Multicast-Based Mobility: A Novel Architecture for Efficient Micromobility

Ahmed A.-G. Helmy, Muhammad Jaseemuddin, *Member, IEEE*, and Ganesha Bhaskara, *Member, IEEE*

***Abstract*—Handover performance is very important when evaluating IP mobility protocols. If not performed efficiently, handover delays, jitters, and packet loss directly impact application performance. We propose a new architecture for providing efficient handover, while being able to coexist with other protocols. We propose a paradigm for multicast-based micromobility (M&M), where a visiting mobile is assigned a multicast address to use while moving within a domain. The multicast address is obtained using algorithmic mapping, and handover is achieved using multicast join/prune mechanisms.**

This paper outlines a framework for the design and evaluation of micromobility protocols. We define a suite of protocols (called candidate access router set) to enable multiple-access routers to receive traffic for the mobile node. By changing the number of such routers, timing, and buffering parameters, the protocol may be fine-tuned for specific technologies (e.g., 802.11) and handover scenarios.

Extensive NS-2 simulations are used to compare M&M with other micromobility schemes—cellular Internet protocol (CIP) and handoff-aware wireless access Internet infrastructure (HAWAII). For proactive handover scenarios, our results show that M&M and CIP show lower handover delay and packet reordering than HAWAII. M&M, however, handles multiple border routers in a domain, where CIP fails. Also, for scenarios of reactive handover and coverage gaps M&M clearly outperforms CIP and HAWAII.

***Index Terms*—Cellular Internet protocol (CIP), handoff-aware wireless access Internet infrastructure (HAWAII), micromobility, multicast-based micromobility (M&M), performance evaluation.**

I. Introduction

THE GROWTH OF mobile communications necessitates efficient support for Internet protocol (IP) mobility. IP mobility addresses the problem of changing the network point-of-attachment transparently during movement. When the mobile node (MN) moves away from its current network point-of-attachment, handover is invoked to choose another suitable point-of-attachment. In such an environment, handover latency and mobility dynamics pose a challenge for provisioning of efficient handover.

Manuscript received February 16, 2003; revised October 1, 2003. This work was supported in part by the Defense Advanced Research Projects Agency (DARPA) under Grant G408, in part by the National Science Foundation (NSF) Career Award 0134650, and in part by Intel, Pratt and Whitney, and Nortel.

A. A.-G. Helmy and G. Bhaskara are with the Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089-2562 USA (e-mail: helmy@ceng.usc.edu; bhaskara@usc.edu; http://ceng.usc.edu/~helmy).

M. Jaseemuddin is with the WINCORE Laboratories, Department of Electrical and Computer Engineering, Ryerson University, Toronto, ON M5B 2J1, Canada (e-mail: jaseem@ee.ryerson.ca).

Digital Object Identifier 10.1109/JSAC.2004.826002

Several studies [1], [8] show that mobile IP (MIP) [3], the proposed standard, has several drawbacks ranging from triangle routing and its effect on network overhead and end-to-end delays, to poor performance during handover due to communication overhead with the home agent. Several micromobility approaches attempt to modify some mechanisms in MIP to improve its performance [4], [5]. However, as we will show, such approaches suffer from added complexity and, in general, do not achieve the best handover performance.

We follow a different approach to IP mobility using multicast-based mobility (M&M) [1]. In such paradigm, each MN is assigned a multicast address to which it joins through the access routers (ARs) it visits during its movement. Handover is performed through standard IP-multicast join/prune mechanisms. Such approach, however, is not suitable for interdomain IP mobility, for several reasons. First, the architecture requires ubiquitous multicast deployment, which is only partially supported in today's Internet. M&M should be designed for incremental deployment, and to allow coexistence with other IP mobility protocols. Second, the multicast state kept in the routers grows as the number of MNs becomes larger. This problem may be alleviated using state aggregation [38]. Third, allocating a globally unique multicast address for every MN requires a global multicast address allocation scheme, and wastes multicast resources. Furthermore, MNs incur security delay with every handover, which may overshadow architectural mechanisms that attempt to reduce handover delays.

To alleviate these problems, we propose new schemes for intradomain M&M that allow for incremental deployment. In this architecture, a MN is assigned a multicast address within a domain for use with micromobility. The allocated multicast address is locally scoped (i.e., unique only domain-wide). This allows for domain-wide address allocation schemes. Packets are multicast-tunneled to the MN within the domain. The multicast address of a mobile does not change throughout its movement within the domain. This allows for lighter-weight security during handover, as it is used for micromobility (i.e., intradomain).

In this paper, we present two different approaches to M&M, one approach is based on mobility proxies and the other based on a novel scheme for algorithmic mapping. We compare the two approaches and show that algorithmic mapping provides a more scalable and robust approach, and we also develop efficient, yet simple, mechanisms to realize it.

Another main contribution of this paper is providing a framework for the design of micromobility protocols. We introduce a flexible architecture with a suite of protocols that allow groups of ARs, called candidate access router set (CAR-set), to receive

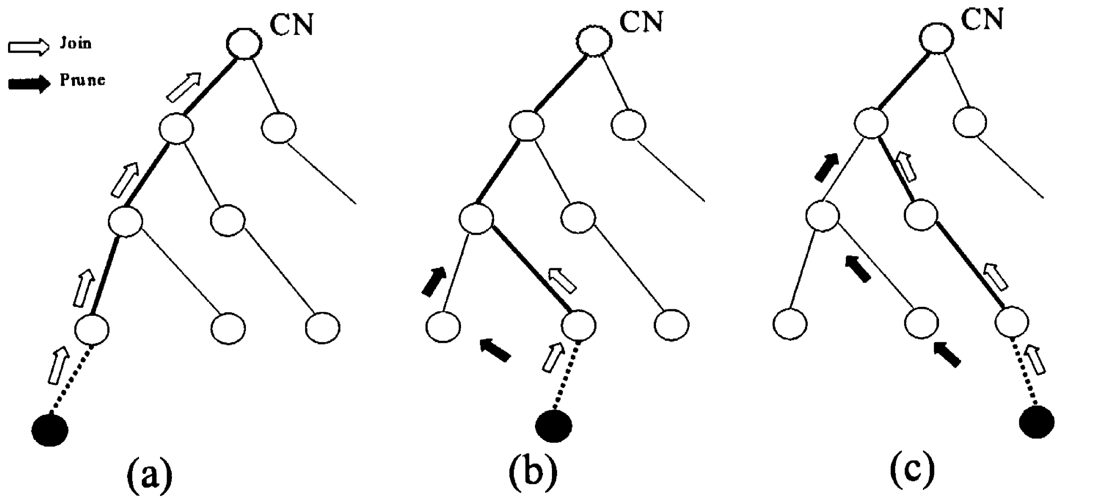

Fig. 1. M&M. As the MN moves, as in (b) and (c), the MN joins the distribution tree through the new location and prunes through the old location.

packets destined to the MN at various times. By changing the number (and timing) of the ARs that simultaneously receive the packets, and buffer parameters, the resulting protocol may be fine-tuned to optimize performance for given scenarios.

We conduct extensive simulations to compare the handover performance of our approach with other routing-based micro-mobility schemes. The proactive handover performance results show that our scheme performs as well as cellular Internat protocol (CIP) and much better than HAWAII. Furthermore, it handles multiple border routers (BR) in a domain where CIP fails. For reactive handover M&M has a clear edge over CIP. Improvement in handover performance is achieved by a slight increase in bandwidth overhead in the wired network.

We provide detailed and systematic analysis of such overhead, and indicate favorable parameter settings of our protocols for various handover scenarios.

The rest of the paper is outlined as follows. Section II introduces M&M. Section III gives overview of the intradomain architecture, and discusses the proxy-based approach. Section IV describes the algorithmic mapping approach in detail. Section V describes the proposed handover framework and the concept of CAR-set. Section VI gives evaluation and comparison results. Section VII discusses related work. Section VIII concludes the paper.

II. Multicast Based Mobility (M&M)

Performance during handover is a significant factor in evaluating performance of wireless networks. IP-multicast [25], [2] provides efficient location independent packet delivery. The receiver-initiated approach for IP-multicast enables receivers to join to a nearby branch of an already established multicast tree. M&M [1], [8] uses this concept to reduce latency and packet loss during handover.

In M&M, each MN is assigned a multicast address.

The MN, throughout its movement, joins this multicast address through locations it visits. Correspondent nodes (CNs) wishing to send to the MN send their packets to its multicast address, instead of unicast. Because the movement will be to geographical vicinity, it is highly likely that the join from the new location, to which the mobile recently moved, will traverse a small number of hops to reach the already-established multicast distribution tree. Hence, performance during handover improves considerably. An overview of this architecture is given in Fig. 1. As the MN moves, it joins to the assigned multicast address through the new AR. Once the MN starts receiving packets through the new location, it sends a prune message to the old AR to stop the flow of the packets down that path, thus completing the smooth handover process. In spite of its promise, we believe that many issues need to be addressed to realize M&M in today's Internet. These issues include scalability, multicast address allocation, multicast deployment and security.

Scalability of Multicast State: The state created in the routers en-route from the MN to the CN is source-group (S, G) state. With the growth in number of MNs, and subsequently, number of groups (G), the number of states kept in the routers increases.

In general, if there are "x" MNs, each communicating with "y" CNs on average, with an average path length of "l" hops, then the number of states kept in the network is "$x.y.l$" states. Clearly, this does not scale well for the number of hosts in the Internet.

Multicast Address Allocation: Interdomain M&M requires each MN to be assigned a globally unique multicast address. Using a global multicast address for each MN may be wasteful and requiring uniqueness may not be practical.[1]

Ubiquitous Multicast Deployment: Interdomain M&M assumes the existence of interdomain multicast routing, which is not widely deployed. We believe, however, that incremental deployment and interoperability should be an integral part of any architecture for IP mobility.

