenance for routes to GWs. Considering multiple GWs in the hybrid network, we have also presented a GW selection algorithm, which considers both hop count and load balancing. We conducted simulations to evaluate and compare different multi-hop handover schemes. The results demonstrate that our proposed approaches can reduce the handover latency and packet loss.

6.2 Future work

While the work in this thesis concerns more on achieving faster handover for delay-sensitive applications, there are a few more issues that can be further explored:

- Optimum values of the interval and the broadcast hops of RA

Under different network scenarios (such as MNs' density, MNs' movement pattern, mobility level of network, etc.), the RA interval and RA flooding range should be adjusted to achieve optimal performance. Therefore, a method to collect real-time network parameters and an algorithm to calculate optimal RA interval as well as flooding range are required.

- Load balancing between mobility agents

The mobility agents are responsible for redirecting traffic for registered MNs.

When the number of registered MNs increases, the workload of a mobility agent will be increased. Balancing the workload between mobility agents can allow more efficient usage of network resources and also prevents the "one-point-of-failure" problem. As discussed in Section 4.3.3, the optimum combination of w_i (i=0 or 1) in GW selection metric is an interesting issue to study, which requires empirical observations.

- Extending MPLS to wireless part of hybrid network

Since MPLS is a packet forwarding scheme with high scalability, it is becoming a key technology for traffic engineering and fast packet forwarding in wired networks. Hence, it is beneficial to extend MPLS to MANETs to achieve both scalable mobility support and QoS support.

REFERENCES

[1] L. Bos, S. Leroy, "Toward an all-IP-based UMTS system architecture", IEEE Network, Vol.15, No.1, Jan/Feb 2001, pp.36-45.

[2] J. D. Vriendt, P. Laine, C. Lerouge, and X. Xu, "Mobile Network Evolution: A Revolution on the Move", IEEE communications magazine, Vol.40, No.4, Apr 2002, pp.104-111.

[3] Y. Kim, B. J. Jeong, J. Chung, C-S. Hwang, J. S. Ryu, K-H. Kim, and Y. K. Kim, "Beyond 3G: Vision, Requirements, and Enabling Technologies", IEEE communications magazine, Vol.41, No.3, Mar 2003, pp.120-124.

[4] I. F. Akyildiz, J. McNair, J. S. M. Ho, H. Uzunalioglu, and W. Wang, "Mobility Management in Next-generation wireless systems", Proceedings of the IEEE, Vol.87, No.8, August 1999, pp.1347-1384.

[5] K. Pahlavan and P. Krishnamurthy, *Principles of Wireless Networks*, Prentice Hall, 2002.

[6] D.Johnson, "Mobility support in IPv6", draft-ietf-mobileip-ipv6-19.txt, October 2002.

[7] T. Narten, E. Nordmark, and W. Simpson, "Neighbor Discovery for IP Version 6 (IPv6)", RFC2461, December 1998.

[8] A. T. Campbell, J. Gomez, S. Kim, and C-Y. Wan, "Comparison of IP Micromobility Protocols", IEEE Wireless Communications Magazine, Vol.9, No.1, Febrary 2002, pp.72-82.

[9] E.Rosen, "Multiprotocol label switching architecture", RFC3031, January 2001.

[10] H.Soliman, "Hierarchical Mobile IPv6 mobility management", draft-ietf-mobileip-hmipv6-07.txt, October 2002.

[11] R. Koodli, "Fast handovers for Mobile IPv6", draft-ietf-mobileip-fast-mipv6-05.txt, September 2002.

[12] A. G. Valko, "Cellular IP: A New Approach to Internet Host Mobility", ACM SIGCOMM Computer Communication Review, Vol.7, No.4, August 2000, pp.42-49.

[13] R. Ramjee, T. L. Porta, S. Thuel, K. Varadhan, and S. Wang, "HAWAII-A Domain-based Approach for Supporting Mobility in Wide-Area Wireless Networks", in IEEE 7th International Conference on Network Protocols, Toronto, Canada, Oct 31-Nov 3, 1999.

[14] E. Gustafsson, A. Jonsson, and C. E. Perkins, "Mobile IP Regional Registration", Internet draft, draft-ietf-mobileip-reg-tunnel-03, July 2000.

