Mobile Networks and Applications 7, 127–141, 2002

P-MIP: Paging Extensions for Mobile IP

XIAOWEI ZHANG, JAVIER GOMEZ CASTELLANOS and ANDREW T. CAMPBELL
Comet Group, Columbia University, New York, NY 10027, USA

Abstract. As the number of Mobile IP users grows, so will the signalling overhead associated with Internet mobility management in the core IP network. This presents a significant challenge to Mobile IP as the number of mobile devices scale-up. In cellular networks, registration and paging techniques are used to minimize the signalling overhead and optimize the mobility management performance. Currently, Mobile IP supports registration but not paging. In this paper, we argue that Mobile IP should be extended to support paging to improve the scalability of the protocol to handle large populations of mobile devices. To address this, we introduce P-MIP, a set of simple paging extensions for Mobile IP, and discuss the construction of paging areas, movement detection, registration, paging and data handling. We present analysis and simulation results for Mobile IP with and without paging extensions, and show that P-MIP can scale well supporting large numbers of mobile devices with reduced signalling under a wide variety of system conditions.

Keywords: Mobile IP, paging, mobility management

1. Introduction

Mobile IP (MIP) [11] represents a global mobility solution that provides host mobility support for a wide variety of access technologies, devices, and applications. Mobile IP supports a simple mobility mechanism. A mobile node's location is tracked by its home agent which binds the care-of address (CoA) or co-located care-of address used by the mobile node in a visited network to the mobile node's home address. When a mobile node moves to a new network, it "registers" its new care-of address with its home agent. This basic mechanism presents a number of challenges to the widespread deployment of Mobile IP with respect to handoff performance and scalability issues. In this paper, we address Mobile IP scalability in support of very large numbers of mobile nodes. For a survey of solutions that improve Mobile IP handoff performance, see [7].

In cellular networks, registration and paging techniques are used to minimize the signalling overhead and optimize mobility management performance. As a result, these mechanisms compensate each other to improve the overall system performance. Currently, Mobile IP supports registration but not paging. We observe that, as in the case of cellular network subscribers, Mobile IP users will not be actively communicating most of the time [3], rather, they will be idle. We argue in this paper that as wireless access to Internet becomes the norm, Mobile IP will need to provide a more scalable and efficient location tracking scheme that distinguishes between mobile nodes that are actively communicating or idle. In this case, it would be sufficient for Mobile IP only to know the approximate location of idle users to resolve this scalability issue. The exact location of idle mobile nodes can be found by using *paging*.

Paging is a procedure that allows a wireless system to search for an idle mobile host when there is a message destined for it, such that the mobile user does not need to register its precise location to the system whenever it moves. Paging has a number of benefits. First, it reduces signalling overhead associated with registration and location system database updates. The power consumption of mobile nodes is also reduced because idle mobile nodes no longer have to register their exact location with the system. Paging, however, introduces uncertainty about the location of mobile hosts, and as such, can result in additional delays associated with message delivery. Paging also introduces additional signalling to the system that consumes system resources. Thus, it is important to balance the paging and registration processes in order to optimize system performance.

The contribution of this paper is as follows. We address Mobile IP scalability and propose P-MIP, paging extensions for Mobile IP. We show that the addition of a set of simple paging extensions can provide significant savings in signalling overhead and bandwidth, helping Mobile IP to scale to support large numbers of mobile users. This paper is structured as follows. Section 2 discusses related work and section 3 provides an overview of P-MIP and its operations. Section 4 presents the detailed design of the protocol, which assumes the use of foreign agents. We discuss P-MIP's support for paging area construction, movement detection, registration, paging and data handling. In section 5, we evaluate the signalling cost of Mobile IP with and without the P-MIP paging extensions. We study how paging area size, mobile speed, mobile density and data session characteristics can impact the signalling cost under a variety of system conditions. We use a combination of analysis and simulation to evaluate the system performance and present our results. Finally, in section 6 we make some concluding remarks.

2. Related work

Paging is used widely in cellular systems [6,9] to locate idle mobile stations prior to establishing incoming calls. A paging area is constructed by a group of base stations and, typically, these base stations are under the control of the same mobile switching center (MSC). When an incoming call for a mobile user is received, the mobile switching center sends a paging message to all the base stations in the same paging area. Each base station broadcasts the paging message in its own cell. The system determines the mobile station's accurate location after receiving a paging response message from the paged mobile device. The precise location information is then used to establish the call.

Cellular networks are connection-oriented and paging is only necessary in support of incoming calls. Mobile IP, however, is based on a connectionless service paradigm where datagrams are treated independent from one another. The interpretation of paging in IP networks is, therefore, an open issue. Clearly, paging would represent a considerable overhead if mobile hosts were paged on each incoming packet. This issue is resolved in P-MIP by using mobile host state and the concept of a data session. Mobile node state consists of a per-node tuple (operational mode, active state timer) where mobiles can be in an *active* or *idle* operational mode. A mobile node is in active mode if it has recently sent or received IP data. A mobile node is considered to be active for an active timer period starting from the instance the node sent or received data. Each time the mobile node sends or receives data, the timer is reset in both the mobile node and the serving foreign agent. When the active timer expires, the mobile node enters idle state. Whenever an idle node sends or receives data, its operational mode is changed to active and its active timer started. The active state timer value is implementation dependent. If packets in a packet train are close enough, then a mobile node's state will remain active during the packet train transmission. In P-MIP, we call this a *data session*. We can broadly consider the "holding time" of data sessions as analogous to call holding times found in connection-oriented networks. However, in some cases (e.g., Web interactions), these are very short-lived sessions (e.g., microflows). As a general design rule, P-MIP only needs to page a mobile node on the first packet of a data session.

Recently, a number of micro-mobility protocols [7] (e.g., Cellular IP [3], Hawaii [12] and Regional Registration [8]) have been primarily developed to support better handoff performance in environments where mobile hosts handoff frequently. Micro-mobility protocols reduce registration signalling and improve performance by minimizing delay and packet loss during handoff. However, these fast handoff approaches do not improve the power consumed by mobile nodes or the bandwidth consumed over the air interface. In contrast, paging does not reduce the registration delay. Rather, paging reduces the number of registrations that mobile nodes make. Paging can reduce the power consumption of mobile nodes, and possibly the bandwidth used over the air interface. In this paper, we focus on paging and not fast handoff control. By extending Mobile IP to support paging, P-MIP can reduce the signalling load in the core IP network, power consumption of mobile nodes, and potentially radio resources.

Two micro-mobility protocols, namely, Cellular IP [3] and Hawaii [12], integrate support for fast handoff and paging. These protocols, which operate in wireless access networks and interwork with Mobile IP via mobile gateways, represent independent protocols that are rather different in comparison to the base Mobile IP protocol [11]. We take a different approach and propose P-MIP, a simple set of paging extensions to the base Mobile IPv4 protocol.

In [4], Mobile IP is extended with an adaptive paging scheme. In this approach, mobile nodes support optimum paging areas, where a paging area is adaptive on a per-mobile basis. Mobile nodes continuously compute optimal paging areas. This operation can, however, adversely impact the power consumption of mobile nodes. In addition, some of the input parameters used by the optimization process are difficult for a mobile to determine. The adaptive paging scheme also calls for protocol modifications that are incompatible with Mobile IP. P-MIP defines a minimal set of paging extensions to the base Mobile IP protocol and is backward compatible with Mobile IP. We believe P-MIP represents a more synergistic approach to the base protocol for the delivery of value-added IP paging services.