Security Overhead: Security is critical for mobility support, where continuous movement of mobiles is part of the normal operation. Such setting is prone to remote redirection attacks, where a malicious node redirects to itself packets that were originally destined to the mobile. The problem is even more complex with multicast, where any node may join the multicast address

[1]Multicast address allocation is an active area of research [15]. We envision the number of MNs to grow tremendously.

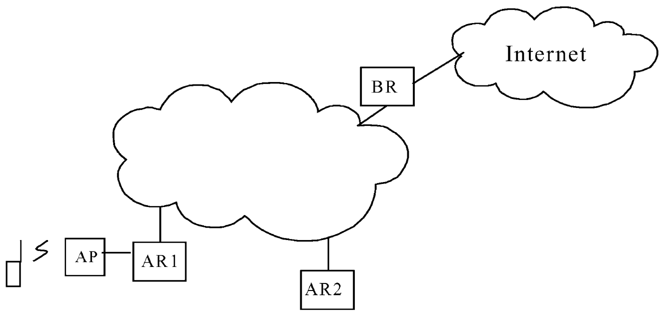

Fig. 2. Reference mobility domain network.

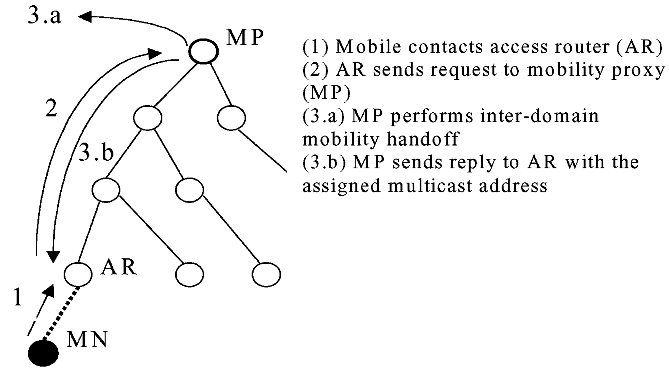

Fig. 3. Event sequence as the MN moves into a domain.

as per the IP-multicast host model. These security measures are complex and may incur a lot of overhead. If such measures are invoked with every handover, however, it may overshadow the benefits of efficient handover mechanisms.[2]

To address the above issues, we propose a new approach for intradomain M&M.

III. Intradomain Architectural Overview

Our intradomain architecture, a MN is assigned a multicast address to which it joins while moving. The multicast address, however, is assigned only within a domain and is used for *micro* mobility. While moving between domains, an interdomain mobility (e.g., MIP) protocol is invoked. In MIP [3], every MN is assigned a home address and home agent (HA) in its home subnet. When the MN moves to a foreign subnet, it acquires a care-of-address (COA) through a foreign agent (FA). The MN informs the HA of its COA through a registration process. Packets destined to the MNs home address are intercepted by the HA in the home subnet, then it tunnels them to the MNs COA. This is known as triangle routing. We will use the MIP model to discuss interdomain routing in the following sections.

Several mechanistic building blocks are needed to realize our proposed architecture. First, when the mobile moves into a new domain it is assigned a multicast address. What is the address allocation scheme? Second, packets destined to the mobile are multicast-tunneled by an encapsulator to the MN. How are the encapsulator(s) selected and where are they placed? To answer these questions, we investigate two different approaches: 1) *proxy*-based architecture and 2) *algorithmic mapping* architecture. We first describe below the reference architecture of a mobility domain, and then discuss the proxy-based scheme. We discuss the algorithmic mapping scheme in Section IV.

A. Reference Architecture

We consider an IP network for a single domain, as shown in Fig. 2. The network is connected to the Internet through BRs. An access point (AP) is the radio point of contact for a MN. A number of APs are connected to an AR. From the AR's point of view, each AP is a node on a separate subnet. When a mobile moves from one AP to another without changing AR is an intra-AR handover case that can be specific to AR implementation and is not considered in this paper. In this paper, we discuss the inter-AR handover when a mobile is said to handover from old AR ($\mathrm{AR}_{\mathrm{old}}$) to new AR ($\mathrm{AR}_{\mathrm{new}}$). We use this terminology throughout the rest of the paper.

When a mobile is connected to a domain it is assigned a unicast care of address that is unique within the domain, called regional care-of-address (RCOA). It is also assigned a multicast care-of-address (MCOA). The RCOA is a globally routable address, which is used to route packets (destined to the mobile) in the Internet to the domain. The MCOA is used for routing packets within the domain, hence, the scope of the MCOA is local to the domain and there is no need to assign a new unicast COA at every subnet when the mobile moves from one AR to another. The RCOA is a unique unicast address that can be assigned from any subnet in the domain not necessarily from the subnet of the first AP, where the mobile is connected when it moved to the domain. This makes the RCOA assignment simple because it does not run into the risk of address exhaustion at a subnet that is the first subnet for a large number of mobiles when they moved to the domain [42].

If the domain is not the home domain for the MN, then it performs interdomain handover using interdomain mobility protocol such as MIP. In this case, it registers the RCOA with the home agent.

Before explaining the address allocation and management we describe the proxy-based approach and discuss the problems associated with it.

B. Proxy-Based Architecture

When a MN moves into a new domain, it contacts its AR. The AR performs the necessary per-domain authentication and security measures, and then assigns RCOA for the MN. As shown in Fig. 3, the AR then sends a *request* message to the mobility proxy (MP) to obtain a multicast address for the visiting MN. The request message includes the home address of the MN and its home agent's address. Upon receiving the request the MP performs two tasks. The first is to register on behalf of the MN its own address as COA with the MNs home agent. The second task is to assign a multicast address for the visiting MN, send a *reply* message to the AR and keep record of this mapping. The mapping is used for packet encapsulation later on. In this scheme, the MP remains transparent to the MN, which makes

[2]Providing a comprehensive security solution for IP mobility is beyond the scope of this work. We believe, however, that our scheme relaxes security requirements during handover.

the placement of MPs within the domain flexible without notifying every MN.

Once this step is complete, the visiting MN joins the assigned multicast address (G). The joins are sent to the proxy-group pair (MP, G) and are processed as per the underlying multicast routing. The MN continues to move within the domain using the same multicast address. The scope of the assigned multicast address is local to the domain. Handover is performed using standard join/prune mechanisms and only lightweight intradomain security is required in this case.

Packets sent to the MNs home address are tunneled by the HA to the MP using interdomain mobility. The packets are then encapsulated by the MP, based on the recorded mapping, and sent down the multicast tree to the MN. The MN uses the unicast RCOA for sending packets. To avoid single-point-of-failure scenarios multiple MPs are used. These MPs are typically placed at the border of the domain or at the center of the network.[3] An algorithm similar to [24] may be used for dynamic MP liveness and election mechanisms.

Several issues need to be addressed in the above architecture. First, the MPs need to maintain unicast-to-multicast address mapping for all visiting MNs. The scalability of such a scheme is of question. Second, complex robustness algorithms are needed to maintain MP liveness information, requiring initial configuration and setup. Third, the service disruption effect of MP failure is not clear. Since the MP registers its own address with the home agent and is used to encapsulate incoming packets, this introduces a third-party-dependence problem that is undesirable. In addition, MPs should run a multicast address allocation scheme to ensure collision-free address assignment. This scheme needs modification in MIP to handle indirect home registration via MP.

To address these problems we propose a novel approach based on *algorithmic mapping* that obviates the need for explicit unicast-to-multicast mapping, and eliminates the need for complex address allocation.

IV. Algorithmic Mapping Architecture

This section provides detailed address management, duplicate address detection, and inter-AR handover.

A. Overview

In this scheme, we assume there is a one-to-one mapping between an RCOA and MCOA. When a MN moves into a new domain it is assigned RCOA in the domain and it performs interdomain handover; i.e., it registers the RCOA with its home agent. The AR automatically infers the multicast address (MCOA) for the MN from the assigned unicast address (RCOA) through a straight forward *algorithmic mapping*, described later in this section. The AR then triggers a Join message for MCOA to establish the multicast tree. Packets destined to the MNs home address are tunneled to its RCOA by the HA. When these packets arrive in the foreign domain they are identified by the BR as being destined to a node on a subnet in the domain. As shown in Fig. 4, the BR maps the destination unicast address to the multicast address and transmits the packets to the MN down

[3]Network center are nodes with min(max distance) to any other node [26].

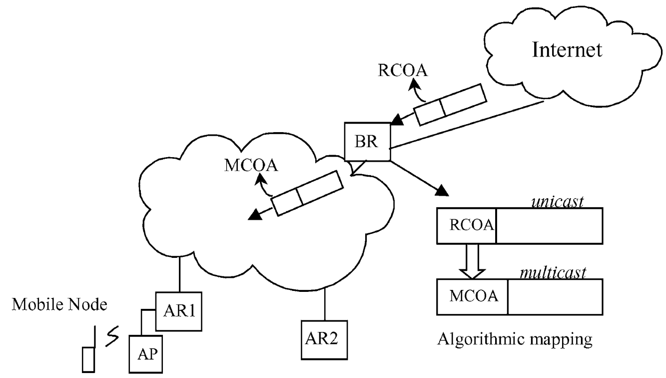

Fig. 4. Architectural view: A packet is unicast to the RCOA and arrives at the BR for the MN. The BR intercepts the packet and performs algorithmic mapping from the RCOA to MCOA. The packet is then multicast within the domain.

the multicast tree. The serving AR changes the destination address from multicast to the unicast address. The mapping is constrained to allow restoration of RCOA. Since the destination address is modified twice within the network and restored to the RCOA by the AR, the packet does not cause security association violation at the MN.

This architecture provides several advantages over the proxy-based approach. First, it avoids the third party dependence on the MP. Second, since algorithmic mapping is used, no explicit RCOA-MCOA mapping is kept or maintained at any node in the network, which solves the mapping scalability problem and provides a more robust mechanism. The mapping function can be performed by any node in the network including BR. Hence, this scheme allows forming multicast shared tree rooted at BR, where all the multicast enabled routers keep $(*, \mathrm{G})$ state. In this case, a router keeps the amount of state proportional to the number of MNs connected through the router.