[15] C. E. Perkins, *Mobile IP: Design Principles and Practices*, Addison-Wesley, 1998.

[16] N. Montavont, T. Noel, "Handover management for mobile nodes in IPv6 networks", IEEE Communication Magazine, Vol.40, No.8, August 2002, pp.38-43.

[17] Z. Ren, C-K Tham, C-C Foo, and C-C Ko, "Intergration of Mobile IP and Multi-Protocol Label Switching", IEEE Internation Conference on Communications Networks (MWCN 2003), Singapore, Oct 27-29, 2003.

[18] T. Yang, D. Makrakis, "Hierarchical Mobile MPLS: supporting delay sensitive applications over wireless internet", International Conference on Info-Tech & Info-Net (ICII 2001), Beijing, China, October 2001.

[19] T. W. Um, J. K. Choi, "A study on path re-routing algorithms at the MPLS-based Hierarchical Mobile IP network", TENCON'2001, August 2001.

[20] P. Tan, "Recommendations for achieving seamless IPv6 handover in IEEE 802.11 networks", draft-paultan-seamless-ipv6-handover-802-00.txt, February 2003.

[21] Q. Y. Xie, H. M. Nguyen, Winston K.G. Seah, "Handoff supporting QoS in MPLS-based Hierarchical Mobile IP networks", in IEEE 58th Vehicular Technology Conference (VTC Fall 2003) , Orlando, FL, USA, Oct 6-9, 2003.

[22] P. Tan, Q. Xie, H. M. Nguyen, J. Shankar, W. K. Seah, "Achieving Seamless IP Handover and Dynamic Neighborhood Discovery in IEEE 802.11 Networks", in 5th IFIP TC6 International Conference on Mobile and Wireless Communications Networks (MWCN 2003), Singapore, Oct 27-29, 2003.

[23] H. M. Nguyen, F. Li, and Q. Xie, "Integration of Micro-Mobility with QoS in IP/MPLS-based Radio Access Networks", in IEEE 57th Vehicular Technology Conference (VTC Spring 2003) , Jeju, Korea, April, 2003.

[24] U. J¨onsson, F. Alriksson, T. Larsson, P. Johansson, and G. Q. M. Jr., “MIPMANET - Mobile IP for Mobile Ad Hoc Networks”, ACM MobiHOC'00, Boston, Massachusetts, August 2000, pp.75–85.

[25] Y. Sun, E. M. Belding-Royer, and C. E. Perkins, “internet connectivity for ad hoc mobile networks”, International Journal of Wireless Information Networks special issues on mobile ad hoc networks, Vol.9, No.2, April 2002, pp.75-88.

[26] Y. C. Tseng, C. C. Shen, and W. T. Chen, “Mobile IP and ad hoc networks: an integration and implementation experience”, IEEE Computer, 2002.

[27] J. Xi and C. Bettstetter, “Wireless Multihop Internet Access: gateway discovery, routing, and addressing”, proceeding of the International Conference on Third Generation Wireless and Beyond (3Gwireless’02), San Francisco, CA, USA, May 28-21, 2002.

[28] R. Wakikawa, J. T. Malinen, C. E. Perkins, A. Nilsson, and A. J. Tuominen, “Global connectivity for Ipv6 mobile ad hoc networks”, draft-wakikawa-manet-globalv6-02.txt, Nov.2002.

[29] E. Nordstrom, P. Gunningberg, C. Tschudin, “Gateway forwarding strategies in ad hoc networks”, in proceeding of the 4th Scandinavian Workshop on Wireless Ad-hoc Networks, May 4-5, 2004.

[30] C. Perkins, E. Belding-Royer, and S. Das, “Ad hoc On-Demand Distance Vector (AODV) Routing”, RFC3561, July 2003.

[31] P. Johansson, T. Larsson, N. Hedman, B. Mielczarek, and M. Degermark, “Scenario-based performance analysis of routing protocols for mobile ad-hoc networks”, in proceding of the 5th Annual International Conference on Mobile Computing and Network, Aug. 1999.

[32] 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”, in proceeding of the 4th Annual ACM/IEEE International Conference on Mobile computing and Networking, Oct 1998, pp. 85-97.