3. P-MIP overview

In what follows, we present an overview of P-MIP, which extends the Mobile IP [11] agent advertisement message to include a "P bit", as shown as in appendix A.4. P-MIP assumes the use of foreign agents and that mobile nodes register through foreign agents[1]. Foreign agents supporting paging set the P bit in agent advertisements. As a default, however, the P bit is not set. When a mobile node first receives an agent advertisement, it checks if paging is supported by looking at the "P bit". In this paper, we assume that foreign agents support paging. Note that there is also a P bit in the registration request message, which is used to indicate if mobile nodes support paging or not. P-MIP message formats are shown in the appendix.

An active mobile node operates in exactly the same manner as in Mobile IP [11]. When a mobile node changes its point of attachment, it registers. When an idle mobile node moves to a new paging area, it registers. Idle mobile nodes do not register when moving within a paging area.

When packets are destined to mobile nodes, home agents forward them to the foreign agents registered by the mobile nodes, known as registered foreign agents. A registered foreign agent first determines if it has any information on record[2] for the mobile node. If a record exists, then the registered foreign agent checks if the mobile node supports paging or

[1] Future work will consider paging in environments that do not support foreign agents (e.g., when collocated care-of addresses are used).

[2] The term "record" refers to Mobile IP mobility management related information maintained by home and foreign agents, and mobile nodes. A record comprises transient state information including the mobile node's home address, home agent address, care-of address, and various timer related information associated with control messaging (e.g., registration lifetime, etc.).

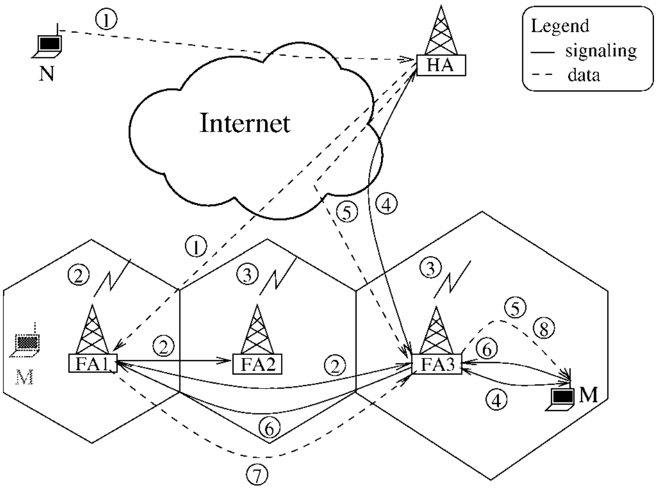

Figure 1. An example P-MIP paging scenario.

not. If the mobile node supports paging, then the registered foreign agent checks the mobile node's operational state. If the mobile node is in active mode, then the registered foreign agent decapsulates packets and forwards them to the mobile node, as in the case of Mobile IP [11]. If the mobile node is in idle mode, then the registered foreign agent sends a *paging request* message to all other foreign agents that reside in the same paging area, as well as broadcasting the paging request on its own network.

We use the following terminology to clarify the interaction between system components that implement P-MIP. The foreign agent serving the network that a mobile host is currently visiting is called the *current foreign agent*; the foreign agent a mobile host registers with is called the *registered foreign agent* which may or may not be on the same network that the mobile is currently located; the foreign agent that originates the paging request is called the *paging foreign agent*; and a mobile node that is paged is called the *paged mobile node*.

When a mobile node receives a paging request message, it registers with its home agent through the foreign agent on the visited network (i.e., the current foreign agent). After receiving a registration reply message, the mobile node sends a paging reply message back to the foreign agent it had previously registered with through its current foreign agent to inform the previously registered foreign agent of its current location. Note that a mobile node's current foreign agent and its previously registered foreign agent are the same foreign agent if it has not moved out of the network since the last registration. When the previously registered foreign agent receives a paging reply message, it forwards any buffered packets toward the mobile node through its current foreign agent.

Figure 1 illustrates a simple paging scenario where foreign agents FA1, FA2, and FA3 form a paging area. An idle mobile node (M) moves from the FA1 cell to the FA3 cell without registering. The care-of address record maintained at the mobile host's home agent points to the foreign agent FA1. In this scenario, a correspondent node (N) sends data ① to mobile node M. The home agent (HA) encapsulates the data packets ① and tunnels packets to the foreign agent FA1. After foreign agent FA1 receives data destined to mobile node M, it checks if it has a record for the mobile node. If it has a record, then it determines if the mobile node supports paging. If this is the case, then FA1 checks the mobile node state. In this scenario, the mobile node (M) is in an idle state. As a result, the paging foreign agent FA1 starts to buffer packets and sends a paging request message ②. This paging request is broadcast in foreign agent's (FA1) cell and unicast to foreign agents FA2 and FA3. Following this, foreign agents FA2 and FA3 broadcast the paging request message ③ in their cells. The mobile node receives the paging request (which includes its home address) in the FA3 cell and registers its location with its home agent, as shown by ④ in the figure. The home agent starts forwarding data toward foreign agent FA3 and the mobile node (M) once the registration process is complete, as shown by ⑤ in the figure. In addition to action ④, the mobile node sends a paging reply ⑥ to foreign agent FA3. Foreign agent FA3 forwards the paging reply message to the paging foreign agent FA1. Foreign agent FA1 forwards[3] any buffered data toward foreign agent FA3 (as indicated by ⑦ in the figure) and deletes the mobile node's record. Following this, foreign agent FA3 forwards data packets ⑧ to the mobile node.

4. P-MIP design

The design of P-MIP encompasses paging area construction, movement detection, registration, paging and data handling.

4.1. Paging area construction

In P-MIP, a paging area is identified as having a unique paging area ID (PAI) and consists of two or more networks which are identified by the network prefix as part of their Internet addresses. Paging areas can be configured based on a number of criteria (e.g., node mobility, traffic patterns, mobile density, etc.) so that most mobile nodes move within the same paging area. Paging areas can be configured manually by administrators by setting parameters at each foreign agent or more automatically by having foreign agents interact with paging servers. The benefit of paging servers is that administrators only need to configure the servers. In this case, each foreign agent acquires its paging area information directly from a paging server.

We consider two types of paging area construction in P-MIP that encompass *non-overlapping* and *overlapping paging areas*[4]. In the case of non-overlapping paging areas, a network can only be associated with one paging area, as shown for paging areas PA1, PA2, and PA3 in figure 2. Note that the

[3] Out of sequence data may be experienced by mobile nodes during paging due to buffering delays at paging foreign agents. Such conditions may adversely impact the performance of UDP and TCP applications. This topic is for further study.

[4] Our approach to overlapping paging area construction is similar to the work on adaptive paging areas [4]. In P-MIP, however, each paging area is based on a foreign agent rather than a mobile node. In addition, P-MIP paging areas are preconfigured and fixed which simplifies the protocol, rather than being adaptive [4].