B. Address Management

The number of multicast addresses required is proportional to the number of MNs in the domain. The scope of an MCOA is local to the domain where it is used. Both IPv4 and IPv6 multicast addressing provides facility to define scope within the address [32], [43]. In this section, we describe a mapping scheme for both IPv4 and IPv6 address architecture.

1) IPv4 Addressing: The standard IPv4 address is composed of three subblocks, namely, Net ID, Subnet ID, and Host ID [44]. The Net ID is common for the addresses assigned within the domain, and is mainly used for routing outside the domain. Hence, losing the Net ID when mapping from RCOA to MCOA is restorable at the AR. The MCOA is assigned from $239.0.0.0/8$ multicast address block that is set aside for administratively scoped multicast addresses [43]. A simple mapping between RCOA and MCOA is shown in Fig. 5(d).

2) IPv6 Addressing: The standard IPv6 unicast and multicast address architectures [32] are shown in Fig. 5(a) and (b). We modify the group bits to include interface ID as the group ID. The remaining reserved bits of the group ID are ignored by multicast routing. The 64-bit interface ID address space is large enough for all the mobiles within a domain, hence, a single subnet ID, called *m-subnet*, can be set aside for assigning RCOA. We also define a new scope: micromobility scope with

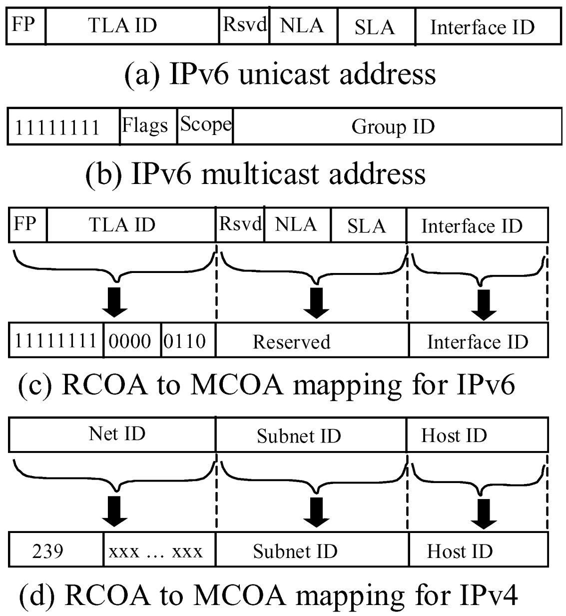

Fig. 5. Algorithmic mapping.

value 0×6. The SLA is a 16-bit long field, used to create local hierarchy and identify subnets [33].

The TLA ID is common to all the addresses assigned within the domain, and is mainly used for routing outside the domain. Hence, losing the TLA ID when mapping from RCOA to MCOA is restorable at the AR. The BR forms the MCOA by replacing the ⟨FP, TLA ID⟩ bits of the RCOA with the multicast ⟨FP, flag (0000), scope (0110)⟩ values. When the packet arrives at the serving AR it recovers the RCOA by replacing the ⟨FP, flag, scope⟩ bits with the ⟨FP, TLA ID⟩ bits, as shown in Fig. 5(c). This provides a simple, yet very efficient and unique *algorithmic mapping*.

The mobile acquires RCOA on the m-subnet through either DHCP [34] or auto-configuration. For auto-configuration m-subnet address needs to be advertised by every AR in their router advertisements (RA) [37]. The auto-configuration requires duplicate address detection (DAD) [35] on every subnet. In our scheme, the mobile obtains RCOA and MCOA once it is connected to the network. We propose a scheme in [39] that detects address duplication within the m-subnet, which is performed once at the AR during initial address assignment. The mobile afterward is able to move freely without running DAD at any other AR. In any case, when a mobile first connects to the network it performs high latency operations such as reconfiguration of the node itself and in some cases interdomain handover, which overshadows the duplicate address resolution latency.

C. Implementation Outline

The M&M solution using algorithmic mapping can be implemented using multicast shared tree, e.g., PIM-SM [2], where the shared tree is rooted at the rendezvous point (RP). For M&M implementation the RP can be co-located at a BR. All CAR-set member routers join the shared tree to the BR. Alternatively, the RP can be any router in the network, and the BR simply acting as the source for all the packets entering the network deliver them to the RP to forward the packets down the shared tree. By separating the node performing algorithmic mapping (BR) and the root of the shared tree (RP), we can accommodate multiple BRs in a network. The flexibility comes at the expense of possible reduction of routing efficiency, because packets are first tunneled to the RP and then delivered to the MN through the multicast tree. To alleviate this situation the BRs may be configured as candidate RP for the MCOA prefix, thus ensuring that one of the BRs becomes the RP. However, both HAWAII and CIP do not handle well the case where a domain contains multiple BRs. In particular, if packets enter the domain through one BR and leave through another BR, routing in CIP fails. M&M relies on the underlying multicast protocol to handle multiple BRs in a domain, which is often the case.

V. Handover Framework

When a mobile moves from one AR to another, a handover event takes place between the two routers. The network-level handover to the new AR involves: 1) the router detection (and possibly a new subnet detection); 2) association at the new AR; and 3) *route repair* that is path setup inside the network to redirect the incoming traffic flow to the new AR. In *proactive handover*, the identification of the new AR is known to the MN *a priori* to its disconnection from the old AR. This advanced knowledge of the new AR can be used to initiate a mechanism for avoiding packet losses and accomplishing a *smooth handover*, i.e., handover with low packet loss. For example, a tunnel between the old and new AR can be setup to move the packets to the new AR as soon as the MN is disconnected from the old AR and before any route repair is performed inside the network [21], [31]. Alternatively, the route repair process may be initiated directly by bicasting packets to both ARs. In *reactive handover*, an abrupt disconnection from the old AR may cause the MN to switch over to the new AR, which may happen in some break-before-make handover model [42]. It occurs when a MN moves out of the coverage area of the old AR; due to obstacles, lack of cell overlap, etc.; and then enters the coverage area of the new AR. Some wireless technologies, e.g., IEEE 802.11, only support reactive handover. The route repair in this case can only be initiated from the new AR, hence tunneling and bicasting (from the old AR) cannot avoid packet loss.

Dealing with reactive handover requires predicting the new AR and setting up path proactively to it. In some cases, the new AR can be predicted with some degree of accuracy, for example when the vehicles move in a known trajectory. To improve the degree of accuracy a set of potential new access routes can be formed. Since bicasting is limited to sending packets to one new AR, it cannot be used for proactive path setup in many circumstances, and extending bicasting to send packets to multiple ARs is basically reinventing multicasting. The M&M, on the other hand, by virtue of being a multicast protocol, is able to send packets to multiple (two or more) ARs simultaneously. Multicasting allows proactive path setup to the new AR before the mobile is actually connected to it. This can minimize packet losses in reactive handover, where bicasting fails. Moreover, since bicasting is a special case of multicasting, multicasting-based solution, e.g., M&M, performs equally

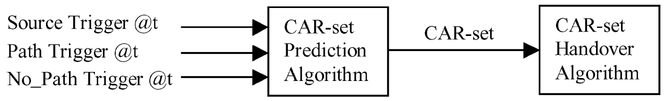

Fig. 6. M&M handover framework.

well for achieving proactive handover. Thus, M&M provides a comprehensive mechanism of achieving smooth handover in all handover situations. In this section, we describe one handover scheme where proactive path setup is used to achieve smooth reactive handover.

A. Framework

We define the set of potential new ARs as the candidate access router set (CAR-set). The routers in the CAR-set are adjacent to the serving AR. The adjacency can be established based on the adjacency of the radio coverage area of the serving AR, for example in the case of cellular wireless network. The serving AR is called the Head of the CAR-set. Thus, there is a unique CAR-set defined for every AR. For example, in Fig. 7, AR1–AR7 constitute a CAR-set for AR1, which is the serving AR for the mobile. The CAR-set members join the multicast group, MCOA, to receive the packets destined to the MN. Hence, the mobile can move to any router in the CAR-set without interruption in the packet flow.

The membership of CAR-set is dependent upon a number of factors, such as the handover type and the desired prediction accuracy, among others. For example, if the new AR is known *a priori* of handover, then the CAR-set contains only one member that is the new AR. The membership is determined by the CAR-set prediction algorithm. Fig. 6 shows a CAR-set based handover framework. The CAR-set prediction algorithm is triggered at certain time t_e by an event e, where e is called as the trigger and t_e the trigger time. In Fig. 6, three triggers are shown to illustrate the proactive and two reactive handover cases. They are defined as follows.

1) *Source Trigger*: As defined in [31], the source trigger is generated at the old AR prior to the mobile's disconnection. It indicates the onset of handover condition and contains the identification of the new AR.
2) *Path Trigger*: It is generated at the old AR anytime, while the mobile remains connected to the old AR. It contains the path vector of the mobile indicating future ARs (cells) to be visited. For example, routes taken by vehicles on the road are either often known or predicted with good accuracy.
3) *No-Path Trigger*: It is generated at the old AR anytime, while the mobile remains connected to the old AR. It contains no indication for future ARs (cells) to be visited by the mobile. This indicates an extreme case for proactive handover where all the adjacent routers (cells) of the old AR are potential new AR.

The trigger contains hints for the potential new AR(s) and is used to launch a particular prediction algorithm. Multiple triggers may be used in a single system, with different triggers under different handover conditions. The CAR-set member routers remain joined to the multicast tree between the trigger time t_e and the time the mobile is connected to the new AR, called t_c.

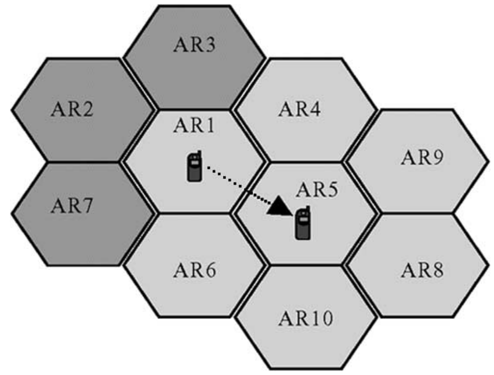

Fig. 7. Handover across CARS.

Hence, the trigger time t_e plays a significant role in achieving the quality of handover and system efficiency. This is further discussed in the Section V-C.

B. CAR-Set Handover Algorithm

The CAR-set handover algorithm described in this section accomplishes proactive route repair by initiating multicast join from the old AR ahead of link-level handover. All member routers of the CAR-set join the multicast tree corresponding to the mobile's MCOA. They remain joined to the tree as long as the mobile is not connected to the new AR. The prune process is initiated by the new AR after the mobile achieves a stable network-level connectivity in its coverage area. To simplify the discussion of the handover algorithm in this section we consider that the CAR-set is formed as a result of *No-Path* trigger and it constitutes all adjacent routers of the serving router that can be determined statically. A site-local multicast group address is assigned to the CAR-set, called the CAR-set group address (CGA). Every AR that is a member of the CAR-set must join the corresponding CGA, which serves as a control channel for the members to exchange the control signals. For example, in Fig. 7, all the ARs surrounding AR1 join CGA1 to become members of AR1's CAR-set (CGA1). Similarly, AR1 must also join six other CAR-sets corresponding to adjacent routers AR2–AR7.

We define four control signals as follows.

J-message causes the receiving router to *join* the multicast group identified in the message. It is generated by the router receiving the trigger at time t_e and is sent to the CAR-set.

1) *HO* message causes the receiving router to generate *L-message*. It is transmitted by the new AR to the old RA. Its parameters include the mobile's RCOA and MCOA.
2) *L-message* causes the receiving router to *leave* the multicast group identified in the message. It is generated by the old AR after receiving the *HO* message and is sent to the CAR-set group.
3) *HOA* message is the acknowledgment of HO message causing the receiving router to stop retransmitting HO message. It is transmitted by the old AR to the new AR.

We explain the handover algorithm by using the example depicted in Fig. 7. Consider the MN moving from AR1 to AR5. When connectivity is established between the MN and AR5,

MCOA	Serving Router	CGA	State
MG	*AR1*	*CGA1*	*Joined*

(a)

MCOA	Serving Router	CGA	State
MG	*AR5*	*CGA5*	*Joined*

(b)

Fig. 8. Handover state at AR4 (a) when MN1 is connected to AR1 and (b) after MN1 moved to AR5.

AR5 checks for an active trigger corresponding to the CAR-set join for the MN. If it finds an active trigger, then it multicasts a J-message ⟨MCOA⟩ to the members of its CAR-set (CGA5) requesting them to join the mobile's MCOA. In this case, t_e at AR5 is the time when the MN is connected to AR5. It then sends HO ⟨RCOA, MCOA⟩ message to AR1 to initiate the prune process. When AR1 receives HO message it multicasts an L-message ⟨MCOA⟩ to members of its CAR-set (CGA1) requesting them to leave the MCOA. It also sends HOA message to the new AR acknowledging the reception of HO message. The reception of HO message by the old AR signifies that the mobile is connected to the new AR. The initiation of L-message at the old AR cannot be triggered by the reception of J-message from the new AR. The distinct HO message is required to communicate this event to the old AR because the old AR is not guaranteed to be a member of the CAR-set of the new AR so to receive its J-message. Furthermore, the CAR-set prediction algorithm at the new AR may be triggered at some later time causing unnecessary delay in the prune process resulting in the multicast overhead traffic as discussed in the next section.

Although the ordering of $J \rightarrow HO \rightarrow L$ messages ensures that L-message is initiated after the J-message, the order of message reception, however, is not guaranteed to both CAR-sets. Depending upon the order of arrival of J and L messages at an AR that is a member of both CAR-sets, it may leave the MCOA whereas it is supposed to have remained joined to that group. To ensure consistency between the Join and Leave messages we introduce the following mechanism. Each AR maintains handover state corresponding to its CAR-set membership status in a 4-tuple ⟨MCOA, serving AR (SR), CGA, state⟩ table. The table contains an entry corresponding to every mobile roaming in a CAR-set of which the AR is a member. There are two states defined: *Joined* and *Left*. The rules for updating the table specify that an AR only accept L-message for a MCOA, if the source of the L-message matches the SR in the MCOAs entry (i.e., the AR has joined the MCOA on the request of the same SR). Otherwise, the L-message is discarded. The AR accepts all J-messages and creates/updates the related MCOA entry to include the source of the J-message (as the SR), CGA to the SRs CGA (as the entry's CGA), and the state to *Joined*.

Consider the example shown in Fig. 7. Assume that the mobile's MCOA is MG and after power up in the domain it connects to AR1, which then multicasts a J-message to its CAR-set (CGA1). When AR4 receives the J-message, it joins MG and creates an entry corresponding to the MCOA in *Joined* state as shown in Fig. 8(a). Later, when the MN moves to AR5 it becomes the new serving router. Then, AR5 sends a multicast J-message to its CAR-set (CGA5) followed by a HO message to the old serving router AR1. Since AR4 is a member of both CGA1 and CGA5, it receives both J-message from AR5 and L-message from AR1. After receiving the J-message the table entry is updated as shown in Fig. 8(b). If received after the J-message, the L-message is discarded. Thus, AR4 remains joined to MG. If received before the J-message, however, the L-message may cause AR4 to leave the MG, which interrupts packet flow to AR4 until it receives the J-message and joins the MG group. In case of unintended interruption, it may be minimized by delaying the leave operation. Any delay introduced in the leave operation must be controlled considering the multicast overhead as discussed in the next section. In most cases, the HO message delay is sufficient to minimize the unintended interruption. In case of t_e being some later time after the MN is connected to the new AR, the interruption is intended.

The reliable exchange of HO-HOA messages ensures the initiation of L-message. To deal with the loss of L-message or crash of the SR, a soft-state mechanism is used for maintaining handover state. A timeout is associated with the *Joined* state upon whose expiry the AR prunes the multicast tree and changes the state to the *Left* state. The SR refreshes the *Joined* state at the CAR-set member routers by periodically sending J-message. It can send a single J-message for a number of mobiles having the common CAR-set. To avoid wasting the radio bandwidth due to control traffic caused by the periodic J-message transmissions, the SR can include the timeout for the joined state in the J-message equal to an estimate of the sum of handover latency and remaining dwell time. It then sets up the J-message period equal to less than the timeout value. If it does not receive HO message during that time, then it retransmits the J-message with the new timeout value based on the current estimation of handover latency and the remaining dwell time. This scheme will reduce the number of J-message retransmission for refreshing the *Joined* state. A garbage collection scheme can be employed that periodically checks the table to purge all the entries that are in the Left state and consequently prune the corresponding multicast trees.

C. CAR-Set Multicast Overhead Control

Multicasting packets to the CAR-set improves handover performance at the cost of added overhead of packet replication over wired links leading to the ARs that belong to the CAR-set. The extent of overhead depends upon the network topology, the size of the CAR-set, and the duration for which member routers remain joined to the multicast tree. The CAR-set protocol does not use any bandwidth in the wireless links of the ARs. For a given MN, let the path from the BR to the MN contain L links on average. This is the path from the BR to the serving AR to the MN. Also, let n be the number of ARs other than the serving AR in the CAR-set. Hence, the CAR-set is $\{\mathrm{AR0}, \mathrm{AR1}, \ldots, \mathrm{AR}n\}$, where AR0 is the serving AR. Furthermore, let L_i be the number of links leading from $\mathrm{AR}i$ to the nearest point already branch from the BR to AR0, where $i = 1, 2, \ldots, n$. If we measure the overhead by the number of additional links traversed by the replicated packets, then the overhead is ΣL_i, called Lsum. The ratio Lsum/L gives the measure of additional links carrying replication traffic due to packet replication for a given MN connected to AR0. The upper bound for the total replication traffic on the additional links for AR0 is $m * \mathrm{Lsum}$, where m is the number of MNs connected to AR0.

The replication traffic on a link consumes link bandwidth proportional to $k \cdot b$, where k is the number of ARs for which the link is an additional link carrying replication traffic and b is the wireless bandwidth for each AR. Typically, wireless bandwidth is much smaller than the bandwidth of the wired links and it constrains the traffic (including the replication traffic) over the wired links.

We adopt a two-dimensional approach to reduce the replication traffic by limiting the size of the CAR-set (spatial control) and the duration (temporal control) for which the replication is performed in the network. In this paper, we illustrated the handover algorithm with a simple static CAR-set prediction algorithm using no-path trigger, however more dynamic algorithms can be designed by identifying the highly probable new ARs. The trigger time t_e can be used to reduce the duration of the replication traffic. In situations where t_e can precisely define the handover condition, the temporal control will cause replication traffic for short time. For less precise t_e we present following heuristics that can potentially reduce the overhead significantly.

1) When the old AR detects that the received signal strength (RSS) at the MN is below certain threshold defined to indicate the imminent handover condition, it then triggers the prediction algorithm causing the ARs in the CAR-set to join the multicast group. To avoid packet loss, the handover condition must be detected early enough to provide time margin before the actual handover, to cover the multicast join latency. Once the MN is connected to the new AR, the CAR-set members leave the group as described above. Thus, the overhead due to the replication traffic is reduced to only the fraction of the time during which the CAR-set remains joined to the multicast group, that is only as long as the handover condition exists. This is different from the proactive handover because it does not require precise identification of the new AR.
2) In situations, such as 802.11, where the MN does not report RSS to the AR, lack of a number of acknowledgments over a short time can be used as the trigger for the imminent handover condition.
3) In many cases, the current velocity v of the MN is known, for example, in case of moving vehicles on the road [45], which can be used to compute the expected dwell time as $t_w = v/s$, where s is the size of the coverage area of the AR. The expected departing time, $t_d = t_a + t_w$, can be used to compute the trigger time $t_e = t_d - \Delta$, where Δ is the estimated multicast join latency and t_a is the time when the MN is first connected to the AR.