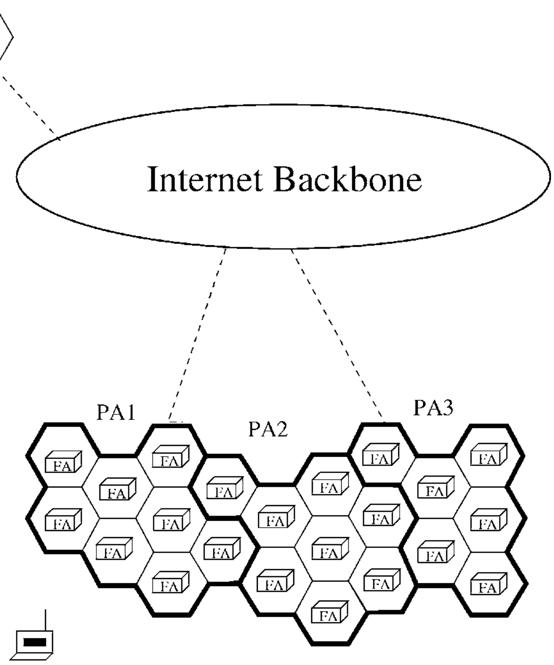

Figure 2. Non-overlapping paging areas.

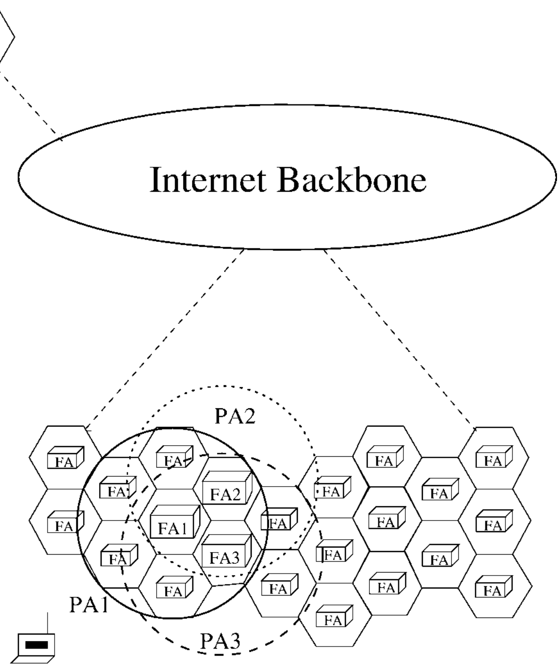

Figure 3. Overlapping paging areas.

Table 1
Paging table for the non-overlapping paging scheme.

PAI	XXXX XXXX
FA1	FA1's IP address
FA2	FA2's IP address
FA3	FA3's IP address
⋮	⋮

Table 2
Paging table for the overlapping paging scheme.

FA1	FA1's IP address
FA2	FA2's IP address
FA3	FA3's IP address
⋮	⋮

figure does not show any specific connections within or between paging areas. A number of possible connectivity scenarios can be considered. For example, all base stations in a paging area are directly connected to the Internet, or all access networks in a paging area are connected to the Internet through one or more IP gateways.

In the non-overlapping scheme, each paging area has a unique paging area ID. Paging area IDs can be reused in the same sense as spatial reuse with adjacent paging areas having different IDs. Paging area ID assignment could follow the same lines as IP addresses, with authorized organizations managing allocation and assignment.

A paging table is maintained by each foreign agent and can be manually set or configured by a paging server or an administrator. In P-MIP, the agent advertisement message is extended to carry the paging area ID, as shown in appendix A.1. The agent advertisement message must be broadcast periodically so that the paging area ID can be periodically distributed. The format of the paging tables for non-overlapping scheme is shown in table 1.

In the case of overlapping paging areas, the paging area ID is made up of the entire list of foreign agent IP addresses in the paging area. The paging area ID does not need to be periodically broadcast as is the case with non-overlapping paging IDs. Rather, the paging area ID (shown in appendix A.3) is sent to mobile nodes using an extended registration reply message. Each foreign agent maintains a paging table in the case of the overlapping scheme. The paging tables for the overlapping scheme can be manually configured, as shown in table 2. Note that the difference between the two tables is that there is no specific paging area ID value associated with the overlapping scheme.

The maximum benefit of the overlapping scheme is achieved when each foreign agent is located at the center of its paging area. As shown in figure 3, foreign agent FA1 is located at the center of paging area PA1, FA2 is located at the center of PA2, and so on. By the term "center" we mean the optimum position where a foreign agent is positioned in its paging area such that a mobile node in this network minimizes its chance of moving out of the current paging area. In this manner, the signalling generated by boundary crossings can be reduced to a minimum. Whenever a mobile node registers, it will receive a new paging area ID (i.e., the list of foreign agent IP addresses in the paging area). This allows the mobile node to position itself at the center of the new paging area when it registers. In this case, registration initiated by crossing new paging area borders is reduced, allowing for an even distribution of registration traffic. This approach would further reduce the signalling associated with mobile nodes that frequently move around paging area borders, which may occur in the case of non-overlapping paging areas.

Each paging scheme has a number of advantages and disadvantages. In the case of non-overlapping paging areas, a paging ID is much shorter than that in overlapping paging areas. Therefore, the bandwidth and power consumed to manage paging IDs will be more efficient in the case of non-overlapping areas. However, this scheme cannot take advantage of overlapping paging areas. As a result, there will be a disproportionate amount of registration traffic generated at paging area borders. With overlapping paging areas, the paging area ID could be rather large when there is a large number of foreign agents in a paging area. In this case, each mobile node must be aware of all the foreign agent IP addresses in a paging area. Mobile nodes need to be capable of maintaining this state information in order to detect if they have moved to a new paging area. Overlapping paging area information is not transmitted periodically so that the impact on radio resources and mobile node resources is limited.

Based on the paging area scheme selected, the P-MIP protocol mechanisms operate differently (e.g., in the case of movement detection). In this paper, we highlight these operational differences by referring to the two paging schemes discussed above.

4.2. Movement detection

The algorithms used to detect whether a mobile node has moved to a different network are the same for P-MIP as Mobile IP. Two algorithms are used for movement detection in Mobile IP. In the first approach, mobile nodes detect movement using agent advertisement messages and the advertisement lifetime. When a mobile node receives an agent advertisement, it extracts the advertisement lifetime included in the message. If a mobile node does not receive another advertisement before the advertisement lifetime expires, then the mobile node assumes that it has moved to a new network. In the second approach, a mobile node compares the network prefix obtained from newly received agent advertisement messages to the current care-of address. If they are different, the mobile node assumes it has moved to another network.

If a mobile node is in an active mode when it detects it has moved, then handoff is initiated which invokes the registration process. In this case, the mobile node's location will be updated at both the foreign and home agents. In contrast, if a mobile node is in an idle state and it detects it has moved to a new network within the same paging area, then registration is unnecessary. However, if an idle mobile node does not have a valid foreign agent on record, then it should locate one for further movement detection and registration purposes.