The combination of temporal and spatial controls can reduce the overhead significantly. For example, the *path* trigger and the t_e estimation using mobile's velocity (as explained above) can be used to reduce the replication traffic. In Section VI, we simulate two variants of the CAR-set protocol and discuss their performance.

VI. Evaluation and Comparison

In order to evaluate the performance of M&M and compare it with other known schemes, and conducted detailed simulations for CIP [20], HAWAII [21] and M&M—the three routing-based mobility solutions.[4] We modified the network simulator, ns-2 [17] to incorporate M&M. We changed the implementation of MN and AR to add mobility detection, handover algorithm, and multicast routing.

Before describing the details of our performance evaluation, we briefly summarize cellular IP (CIP) [20] and HAWAII [21]. Within a domain, CIP installs unicast mobile specific routing entries on the path from the root (or border) router to the local area network (LAN) in which the MN resides. The path is updated as the MN moves from one LAN to another. During handover, the MN may have routing entries to both the LANs, thus enabling the MN to receive packets from both the LANs (called bicasting). HAWAII also installs mobile specific routing entries. It supports a number of handoff optimization schemes. Here, we focus on multistream forwarding scheme (MSF). In MSF, when the mobile gets out of range of an AR, the AR buffers the packets. When the handover occurs, the MN sends a message to the old AR and the old AR forward the buffered packets to the new AR serving the MN.

A. Performance Metrics

We used the following performance metrics to evaluate the performance of M&M and compare it with CIP and HAWAII.

- *Handover delay* is defined as the difference between the time at which the MN received the last packet from the old AR and the first packet from the new AR.
- *Depth of packet reordering* is measured as the maximum difference in the sequence numbers of adjacent packets. This is a rough indicator of the size of the buffer needed to resequence the out of order packets.
- *Packet duplication* is the total number of packets duplicated in a single handover. For bicasting and multicasting schemes, this is measured as the duration for which reordering occurs. Since CBR traffic is used, reordering duration (duration for which the MN is in the radio coverage of more than one BS) gives an estimate of how many packets can be duplicated irrespective of the packet rate at the source.
- *Routing efficiency* is defined as the ratio of the number of hops between the root of the tree and the MN to the number of hops on the shortest path between the two. This gives a qualitative comparison of routing efficiency.
- *Packet loss* is the measure of the number of packets lost during handover.
- *Bandwidth overhead* due to handover is the measure of the extra bandwidth used by the CAR-set based handover

[4]We have also compared our scheme with hierarchical MIP [27] and seamless handoff [31] schemes using route-based analysis. Please refer to [39] for details. As was shown in [39], M&M achieved the min handover delay and min overhead among the three classes.

scheme, over the bandwidth actually used to deliver packets to the MN.

Mobility detection need not necessarily be a part of the micromobility protocol as this can be better achieved with additional information from lower layers.

B. Simulation Scenarios

To study the factors affecting the performance of the micromobility protocols, we simulated a rich set of scenarios including tree topologies of varying depth ranging from 3 to 6. The link bandwidths were fixed at 10 Mb/s for wired links with delays varied from 10 to 5 to 2 ms for all links. Detailed 802.11 models in ns-2 were used. For scenarios involving overlap in wireless coverage, the overlap was set to 30 m. We also simulated scenarios with varying gaps in radio coverage. To take into account the velocity v of MN and the actual gap g in coverage, we consider the time (g/v) for which the MN is out of range of both BSs. Beacons spacing 20 ms apart are used for mobility detection. State timeout of 1 s (as lower bound) is set for the multicast protocol. We have used CIMS extensions of ns-2 that implement CIP and HAWAII.[5] In addition, we developed our own extensions of ns-2 to support M&M. Bicast and CAR-set (passive and triggered) handover mechanisms were employed for M&M. Semi-soft handover and multistream forwarding (MSF) [21], respectively used with CIP and HAWAII. When applicable, both M&M and CIP use bicast technique whereby packets are bicast to both old and new ARs from a crossover point within the network. In contrast, HAWAII uses buffer and forward technique where the old AR buffers the packets and forward them during route repair. We used random mobility model with mobile velocity at 30 m/s to generate the mobility pattern for the MN. A 400 kb/s CBR traffic source placed at the root of the tree, with packet size of 512 bytes and 10 ms/packet was used. To avoid the side effects of mechanisms of other protocols (like congestion control mechanism of TCP) affecting the handover delay and packet delivery performance, we chose CBR over UDP.

C. Simulation Results

We conducted simulations over different topologies, varying parameters like link delays and gaps in radio coverage. Since mobility detection mechanism is not a part of the protocol, simulations were set-up such that mobility detection always succeeded (when there were no gaps in radio coverage) when the MN moved from one AR to another. This was to prevent loss of packets due to failure of the underlying mobility detection scheme. To evaluate scenarios involving gaps in coverage, we consider packet loss and handover delay as the metrics.

Results for different topologies show similar trends. Hence, and for illustration, we only show results for a simple tree topology with depth 3, shown in Fig. 9.

The graphs in Figs. 10–13 for scenarios involving no gaps in radio coverage follow a common format. Each graph shows data for M&M, CIP and HAWAII (in that order from left to right). The x axis shows three sets of data corresponding to link delays of 10 ms, 5 ms, and 2 ms (again from left to right) for each protocol. Path lengths from the fork (crossover) router to old and new ARs vary along y axis. For example, "3, 2" means path length of three hops from the fork router to the old ARs and two hops from the fork router to the new AR. The z axis shows the performance parameters under evaluation.

[5]We used the Columbia IP Micromobility Suite (CIMS) at http://comet.ctr.columbia.edu/micromobility/software.htm

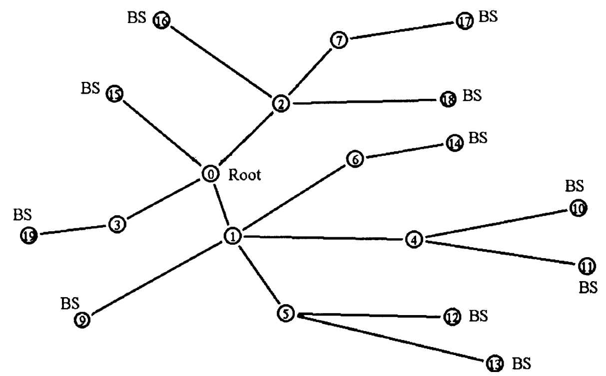

Fig. 9. Simple tree topology.

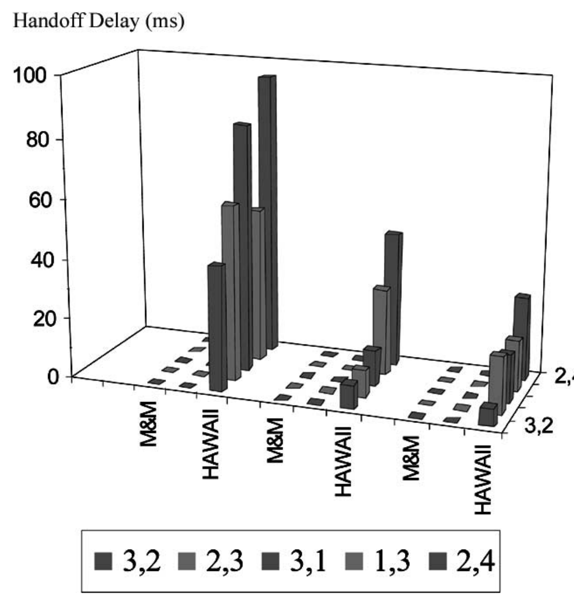

Fig. 10. Handover delay.

Fig. 10 illustrates the handover delays incurred by M&M, CIP, and HAWAII with link delays 10, 5, and 2 ms. From the graphs, we observe that the handover delay for M&M and CIP is small as compared with that of HAWAII. Both CIP and M&M use bicast, which causes smooth handover with negligible handover delay. Whereas, HAWAII using the MSF, a buffer and forward scheme that consistently incurs long handover delays.

Fig. 11 shows the depth of reordered packets. We measured depth of reordering instead of the number of packets reordered because it indicates the size of buffer needed to resequence the out of order packets. It is obvious from the graph that the depth of reordering is small for M&M and CIP, whereas it is large for HAWAII. For M&M and CIP, it hardly goes beyond one, whereas for HAWAII up to five packets are reordered. The out of sequence packets in M&M and CIP are dependent on the difference in the link delays from fork router to old and new ARs. The greater the difference, the greater will be the depth of reordering. (The minor differences in out of sequence depth and reordering duration of M&M and CIP in Figs. 11 and 12 can be attributed to minor implementation differences). In case of HAWAII, the

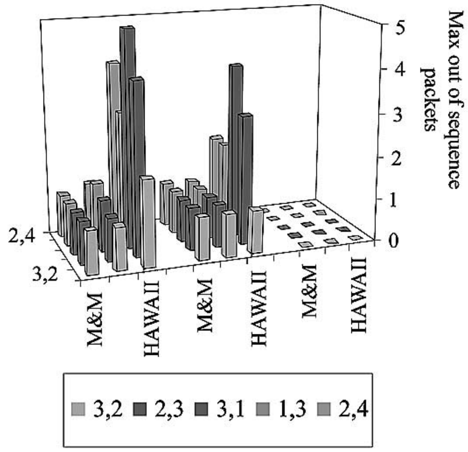

Fig. 11. Maximum difference in sequence numbers of consecutive packets.

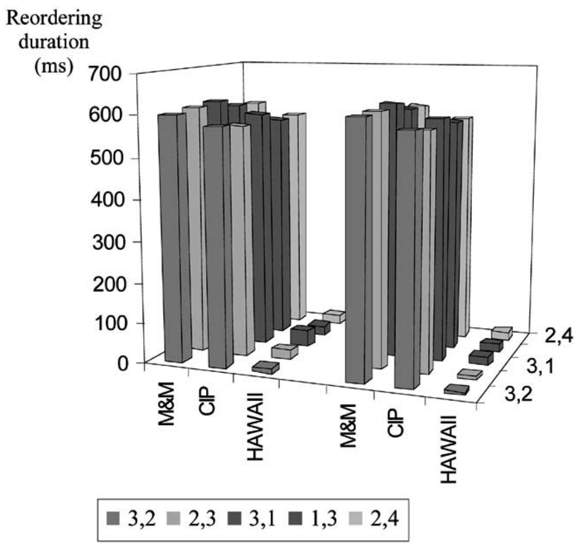

Fig. 12. Reordering duration.

depth is large because the old AR buffers packets and then forward it to the new AR via the crossover router. The crossover router also forward the incoming packets to the new AR at the same time. This results in packets reaching the new AR out of order. Depth of reordering is dependent on the buffering duration and the link delays from the cross over router to the old AR. It is important to observe the duration for which reordering of packets occur. In M&M and CIP, reordering occurs as long as bicasting is done. However, in HAWAII, reordering duration depends on the number of packets buffered at the old AR and the link delay from the old AR to the crossover point.

The duration for which reordering of packets occurs indicates an estimate of the amount of packet duplication caused by a scheme. Fig. 12 shows the reordering duration incurred by the three schemes. As previously mentioned, in case of M&M and CIP, the reordering occurs as long as bicasting lasts causing large number of packet duplication as shown in the figure. For HAWAII, reordering duration depends on the number of packets buffered at the old AR and the link delay from the old AR to the crossover point, which shows relatively low number of duplications.

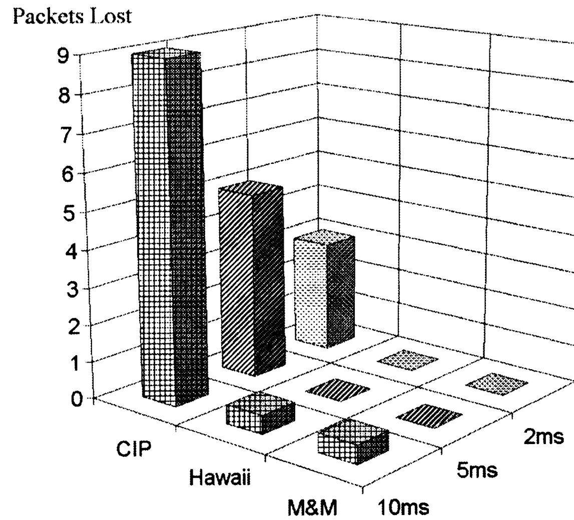

Fig. 13. Packet loss for reactive handover.

M&M uses the multicast path to route packets to the MN. In many cases (and in our simulations), the BR acts as the root (RP) of the multicast tree. CIP uses the shortest path along the reverse path from the MN to the BR to route packets from the BR to the MN. In most cases, the routing in M&M is as efficient as CIP. In the case of HAWAII, routing is a function of topology and node mobility, which is generally less efficient than that of M&M and CIP.

D. *Reactive Handover*

M&M has a clear edge over any other unicast based micromobility protocol (e.g., CIP and Hawaii) during reactive handover. This is mainly due to the CAR-set protocol that enables multicast to multiple ARs. In this section, we evaluate two versions of CAR-set protocols. The first version is the no-path-trigger (described earlier) where packets are sent to the CAR-set ARs as long as the MN is being served by that CAR-set's head AR. This version is used for the reactive handover scenarios (as in 802.11). The second version is source-trigger in which packets are sent to the CAR-set only when the head/serving AR senses MN disconnection. This second version is used for scenarios with coverage gaps. Fig. 13 shows the number of packets lost during no-path-trigger reactive handover (with no gap in coverage). HAWAII and M&M incur very little or no packet loss, whereas CIP incurs significant packet loss. M&Ms and HAWAII's reduced packet loss is due to the buffering of packets. Unlike HAWAII, M&Ms reduced packet loss is not at the expense of higher handover delay, as the packets are buffered in each of the CAR-set ARs (as opposed to only in the old AR in HAWAII). Handover delay is similar to that given in Fig. 10. The handover delay for M&M includes only the delay incurred in MN sending control messages to the new AR over the wireless link. In HAWAII, the handover delay will include the round trip delay from the new AR to the old AR. Thus, the reduced packet loss in M&M is not at the expense of increased handover delay. The handover delay for CIP is less than that of HAWAII, but still significantly greater than that of M&M. This is due to the delay taken for control/route repair messages to setup the new route to the new AR. It is clear that M&M has the best performance in

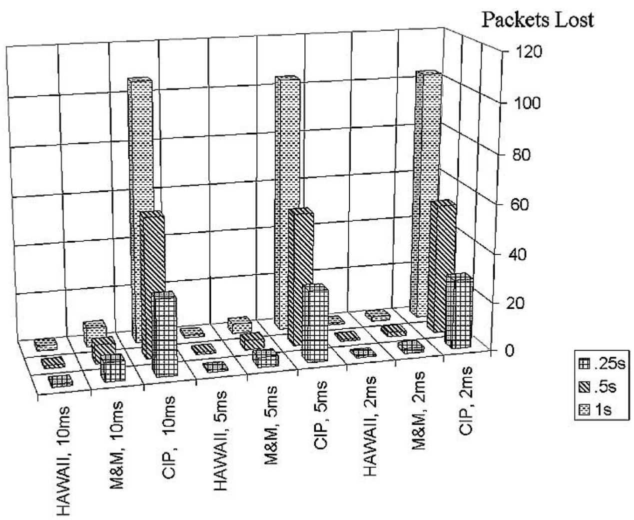

Fig. 14. Packet loss for scenarios with gap in radio coverage.

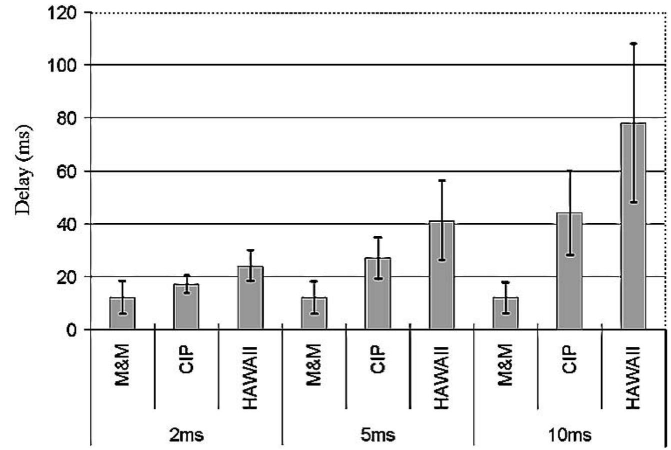

Fig. 15. Handover delays for scenarios with coverage gaps.

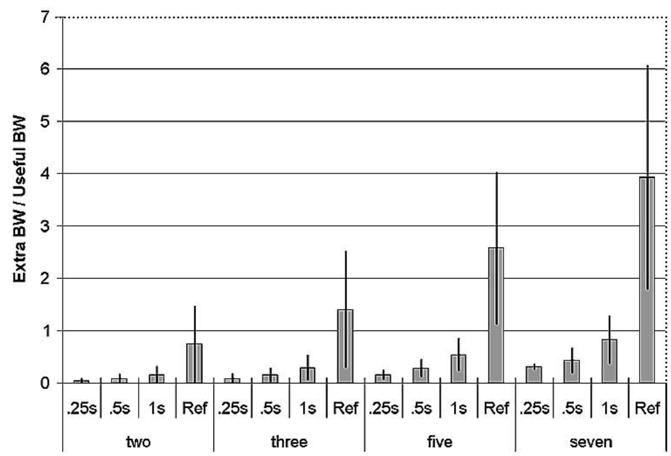

Fig. 16. Overhead for the CAR-set protocols.

terms of both packet loss (clearly outperforming CIP) and handover delays (clearly outperforming HAWAII) during reactive handover.

With gaps in radio coverage, *source-trigger* CAR-set protocol can be used to significantly reduce packet loss. Fig. 14 shows the average packet loss incurred by the three protocols for varying link delays (2, 5, and 10 ms) and gaps (0.25, 0.5, and 1 s) in coverage.

CAR-set is triggered only when the old AR notices that the MN is out of range. Due to the delay to send control messages to the CAR-set ARs, packet loss occurs until the new routes are setup. This is a function of the number of hops from the old AR to the CAR-set ARs and the link delays. This however is significantly less compared to that of the gap itself. With increase in link delay, the packet loss slightly increases. However, the triggered CAR-set protocol may be modified so that the old AR forward the buffered packets to the CAR-set ARs to minimize or eliminate packet loss. In this case, the performance of M&M will be as good as HAWAII in packet loss and significantly better than that of both HAWAII and CIP in terms of handover delay.

The handover delays in coverage gap scenarios consist of the "gap duration $+ \xi$," where ξ is the delay due to the micromobility protocol. We plot ξ for the different protocols in Fig. 15. Average values are given with standard deviation shown in black bars. As can be seen, the delays for M&M are much less than those for CIP and HAWAII. Also, M&Ms handover delays are due to control messages getting from the MN to the AR and are not a function of the link delays. For CIP and HAWAII, the handover delay increases with the wired network delays. This shows a clear advantage in using the CAR-set for scenarios with coverage gaps.

E. CAR-Set Multicast Overhead

Multicasting packets to the CAR-set causes overhead of packet replication over wired links leading to the ARs that belong to the CAR-set. The extent of overhead depends upon the network topology and the size of the CAR-set. The CAR-set protocol does not use any bandwidth in the wireless links of the ARs. Fig. 16 illustrates the overhead incurred for different CAR-set sizes and different gaps in radio coverage.