The algorithms used to detect movement between paging areas are different based on the paging area scheme adopted. In the case of non-overlapping paging areas, a mobile node records the advertisement lifetime from the agent advertisement message. If a mobile node fails to receive another advertisement from the same agent within the specified lifetime, then the mobile node assumes that it has lost contact with its registered foreign agent. If the mobile node has previously received an agent advertisement from another foreign agent for which the lifetime field has not yet expired, then the mobile node should check if the paging area ID from the new agent is the same as its current paging ID. If the paging IDs are the same, then the mobile node will do nothing except track the foreign agents identified using agent advertisement messages. If the paging IDs are different, then idle mobile nodes should immediately register with the new foreign agent. If the mobile node is not in contact with any foreign agent, then it sends an agent solicitation message. Upon reception of an agent advertisement message, the mobile node checks if the new paging area ID is the same as the old ID. The mobile node uses this test to determine whether it has moved to a new paging area or not.

In the case of overlapping paging areas, a mobile node records the advertisement lifetime received from the agent advertisement message. If a mobile node fails to receive another advertisement from the same agent within the specified lifetime, then the mobile node assumes that it has lost contact with the registered foreign agent. If the mobile node has previously received an agent advertisement from another foreign agent for which the lifetime field has not yet expired, then the mobile node checks if the source IP address of the agent advertisement sent by the new foreign agent is in the current paging ID's list of foreign agents. The current foreign agent list is obtained in the registration reply message. If the new foreign agent's IP address is not in the current foreign agent list, then the idle mobile node registers with the new foreign agent. If the mobile node is not in contact with a foreign agent, then it should send an agent solicitation message. Upon reception of an advertisement message, the mobile node should check if the new foreign agent's IP address is in the current foreign agent list or not. It uses this test to determine whether it has moved to a new paging area or not.

4.3. Registration

Registration in P-MIP follows the same procedure as Mobile IP. However, after receiving a registration reply message in P-MIP, both the foreign agent and the mobile node update the mobile node state by setting the operational mode to active and starting the active timer. In the case overlapping paging areas, the registration reply message includes the paging area ID. The list of foreign agents in the paging area is inserted by the foreign agent in the registration reply message returned to the mobile node. The mobile node records the paging area ID.

An important performance difference between Mobile IP and P-MIP relates to the number of registration signalling messages generated. In Mobile IP, a mobile node registers when:

- a mobile node changes its point of attachment;
- a mobile node's registration lifetime is about to expire (we call this "registration refresh"); or
- a mobile node detects that its current foreign agent has rebooted.

In P-MIP, a mobile node registers when:

- a mobile node detects it has moved to a new paging area;
- a mobile node's registration lifetime is about to expire;
- a mobile nodes detects that its current foreign agent has rebooted;
- a mobile node is paged and is not on the same network as the paging foreign agent; or
- an idle mobile node is about to transmit data.

Because idle mobile nodes do not register when moving between networks in the same paging area, P-MIP reduces signalling in comparison to the base Mobile IP protocol. We discuss this issue when we present our performance evaluation in section 5.

4.4. Paging

When a foreign agent has to locate an idle mobile node, it broadcasts a paging request message on its own network and to all other foreign agents in its paging area. We use unicast[5] as a basis for broadcasting paging messages to all other foreign agents in the paging area. Foreign agents broadcast the paging request within their own wireless networks. P-MIP supports the optimization that a foreign agent may have a number of paging requests to make at any one moment. In this case, P-MIP aggregates all paging requests into a single paging message. This helps to minimize the paging overhead as the number of paged mobile nodes grows. The paging request message lists the home addresses of all paged mobile nodes, as shown in appendix A.2. In this paper, we focus on the scenario where a foreign agent pages a single mobile node.

When a foreign agent receives a paging request from the paging foreign agent, it first conducts an authentication check if the foreign agents share a mobility security association[6]. If authentication is successful, then the foreign agent broadcasts the paging message on its own network. When a mobile node finds its home address in the paging request and the message is directly from its registered foreign agent, then the registered foreign agent is the same one as the mobile host's current foreign agent. Thus, the mobile node deduces that it is still located in the same network as its registered foreign agent. In this case, the mobile node sends a paging reply back to the paging foreign agent without registering, and sets its operational mode to active and restarts its active timer. Otherwise, the mobile node starts the registration procedure to make sure that the binding between its home address and new care-of address is valid. Following this, the mobile node responds to the paging request by sending a paging reply. The mobile node will insert its care-of address in the paging reply. The transmission time for each registration and paging reply to a foreign agent is randomized to avoid potential collisions between signalling messages (e.g., registrations and paging reply messages) from multiple paged mobile nodes in the same network.

When the foreign agent receives the registration reply from a home agent, it sets the mobile node's operational mode to active and then relays the registration reply to the mobile node. If the mobile node does not receive the registration reply, its operational mode remains idle in its own record. Loss of the registration reply message can result in state inconsistency between the foreign agent and mobile node records. However, this state inconsistency is short lived. When the foreign agent receives the registration reply message, it starts to forward data to the mobile node. If a mobile node receives data packets in idle mode, it checks if this foreign agent is the one to which it sent a registration request. If this is the case, then the mobile node sets its care-of address to the one associated with the foreign agent, changes operational mode to active, and then starts its active timer.

A foreign agent conducts authentication checks on paging reply messages received from paged mobile nodes if the foreign agent and the mobile nodes share a mobility security association. If the foreign agent did not originate paging, then it will forward the paging reply message to the paging foreign agent. The foreign agent takes this action after appending the foreign agent-to-foreign agent authentication extension to the paging reply message. If the paging foreign agent receives a paging reply message directly from a paged mobile node, then it sets the mobile node's operational state to active and starts the active timer. The paging foreign agent removes "records" for all paged mobile nodes that are not located in its network. The paging foreign agent can determine this from paging reply messages. Paging foreign agents take this action because paged mobile nodes located in other paging areas have already initiated registration through foreign agents in the new network.

The paging foreign agent will retransmit a paging request after a "paging period" if no response is received from an earlier paging request. If a mobile node's registration lifetime expires during paging, then the paging foreign agent stops paging. Each successive paging period is at least twice the previous duration. The foreign agent stops paging after sending multiple successive paging requests without receiving a paging reply. The specific number of paging request retransmissions is implementation dependent. However, the paging foreign agent maintains the paged mobile node's record until the lifetime expires. Maintaining the record in this manner counters the situation where mobile nodes are temporarily partitioned in the network.

[5] IP Multicast would provide better performance than simple unicast distribution of paging request messages in paging areas. However, IP Multicast is rather complex to implement in wireless access networks. Therefore, we adopt unicast for simplicity at this stage of our work. Future work will consider more efficient support for the transport of signalling messages in paging area through the implementation of a lightweight multicast service tailored for such a task.

[6] The authentication method used in P-MIP is the same as in Mobile IP [11]. If mobility security associations exist among foreign agents in a paging area, then the foreign agent-to-foreign agent authentication extension is attached at the end of the paging request and reply messages. If a mobility security association exists between a foreign agent and a mobile node, then the mobile node-to-foreign agent authentication extension is attached at the end of the paging request and reply messages. The case when nodes do not share a security association is for further study.