The total useful bandwidth used is calculated as the product of the number of hops from the BR or RP to the serving AR, times the duration of its usage (i.e., time between handovers). The extra bandwidth is calculated as the product of the number of extra links over which the packets are sent to the CAR-set ARs, times the time for which the CAR-set is used. If the number of useful links (from the BR or RP to the serving AR) is "U," the duration for which it is used is "$\mathrm{t_h}$," the number of extra links is "E" and the time for which it is used is $\mathrm{t_g}$, we calculate the ratio $(\mathrm{E} * \mathrm{t_g})/(\mathrm{U} * \mathrm{t_h})$. For any typical scenario, $\mathrm{t_g}$ will be the gap duration in addition to average round-trip time (RTT) from the new AR to the members of the CAR-set, and $\mathrm{t_h}$ will be the time between successive handovers. Here, we consider an extreme scenario in which the MN handovers once every 5 s (assuming 250 m radius, MN will need to travel at the rate of 100 m/s to handover once every 5 s). This extreme scenario is chosen to illustrate that the overhead incurred even in extreme cases is not high. For more moderate scenarios, the overhead will be significantly less than the numbers indicated in Fig. 16.

The CAR-set size is varied from 2 to 7 and the ratio of extra bandwidth to the useful bandwidth is plotted along with their corresponding standard deviations. For each CAR-set size, the "*ref*" on the x axis indicates the overhead when packets are multicast to the CAR-set all the time (i.e., *no-path trigger*). The other values on the x axis indicate the duration for which the CAR-set is used only as long as there is a gap in radio coverage (i.e., *source-trigger*). From the figure, it is clear that the overhead incurred by *source-trigger* CAR-set is substantially less that incurred by *no-path-trigger* CAR-set (indicated by "ref"). As the number of ARs in the CAR-set increases, the overhead incurred by both the versions of the CAR-set protocol increases. For *no-path-trigger* CAR-set the average ratio of extra bandwidth to useful bandwidth goes up to 4 (with 7 ARs per CAR-set and 1 s gap). This is acceptable if the wired network bandwidth is much larger than that of the wireless network (say 10 times larger). This seems to be valid for most practical purposes.

For *source-trigger* the total overhead incurred for the CAR-set of size 7 and gap of 1 s is less than the total useful bandwidth. This clearly indicates that the *source-trigger* CAR-set protocol gives substantial benefits while incurring a very moderate overhead.

VII. Related Work

Several architectures have been proposed to provide IP mobility support. In MIP [3], every MN is assigned a home address and home agent (HA) in its home subnet. When the MN moves to another foreign subnet, it acquires a care-of-address (COA) through a FA. The MN informs the HA of its COA through a registration process. Packets destined to the MN are sent first to the HA, then are tunneled to the MN. This is known as triangle routing, a major drawback of MIP. Route optimization in [4] attempts to avoid triangle routing by sending binding updates containing the current COA of the MN, to the CN. However, communication overhead during handover renders this scheme unsuitable for micromobility.

In [16], end-to-end IP mobility is proposed, based on dynamic DNS updates. When MN moves, it obtains a new IP-address and updates the DNS mapping for its host name. This incurs handover latency due to DNS update delays and is not suitable for delay-bounded applications. Also, the scheme is not transparent to transport protocols.

In [10], the HA tunnels packets using a prearranged multicast group address. The AR, to which the MN is currently connected, joins the group to get data packets over the multicast tree. This approach suffers from the triangle routing problem; packets are sent to HA first and then to MN. M&M is proposed in [1] and [8]. Each MN is assigned only a unique multicast address. Packets sent to the MN are destined to that multicast address and flow down the multicast distribution tree to the MN. The CN tunnels the packets using the multicast address. This approach avoids triangle routing, in addition to reducing handover latency and packet loss. The study in [1] quantifies the superiority of handover performance for M&M over MIP protocols. These schemes, however, suffer from several serious practical issues, including scalability of multicast state, address allocation and dependency on interdomain multicast. We address these issues in our work.

Several approaches have been proposed for micromobility [18]. The general approaches include mobile-specific routing, hierarchical approaches, and seamless handover. Mobile-specific route approaches include cellular IP [20] and HAWAII [21]. A domain-gateway registers its address with the HA (this has similarities to our proxy-based approach) and forward the packets to the MN. The MNs home address is used within the domain. These approaches need special signaling to update mobile-specific routes and require changes in packet forwarding and unicast routing in all the routers. In CIP [20], signaling is data-triggered to create paths by having routers snoop on the data packets. HAWAII [21] proposes a separate routing protocol and requires explicit signaling from the mobiles. In a way, these approaches attempt to create a distribution tree using extra routing entries for the mobile, similar to multicast. Our approach builds upon existing multicast mechanisms as opposed to recreating them.

Approaches based on seamless handover between old and new ARs, involve fairly complex signaling, buffering and synchronization procedures. Router-assisted smooth handover in MIP [5], and edge mobility [22] belong to this category. Fast handover in [31] introduces fast tunnel set-up between $\mathrm{AR}_{\mathrm{old}}$ and $\mathrm{AR}_{\mathrm{new}}$ as soon as the layer 2 handover is detected. The tunnel avoids packet losses caused by path set-up delay inside the mobility domain. In a way, it is complementary to our multicast-based routing inside the mobility domain. Unlike fast handover, however, our m-subnet idea considers the edge of the network as a single subnet and allows MN to carry RCOA and MCOA across ARs, which reduces the handover latency. Approaches using a hierarchy employ a gateway per-domain and need to keep a location database to map identifiers into locations. This mapping suffers from scalability and robustness problems as was noted earlier in this paper. In [12], a hierarchy of FAs is created at the local, administrative domain and global levels. In [19], a multilevel hierarchy is used in which packets from the HA arrive at a root FA, where they are tunneled to a lower level FA and then to the MN. Hierarchical MIP [27] builds a network of tunnels (overlay network) between FAs. Work in [23] and [29] also use a notion of mobility agent for localized handover within a domain. We have shown in [39] that our multicast-based intradomain mobility scheme outperforms seamless handover and hierarchical approaches and is simpler. This result is consistent with the comparison of routing-based (HAWAII and CIP) and tunneling-based (hierarchical MIP) schemes reported in [40]. It is shown that hierarchical MIP performs either equally well or inferior to the routing-based schemes, because it does not take advantage of the proximity of crossover router to the serving AR. In addition, our comparison results for CIP and HAWAII are generally consistent with the above study. However, we have used more complex topologies, scenarios of reactive handover, and we investigated performance in a more detailed manner; instead of looking at averages, we looked at specific metrics as function of the hop distance from the old AR to the fork router and the new AR to the fork router. Our results indicate that M&M performance during proactive handover matches that of CIP and is better than HAWAII. For reactive handover we show that M&M clearly outperforms CIP and HAWAII.

The idea of CAR-set is similar to handoff-affected router group (HARG) proposed in [41]. The HARG is a group of routers in the network that are affected by the handover when a MN moves from one AP to another and need to do route repair. The fundamental difference with HARG is that the CAR-set is the set of only ARs that are selected to receive the packets destined to the MN.

VIII. Concluding Remarks

We have presented a novel approach to IP micromobility using intradomain M&M. Our approach solves major challenging problems facing the deployment of M&M. Our novel algorithmic mapping scheme from unicast to multicast address ensures collision-free assignment by providing unique and consistent mapping throughout the network.

One major contribution of our study is the introduction of a handover framework, using the CAR-set protocols, that may be tuned to perform efficient proactive, reactive or gap handovers.

In addition, we provided a thorough and systematic evaluation of our proposed schemes. Our extensive simulations for proactive handover show the following.

- There is significant difference in handover delay and packet reordering performance between protocols using different types of handover schemes. For example, M&M and CIP use bicast, while HAWAII uses buffer and forwarding.
- In most proactive handover cases, M&M and CIP show comparable routing efficiency and handover performance because both use shortest path routing as opposed to HAWAII. Routing packets on the path that is not the shortest path from the root of the tree to the MN not only increases end-to-end delay, but also wastes bandwidth and creates extra mobile specific routing entries.
- The bicasting scheme masks proactive handover delays, but produces duplicate packets and small reordering depth depending on the difference in the path lengths from the fork router to the old and new ARs.
- Buffering and forwarding incur longer handover delays and produce large reordering depth.

For reactive and gap handover scenarios M&M clearly outperforms CIP and HAWAII because of our CAR-set path setup capability. Evaluations of variants of the CAR-set protocol show that drastic improvement in handover performance (in terms of delay and packet loss) may be achieved with a slight increase in overhead.

M&M provides a flexible framework through which timing and formation of CAR-set and buffering parameters may be easily tuned for different technologies and scenarios. It is promising in managing mobility in the networks where short dwell time makes the handover latency so pronounced that the network-level handover is hard to achieve. For example, it can be used in future networks designed to offer high bandwidth by employing small cells to high-speed MNs. For such networks, the M&M approach of not requiring address change across the subnets and flexibility in proactive path setup are vital in reducing the handover latency.

M&M uses existing multicast routing, e.g., PIM-SM, with readily available robust implementation. All these factors facilitate the deployment of M&M per Internet service provider (ISP) domain. Furthermore, M&M naturally supports efficient multicasting to MNs.

In the future, we plan to conduct further simulations to evaluate the performance of TCP over M&M. We also would like to investigate M&Ms support for efficient multicast mobile-to-mobile communication.

References

[1] A. Helmy, “A multicast-based protocol for IP mobility support,” in *ACM SIGCOMM 2nd Int. Workshop Networked Group Comm., NGC*, Nov. 2000, pp. 49–58.

[2] D. Estrin, D. Farinacci, A. Helmy, D. Thaler, S. Deering, V. Jacobson, M. Handley, C. Liu, and P. Sharma, “Protocol Independent Multicast—Sparse Mode: Protocol Specification,” IETF, RFC 2362, Mar. ’97/’98.

[3] C. Perkins, “IP Mobility Support,” IETF, RFC 2002, Oct. 1996.

[4] C. Perkins and D. Johnson, “Route optimization in mobile IP,” IETF, Internet-draft, Feb. 2000.

[5] ——, “Mobility support in IPv6,” in *Proc. MOBICOM’96*, 1996, pp. 27–37.

[6] D. Johnson and C. Perkins, “Mobility support in IPv6,” IETF, Internet-draft, July 2003.

[7] A. Myles, D. Johnson, and C. Perkins, “A mobile host protocol supporting route optimization and authentication,” *IEEE J. Select. Areas Commun.*, vol. 13, pp. 839–849, June 1995.

[8] J. Mysore and V. Bharghavan, “A new multicasting-based architecture for Internet host mobility,” *ACM MOBICOM*, pp. 161–172, Sept. 1997.

[9] ——, “Performance of transport protocols over a multicasting-based architecture for Internet host mobility,” in *Proc. ICC’98*, vol. 1, June 1998, pp. 584–589.

[10] S. Seshan, H. Balakrishnan, and R. Katz, “Handoffs in cellular wireless networks: The Daedalus implementation and experience,” *Kluwer Int. J. Wireless Pers. Commun.*, vol. 4, no. 2, pp. 141–162, Mar. 1997.

[11] H. Balakrishnan, S. Seshan, and R. Katz, “Improving reliable transport and handover performance in cellular wireless networks,” *Proc. ACM MOBICOM’95*, pp. 2–11, Nov. 1995.

[12] R. Caceres and V. Padmanabhan, “Fast and scalable handovers for wireless internetworks,” *ACM MOBICOM’96*, pp. 56–66, Nov. 1996.

[13] ——, “Fast and scalable wireless handovers in support of mobile Internet audio,” *ACM J. Mobile Networks Appl.*, vol. 3, no. 4, pp. 351–363, Dec. 1998.

[14] A. Acampora and M. Naghshineh, “An architecture and methodology for mobile-executed handover in cellular ATM,” *IEEE J. Select. Areas Commun.*, vol. 12, pp. 1365–1375, Oct. 1994.

[15] S. Kumar, P. Radoslavov, D. Thaler, C. Alaettinoglu, D. Estrin, and M. Handley, “The MASC/BGMP architecture for inter-domain multicast routing,” in *Proc. ACM SIGCOMM*, Aug. 1998, pp. 93–104.

[16] A. Snoeren and H. Balakrishnan, “An end-to-end approach to host mobility,” *ACM MOBICOM’00*, pp. 155–166, Aug. 2000.

[17] L. Breslau, D. Estrin, K. Fall, S. Floyd, J. Heidemann, A. Helmy, P. Huang, S. McCanne, K. Varadhan, Y. Xu, and H. Yu, “Advances in network simulation,” *IEEE Comput.*, vol. 33, pp. 59–67, May 2000.

[18] A. T. Campbell and J. Gomez, IP micro-mobility protocols, ACM SIGMOBILE mobile computer and communication review, 2001.

[19] E. Gustafsson, A. Jonsson, and C. Perkins, Mobile IP regional registration, Oct. 2002. IETF ID, draft-ietf-mobileip-reg-tunnel-07.

[20] A. Campbell, J. Gomez, S. Kim, A. Valko, C. Wan, and Z. Turanyi, “Design, implementation, and evaluation of cellular IP,” *IEEE Pers. Commun.*, vol. 7, pp. 42–49, Aug. 2000.

[21] R. Ramjee, T. La Porta, L. Salgarelli, S. Thuel, K. Varadhan, and L. Li, “IP-based access network infrastructure for next-generation wireless data networks,” *IEEE Pers. Commun.*, vol. 7, pp. 34–41, Aug. 2000.

[22] A. O’Neill, S. Corson, and G. Tsirtsis, “Routing and handoff in the edge mobility architecture,” *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 4, no. 4, pp. 54–66, Oct. 2000.

[23] K. El-Malki, Ed., *Low Latency Handoffs in Mobile IPv4*, June 2003. IETF ID draft-ietf-mobileip-lowlatency-handoffs-v4-05.txt.

[24] D. Estrin, M. Handley, A. Helmy, P. Huang, and D. Thaler, “A dynamic bootstrap mechanism for rendezvous-based multicast routing,” in *Proc. IEEE INFOCOM’99*, New York, Mar. 1999, pp. 1090–1098.

[25] S. Deering and D. Cheriton, “Multicast routing in data internetworks and extended LANs,” *ACM Trans. Comput. Systems.*, pp. 85–110, May 1990.

[26] E. Fleury, Y. Huang, and P. K. McKinley, "On the performance and feasibility of multicast core selection heuristics," in *Proc. 7th Int. Conf. Computer Communications Networks*, Oct. 1998, pp. 289–293.
[27] H. Soliman, C. Castelluccia, K. Elmalki, and L. Bellier, Hierarchical mobile IPv6, July 2002. IETF ID, draft-ietf-mobileip-hmipv6-06.txt.
[28] A. Misra, S. Das, A. Datta, and S. K. Das, "IDMP-based fast handoffs and paging in IP-based 4G mobile networks," *IEEE Commun. Mag.*, vol. 4, no. 3, pp. 138–145, Mar. 2002.
[29] G. Dommetry, Local and Indirect Registration for Anchoring Handoff, Apr. 2002. IETF ID draft-dommetry-mobileip-anchor-handoff-02.txt.
[30] K. Calvert, M. Doar, and E. Zegura, "Modeling Internet topology," *IEEE Commun.*, pp. 160–163, June 1997.
[31] Fast Handovers for Mobile IPv6, R. Koodli, Ed., Sept. 2003. draft-ietf-mobileip-fast-mipv6-07.txt, work in progress.
[32] R. Hinden and S. Deering, "IP Version 6 Addressing Architecture," IETF, RFC 3513, Apr. 2003.
[33] R. Hinden, M. O'Dell, and S. Deering, "An IPv6 aggregatable global unicast address format," IETF, RFC 2374, July 1998.
[34] R. Droms, Ed., "Dynamic host configuration protocol for IPv6 (DHCPv6)," IETF, RFC 3315, July 2003.
[35] S. Thomson and T. Narten, "IPv6 stateless address autoconfiguration,", RFC 2462, Dec. 1998.
[36] *3GPP TS Group Core Network; Numbering, Addressing, and Identification*, June 2001.
[37] T. Narten, E. Normark, and W. Simpson, "Neighbor discovery for IP, Version 6 (IPv6)," IETF, RFC 2461, Dec. 1998.
[38] A. Helmy, "State analysis and aggregation study for multicast-based micromobility," in *Proc. IEEE Int. Conf. Communications*, May 2002, pp. 3301–3306.
[39] A. Helmy and M. Jaseemuddin, "Efficient micro-mobility using intra-domain multicast-based mechanisms (M&M)," USC-Tech. Rep., Aug. 2001.
[40] A. Campbell, J. Gomez, S. Kim, Z. Turanyi, C.-Y. Wan, and A. Valko, "Comparison of IP micromobility protocols," *IEEE Wireless Commun.*, vol. 9, pp. 72–82, Feb. 2002.
[41] H. Li, "A micro-mobility routing protocol for wireless Internet," *IP-Based Cellular Networks (IPCN) 2001*, May 17, 2001.
[42] M. S. Corson and A. O'Neill, "An approach to fixed/mobile converged routing," Univ. Maryland, TR-2000-5, 2000.
[43] D. Meyer, "Administratively Scoped IP Multicast,", RFC 2365, July 1998.
[44] R. Braden and J. Postel, "Requirements for Internet gateways," IETF, RFC 1009, June 1987.
[45] A. Iera, A. Molinaro, and S. Marano, "Handoff management with mobility estimation in hierarchical systems," *IEEE Trans. Veh. Technol.*, vol. 51, pp. 915–934, Sept. 2002.

Ahmed A.-G. Helmy received the B.S. degree in electronics and communications engineering from Cairo University, Cairo, Egypt, in 1992, the M.S. degree in electrical engineering, and the Ph.D. degree in computer science from the University of Southern California (USC), Los Angeles, in 1995 and 1999, respectively.

Since 1999, he has been an Assistant Professor of electrical engineering with the University of Southern California. In 2000, he founded and is currently directing the Wireless Networking Laboratory, USC. His current research interests lie in the areas of protocol design and analysis for mobile Ad Hoc and sensor networks, mobility modeling, design and testing of multicast protocols, IP micromobility, and network simulation.

Dr. Helmy received the National Science Foundation (NSF) CAREER Award in 2002, the USC Zumberge Research Award in 2000, and the Best Paper Award from the IEEE/IFIP International Conference on Management of Multimedia Networks and Services (MMNS) in 2002.

Muhammad Jaseemuddin (M'98) received B.E. degree from N.E.D. University of Engineering and Technology, Karachi, Pakistan, in 1989, the M.S. degree from the University of Texas, Arlington, in 1991, and the Ph.D. degree from the University of Toronto, Toronto, ON, Canada, in 1997.

He was with the Advanced IP Group and Wireless Technology Laboratory (WTL), Nortel Networks. He worked on the prototypes of wireless service delivery platform and UMTS VHE framework. He contributed to the development and implementation of Open IP suite of IP protocols. While with WTL, he worked on QoS, routing and handover issues in mobile wireless IP access network. He has been Associate Professor at Ryerson University, Toronto, since 2002. His research interests include IP mobility, IP traffic engineering, routing protocols, overlay Ad Hoc network, transport issues in ad hoc network, and network autonomics.

Ganesha Bhaskara (M'03) was born in Bangalore, India. He received the B.E. degree in electronics and communication from Bangalore University, Bangalore, India, in 1999, the M.S. degree in computer networks from the University of Southern California (USC), Los Angeles, in 2002. He is currently working toward the Ph.D. degree in electrical engineering at the USC under the guidance of Dr. Helmy and Dr. Gupta.

He is a member of the Network Design and Testing Laboratory, USC. His research interests include protocol design methodologies, formal methods for protocol design and analysis, protocol testing, and verification techniques.