In the case where a mobile node moves to a new network, loss of paging reply messages does not have any serious impact on the mobile node's operation. This is because a mobile node will register with a new foreign agent before sending out a paging reply message. Therefore, the record maintained by a home agent points to the current foreign agent. However, it may cause unnecessary data loss for a mobile node which is in the same network. In this case, the mobile node sends out a paging reply and sets the mode to active. If a paging reply message is lost, the mobile node's operational mode remains idle in the record maintained by the foreign agent. In this case, the foreign agent continues paging the mobile node. The paging foreign agent will stop paging after multiple attempts and discard any buffered data. The following action attempts to resolve this condition. After a mobile node (which is in the same network) sends a paging reply message, it assumes reception of data from the foreign agent. If this does not happen after a certain period, then the mobile node assumes that the paging reply has been lost and sends a paging reply again. If the mobile node still does not receive data after multiple paging reply retransmission attempts, it resets its operational mode to idle. The paging foreign agent will forward any buffered data toward the mobile node once a paging reply has been received by the paging foreign agent.

4.5. Data handling

When there are incoming packets destined to a mobile node, the mobile node's home agent forwards data to its care-of address, associated with the mobile node's registered foreign agent. When the registered foreign agent receives data, it first determines whether it has a record for the mobile node. If a record exists, then the registered foreign agent determines if the mobile node supports paging by checking the P bit in the registration request message. If the P bit is set, then the registered foreign agent checks the mobile node's state. If the mobile node is in an active state, then it is likely that the mobile node is located in the registered foreign agent's network. In this case, the foreign agent decapsulates and forwards packets to the mobile node. As discussed earlier, the registered foreign agent sends a paging request message if the mobile node is in an idle state, and at the same time, buffers the data for the mobile node. When the paging foreign agent (which is also the registered foreign agent) receives a paging reply, it forwards the buffered data to the mobile node. If the mobile node is not in the registered foreign agent's network, then the registered foreign agent removes the mobile node's record.

A foreign agent restarts a mobile node's active timer each time a data packet is sent or received to or from the mobile node, respectively. The mobile host also restarts its active timer under these conditions. In the case that a mobile node does not respond to multiple page request messages, the registered foreign agent discards any buffered data.

A mobile node can simply send data when it is in active mode. If the mobile node is idle, it first determines if it is still located on the same network that it previously registered with by checking the agent advertisement or by sending an agent solicitation message. If the mobile node has moved to a new network, then it initiates a registration procedure to update its location record in its home agent and foreign agent before sending data. It is likely that a mobile node which is sending data (e.g., TCP segments) will also receive data (e.g., TCP acknowledgements) in return. Therefore, the home agent and foreign agent bindings need to be updated to facilitate the delivery of incoming data to the mobile node. If the mobile node resides on the registered foreign agent network, then it does not need to register. Rather, the mobile node only needs to set its operational mode to active and starts its active timer before transmitting data.

5. Performance evaluation

We use a combination of analysis and simulation to evaluate the signalling cost of Mobile IP with and without the P-MIP paging extensions. We study how paging area size, mobile speed, mobile density, and session characteristics can impact the signalling cost under a variety of system conditions. The evaluation only considers the non-overlapping paging approach. The signalling cost is represented as the product of the weighted distance (see section 5.2.1 for details) the signalling message travels and the signalling rate. The unit of cost is weighted hops·pkt/s. For simplicity, the air interface is treated as a single hop in the wireline network. Because agent advertisement and solicitation signalling present the same costs in Mobile IP and P-MIP, we exclude them from the evaluation. Therefore, only registration and paging signalling are considered. Simulation is also used to investigate scenarios that cannot be considered by analysis, for example, using a range of mobile speeds other than a constant speed, which is considered in the analysis. P-MIP is implemented as extensions to the *ns* simulator [5] Mobile IP [17] code. Simulation is also used to validate the analysis. The P-MIP *ns* source code is available on the Web [18]

5.1. Mobility models

Paging aims to reduce the signalling overhead associated with mobility in wireless networks. The signalling analysis is, therefore, closely related to the movement patterns of mobile nodes. In legacy wireless networks where voice is the dominant service, the fluid flow model is widely used to analyze cell boundary crossing related issues, such as handoff [2,10]. For simplicity, we adopt this model to analyze the signalling cost in a Mobile IP environment. We also assume that paging areas and wireless cells are square-shaped where there are n cells in a paging area. The perimeter of a cell is l, so the perimeter of the paging area, denoted as L, is $L = l\sqrt{n}$. Mobile nodes move at an average velocity of v in directions that are uniformly distributed over $[0, 2\pi]$ and are uniformly distributed with density ρ. The cell boundary crossing rate r_c is

$$r_c = \frac{\rho v l}{\pi}, \tag{1}$$

where r_c is the cell crossing rate (mobiles/s); ρ is the mobile density (mobiles/m^2); v is the moving velocity (m/s); and l is the cell perimeter (m).

Mobile devices move across a boundary in two directions. For evaluation purposes, however, only one direction needs to be considered. The paging area boundary crossing rate r_p is

$$r_p = \frac{\rho v L}{\pi}. \quad (2)$$

One drawback of the fluid flow mobility model used in the analysis is that mobile nodes are assumed to move at constant fixed rates. The mobility model used in the simulation supports a wide variety of mobility behavior. The mobility model is based on the random waypoint algorithm [1]. A mobile node picks a random location in the simulation area as the destination and moves towards the location at a speed chosen uniformly between 0 and the maximum speed. When a mobile node reaches a destination point, it stops for the duration of pause time and then picks another destination and speed, and moves again. This cycle is repeated for the duration of the simulation. When the pause time is set to the simulation duration time, the mobile node remains stationary. As the pause time decreases so does the movement of the mobile node. Continuous movement corresponds to a pause time of 0. Node mobility movement patterns are generated as in [5].

5.2. Signalling cost analysis

5.2.1. Mobile IP

The formula used to calculate the signalling cost for Mobile IP is

$$\begin{aligned} C &= d_{\mathrm{HA,FA}}(R_{\mathrm{core}} w_{\mathrm{core}} + R_{\mathrm{local}} w_{\mathrm{local}}) \left[r_c n + \rho \left(\frac{l}{4}\right)^2 n r_r \right] \\ &= d_{\mathrm{HA,FA}}(R_{\mathrm{core}} w_{\mathrm{core}} + R_{\mathrm{local}} w_{\mathrm{local}}) \\ &\quad \times \left[\frac{\rho v l}{\pi} n + \rho \left(\frac{l}{4}\right)^2 n r_r \right], \end{aligned} \quad (3)$$

where:

- C is the signalling cost for Mobile IP (weighted hops·msg/s);
- $d_{\mathrm{HA,FA}}$ is the average distance between home and foreign agent in terms of the number of hops;
- R_{core} is the ratio of the number of hops in the core network to the total number of hops between a home agent and a foreign agent;
- R_{local} is the ratio of the number of hops in local access networks to the total number of hops between a home agent and foreign agent;
- w_{core} is the weight of each hop in the IP core network;
- w_{local} is the weight of each hop in access networks;
- r_c is the cell crossing rate (mobiles/s);
- ρ is the mobile density (mobiles/m^2);
- n is the number of cells considered in a paging area;
- v is the mobile node speed (m/s);
- l is the cell perimeter (m);
- r_r is the average mobile registration refresh rate, which is related to registration lifetime and registrations triggered for other reasons (e.g., by an agent advertisement).

The transport of signalling messages over hops in the IP core network has different impact on the signalling cost in comparison to transporting signalling messages in local wireless access networks. The reason for this is that the signalling load introduced in the core IP network will potentially affect other networks and users, while signalling load on the access networks only has local effect. This is represented as weight factors w_{core} and w_{local} in equation (3). The hop weight relates to the hop length (i.e., distance), CPU packet processing time, queuing delay, etc. The hop ratios R_{core} and R_{local} indicate the percentage of hops in the core network and local access networks, respectively. The "local hops" include those hops located in both the foreign and home agent access networks.

In equation (3), the first term in brackets represents the signalling cost due to mobility. The second term represents the signalling cost due to registration refresh.

5.2.2. P-MIP

The signalling cost associated with P-MIP includes registration and paging signalling, and is

$$\begin{aligned} C_p &= d_{\mathrm{HA,FA}}(R_{\mathrm{core}} w_{\mathrm{core}} + R_{\mathrm{local}} w_{\mathrm{local}}) \\ &\quad \times \left[r_p + (r_c n - r_p)\alpha + \rho \left(\frac{l}{4}\right)^2 n r_r \right. \\ &\quad \left. + \rho \left(\frac{l}{4}\right)^2 n(1-\alpha)(\lambda_d + \lambda_a) \right] \\ &\quad + (n-1) d_{\mathrm{FA,FA}} w_{\mathrm{local}} \rho \left(\frac{l}{4}\right)^2 n(1-\alpha)\lambda_a \\ &= d_{\mathrm{HA,FA}}(R_{\mathrm{core}} w_{\mathrm{core}} + R_{\mathrm{local}} w_{\mathrm{local}}) \\ &\quad \times \left[\frac{\rho v L}{\pi} + \frac{\rho v l}{\pi}(n - \sqrt{n})\alpha + \rho \left(\frac{l}{4}\right)^2 n r_r \right. \\ &\quad \left. + \rho \left(\frac{l}{4}\right)^2 n(1-\alpha)(\lambda_d + \lambda_a) \right] \\ &\quad + (n-1) d_{\mathrm{FA,FA}} w_{\mathrm{local}} \rho \left(\frac{l}{4}\right)^2 n(1-\alpha)\lambda_a, \end{aligned} \quad (4)$$

where:

- C_p is the signalling cost for P-MIP (weighted hops·msg/s);
- $d_{\mathrm{HA,FA}}$ is the average distance between home and foreign agent hops;
- R_{core} is the ratio of the number of hops in the core network to the total number of hops between a home agent and a foreign agent;
- R_{local} is the ratio of the number of hops in local networks, such as access networks, to the total number of hops between a home agent and a foreign agent;

- w_{core} is the weight of each hop in IP core network;
- w_{local} is the weight of each hop in access networks;
- $d_{\text{FA,FA}}$ is the average distance between foreign agents (hops);
- r_{c} is the cell crossing rate (mobiles/s);
- r_{p} is the paging area crossing rate (mobiles/s);
- ρ is the mobile density (mobiles/m^2);
- n is the number of cells in a paging area;
- v is the mobile node speed (m/s);
- α is the ratio of active mobile nodes to total number of mobile nodes;
- λ_{a} is the incoming data session rate for a mobile node, it is also the paging rate for a mobile node (1/s);
- λ_{d} is the outgoing data session rate for an idle mobile node (1/s);
- l is the cell perimeter (m);
- L is the paging area perimeter (m); and
- r_{r} is the average mobile registration signalling refreshing rate which is related to registration lifetime and registrations caused for other reasons.

The first long term, including every short terms inside the brackets, and second long term of equation (4) represent the signalling cost associated with registration and paging. The short terms within the first long term represent the registration signalling cost caused by (i) crossing paging area boundaries; (ii) active mobile nodes crossing cell boundaries; (iii) registration refresh; and (iv), the registration signalling caused by incoming and outgoing data to and from idle mobile nodes, respectively.

For registration refresh, we may adjust the registration lifetime in P-MIP to make the refreshing cost in Mobile IP and P-MIP equivalent. Thus, in the signalling analysis, we ignore the effect of refresh signalling and focus on the signalling difference with and without paging. We normalize the signalling cost to the weighted distance between a home agent and foreign agent. Equations (3) and (4) reduce to

$$C_{\text{n}} = \frac{C}{d_{\text{HA,FA}}(R_{\text{core}}w_{\text{core}} + R_{\text{local}}w_{\text{local}})} = \frac{\rho v l}{\pi} n, \tag{5}$$

$$\begin{aligned} C_{\text{pn}} &= \frac{C_{\text{p}}}{d_{\text{HA,FA}}(R_{\text{core}}w_{\text{core}} + R_{\text{local}}w_{\text{local}})} \\ &= \frac{\rho v l \sqrt{n}}{\pi} + \frac{\rho v l}{\pi}\left(n - \sqrt{n}\right)\alpha \\ &\quad + \rho\left(\frac{l}{4}\right)^2 n(1-\alpha)(\lambda_{\text{d}} + \lambda_{\text{a}}) \\ &\quad + \frac{d_{\text{FA,FA}}(n-1)\rho(l/4)^2 n(1-\alpha)\lambda_{\text{a}}}{d_{\text{HA,FA}}(R_{\text{core}}R_{\text{w}} + R_{\text{local}})}, \end{aligned} \tag{6}$$

where the hop weight ratio R_{w} (discussed in section 5.3) is the ratio of w_{core} to w_{local}.

Table 3
Parameters.

R_{w}	8
R_{core}	0.5
R_{local}	0.5
$d_{\text{HA,FA}}$	16 hops
$d_{\text{FA,FA}}$	$\sqrt{n}$ hops
v	65 mph, i.e., 28.9 m/s
ρ	200 users/km^2, i.e., 0.0002 users/m^2
α	5% [3]
λ_{a}	3/hr, i.e., 0.0008/s
λ_{d}	3/hr, i.e., 0.0008/s

5.3. Analysis results

Wireless access networks can be divided into macro, micro and pico cellular systems representing cells found in traditional cellular networks, campus area networks and local area networks, respectively. We assume the following configurations for these systems. Cellular macro systems have an average perimeter cell size of 4000 m, maximum mobile node velocity of 100 mph (i.e., 44.5 m/s), user density of 200 users/km^2 (i.e., 200 users/cell). Cellular micro systems have an average cell perimeter of 400 m, maximum velocity is 20 mph (i.e., 8.9 m/s), user density is 20 users/cell (i.e., 2000 users/km^2). Finally, pico-cellular systems have an average cell perimeter size of 40 m, maximum velocity is 5 mph (i.e., 2.2 m/s), and a user density of 2 users/cell (i.e., 20000 users/km^2). We assume that the active mobile node percentage and data session rate are the same for all cell sizes.

Paging area size. In what follows, we show the relationship between the signalling cost and the number of cells in a paging area. Most parameters used in this evaluation are set to typical values found in the literature for analyzing cellular systems [14] (see table 3).

The average number of hops between foreign agents, $d_{\text{FA,FA}}$, is dependent on the network topology of the paging area. For simplicity we assume the average number of hops to be $\sqrt{n}$. The number and ratio of hops, $d_{\text{HA,FA}}$, R_{core}, R_{local}, and R_{w} were determined experimentally using the Internet traceroute tool to obtain an approximate picture of hop numbers and core/local hop distribution between Columbia University and other US universities. Typically, w_{core} is greater than w_{local}, resulting in a hop weight ratio (R_{w}) greater than 1. However, to accurately determine the hop weight ratio is difficult. Here, we set R_{w} to 8. The impact of different hop weight ratio R_{w} on the signalling cost is discussed later in this section.

As shown in figure 4, the signalling cost for Mobile IP has a linear relationship with the number of cells. The reasoning behind this is that the more cells considered, the more cell boundary crossings, which causes more registration signalling. When paging is introduced, signalling includes paging and registration. We observe that when the number of cells in the paging area is under a certain value, the signalling generated by paging remains small in comparison to the registration signalling overhead. Therefore, the total signalling

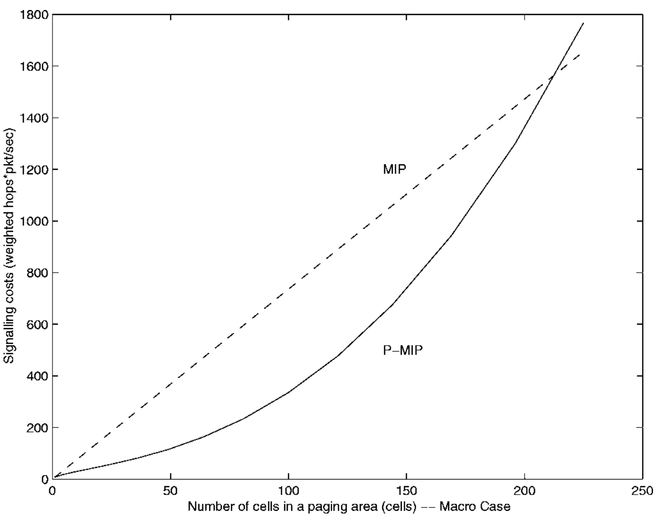

Figure 4. Effect of paging area size on signalling cost for macro cellular systems.

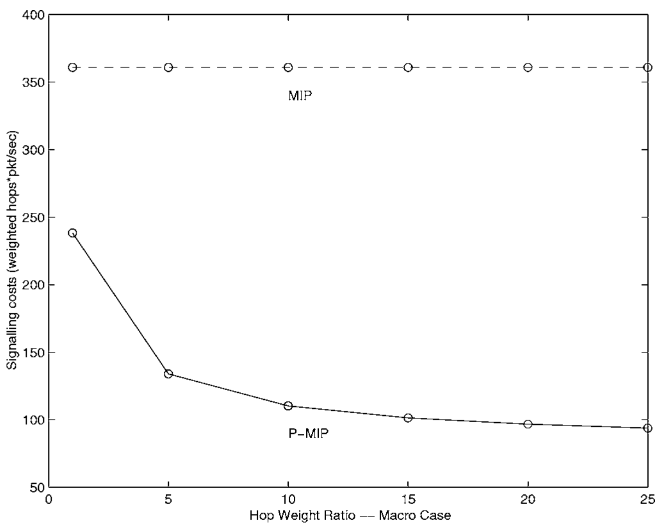

Figure 5. Effect of hop weight ratio on signalling cost for the macro cellular configuration.

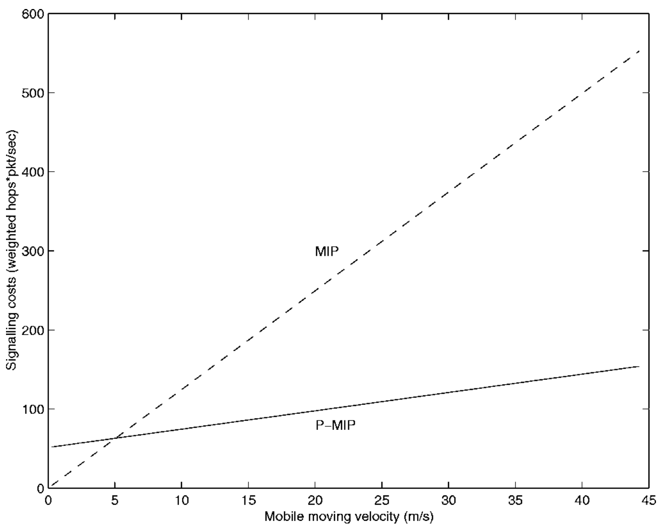

Figure 6. Effect of mobile speed on signalling cost for macro cellular configuration.

cost is reduced more than increased by the addition of paging. When the paging area has a large number of cells, however, the signalling cost introduced by paging grows quickly. This is due to the fact that paging request messages are unicast to all foreign agents in every cell in the paging area whenever paging is initiated. A number of observations can be made regards paging area size. The benefit of paging is limited when there are a small number of cells in a paging area. This is because of the large number of unnecessary registrations caused by boundary crossings. In contrast, the paging cost grows very quickly when there are a large number of cells in a paging area. As a result, this increases the total signalling cost rather than reducing it. Therefore, in order to optimize the total signalling cost and the system performance, it is important to select a suitable paging area size. The addition of multicast in support of paging would also improve the signalling overhead, as discussed earlier.

Figure 4 shows the signalling cost for P-MIP is smaller than Mobile IP when there are fewer than 225 cells in a paging area. However, the cost increases for P-MIP when there are more than 225 cells in a paging area. Today, the number of cells in a paging area varies depending on the service provider. A typical figure for the maximum number of cells in a paging area is around 50 cells [16]. When the paging area comprises 49 cells, the signalling cost is reduced from 400 weighted hops·pkt/s to 100 weighted hops·pkt/s.

Hop weight ratio (R_{w}). We set the number of cells in a paging area to 49 while other parameters remain the same as discussed above. The hop weight ratio (R_{w}) shows the different effects of paging and registration on signalling cost. When the ratio is small, local signalling (i.e., paging signalling) and core network signalling (i.e., registration signalling) have a similar impact on signalling cost. It is necessary to limit both paging and registration signalling in order to reduce the total signalling cost. As the hop weight ratio increases, the effect of paging signalling on signalling cost reduces, while the effect of registration signalling increases. Therefore, it is important to reduce the number of registrations to minimize the total signalling cost. In this case, an increase in paging can reduce the registration signalling. When the ratio increases to a certain level, registration signalling dominates the total signalling cost, and the paging signalling cost becomes negligible.

Figure 5 shows the signalling cost associated with P-MIP decreases as the hop weight ratio increases, while in contrast, the signalling cost of Mobile IP remains unchanged. Only registration signalling is used by Mobile IP where the weight ratio does not impact the total normalized signalling cost, see equation (5). In contrast, paging and registration are used by P-MIP where the weight ratio has impact on the total signalling cost.

Mobile node speed. Figure 6 shows that the impact of mobile node speed on signalling cost for P-MIP is minor in comparison to Mobile IP. If a mobile node moves faster than a certain speed (i.e., 5 m/s), then paging starts to significantly reduce the signalling overhead experienced. In contrast, when the mobile node is moving at a very low rate (e.g., less than 5 m/s), then Mobile IP performs better in terms of signalling cost. This is rather intuitive. Slow moving mobile nodes remain in the same cell for relatively longer periods of time. As a result, the registration signalling due to boundary crossing is small in comparison. Paging represents the major overhead because mobile nodes are likely to be located in the same network. There is a large amount of registration signalling generated by cell boundary crossings for fast moving mobile nodes. Paging brings significant benefits by reducing the number of registrations under these conditions.

Mobile density. In P-MIP, only a small percentage of mobile nodes are active, most nodes are idle. As discussed earlier, active mobile nodes in P-MIP are operationally equivalent to mobile nodes in MIP. The signalling cost of idle mobile nodes is relatively small in comparison to active mobile nodes. In P-MIP, an increase in mobile density only has impact on a smaller percentage of active users. In comparison, this affects all mobile nodes for MIP. As a result, signalling cost due to increased mobile density is more pronounced with MIP.

Operational state. Because Mobile IP does not distinguish between active and idle mobile nodes, signalling cost is unaffected by the number of active mobile nodes. In P-MIP, however, the signalling cost increases as the percentage of active mobile node increases. In the extreme case where all mobile nodes are active, the signalling cost of P-MIP and Mobile IP are equivalent. An estimate for the percentage of active nodes in cellular telephony is between 10–20% [16].

Data session rate. The data session rate has an impact on paging performance. However, the signalling cost is not affected by the data session rate in MIP. In P-MIP, the entire paging area is flooded with a paging request message when a new incoming data session arrives. As the arrival rate of new sessions increases so does the total signalling. For lower session arrival rates, P-MIP reduces the registration signalling with minimal increase in paging overhead. The effect of the outgoing data session rate is similar to that of the incoming session data rate, except that the incoming rate has more impact on the signalling cost. This is because paging is associated with the incoming rate and not the outgoing rate.

5.4. Simulation environment

We now study the performance of P-MIP using our *ns* simulation environment. This allows us to investigate scenarios that cannot be considered by analysis, and to validate our analytical results.

Constant bit rate (CBR) sources are used as traffic sources. The data session rate represents the number of times constant bit rate packet trains are sent to a mobile node during the simulation. Each data session lasts 10 s so that the active mobile node percentage is approximately 5%. The duration of each simulation experiment is 240 s.

The number of cells in a paging area is 4, 9, 16, 25, 36, and 49, respectively, where the paging area is square-shaped. The simulation area in each experiment models two complete paging areas. One base station supports a foreign agent in each cell. The total number of base stations is 8, 18, 32, 50, 72, and 98, respectively. Because setting up cells in the simulations is a time-consuming process, the number of cells in each paging area is not set to more than 49, and the same cell set-up is used for macro, micro, and pico cellular experiments. Velocity is modified proportionally in each case. The diameter of each cell is 20 and distance between two cell centers is 15, which simulates a cell coverage of 15 by 15. The unit of these parameters is dependent on the type of cellular coverage (i.e., macro, micro, or pico cellular) under study. In the macro, micro, and pico cellular configurations, an area of 15 by 15 represents 1000 m by 1000 m, 100 m by 100 m, and 10 m by 10 m, respectively.

A maximum velocity of 100 mph, 20 mph, and 5 mph is selected in the case macro, micro, and pico cellular networks, respectively. If we assume the simulation duration of 240 s is equivalent to one hour in reality, then the maximum velocity for macro, micro, and pico becomes 10 m/s, 5.2 m/s, and 13.3 m/s, respectively. By using the equivalency in cell size and simulation duration, the results for macro, micro, and pico cellular configurations can be seen as changing the mobile node speed while keeping the cell size fixed. Because the maximum speeds in macro and pico cellular cases are very similar, the signalling costs are very similar. Thus, in what follows, we only present the macro and micro cases.

A single mobile node under *ns* simulation can potentially consume all the available bandwidth of a base station [17]. This can result in collisions of signalling messages if two or more mobile nodes are transmitting or receiving at the same time in the same cell. The number of mobile nodes is limited to assure that the majority of registration messages do not suffer retransmission caused by collisions. The mobile node density is set to one mobile node in every four cells and does not change throughout the experiments. All mobile nodes have the same home agent in each simulation.

Even if the number of mobile nodes is not as large in comparison to the number of base stations, it is likely that more than two mobile nodes are in the same cell at the same time. This is because the mobile node's position and movement are generated randomly. If two mobile nodes are in the same cell, their signalling messages may collide. This will cause both mobile nodes to keep retransmitting registration messages until the signalling is successfully received. If more than two mobile nodes remain in the same cell for a long period time, then this situation can cause unproportionately high signalling in this experiment. However, because the same movement and traffic patterns are used for both Mobile IP and P-MIP experiments, the results are still comparable.

The data session rate for each mobile node is set to one in the paging area size experiment; that is, one constant bit rate packet train is destined to each mobile node for the duration of each simulation. We only consider the macro cellular configuration in this experiment and investigate the impact of the data session rate on signalling cost. The number of cells in a paging area is set to 36 and the data session rate ranges from 1–5. Each session lasts 10 s, which is related to the active mobile node percentage discussed in section 5.3.

5.5. Simulation results

We record registration and paging signalling for all experiments where the signalling cost is calculated based on these signalling counts. As discussed in section 5.3, the registration refresh signalling is not considered. In order to validate the analysis, we use the same assumptions and system configuration as discussed earlier whenever possible. The average numbers of hops is $d_{\mathrm{FA,HA}} = 16$ hops, the ratio is $d_{\mathrm{FA,FA}} = \sqrt{n}$ hops, hop weight ratio $R_{\mathrm{w}} = 8$. Half the hops between home agent and foreign agents are in the IP core network and half are in the wireless access networks, where $R_{\mathrm{local}} = R_{\mathrm{core}} = 0.5$.

Paging area size. All mobile nodes operate with a pause time of 0, supporting continuous motion. This movement pattern has some similarity to the fluid flow model used in the analysis in which all mobile nodes move constantly. Some differences exist, however. First, speed and direction are changing for each mobile node under simulation but are constant in the fluid flow model. Next, in the simulation environment, mobile nodes are confined to operate within two paging areas. In contrast, mobile nodes are moving into the paging area in the fluid flow model while other nodes are moving out.

Simulation results for macro and micro-cellular configurations are shown in figures 7 and 8, respectively. The simulation results follow the same trends shown in figure 4 when the number of cells in a paging area is fewer than 50. Differences are likely to be associated with the different mobility models used. Simulation results confirm that in order to improve system performance using paging, it is important to select the appropriate number of cells in a paging area.

Simulation and analysis results cannot be directly compared in a numerical sense because of the differences in the mobility models used. However, some coarse observations can be made. The paging overhead is smaller under simulation conditions when compared with the analysis. The reason for this is that mobile nodes are confined to operate within two paging areas under simulation. Therefore, we observe fewer paging area boundary crossings than analysis highlights. We also observe that paging provides more gain under fast mobility conditions. The reason for this is intuitive. The faster the mobiles move, the more boundary crossings experienced and registration signalling generated. Thus, paging can reduce signalling cost under higher mobility conditions.

We also investigate the impact of the discrete movement of mobile nodes that cannot be observed by using the fluid flow model used in the analysis. All mobile nodes are now set to operate with a pause time of 60, supporting discrete motion, in contrast to the continuous movement discussed above. Mobile nodes move to a destination, stay there for certain period of time and then move again. The fluid flow model is more suitable to movement found in cellular networks, while the way point model with non-zero pause times is more applicable to the mobility that may be typical with Internet.

Figure 9 shows that paging can reduce the signalling cost for mobile nodes that move infrequently. However, comparing the results shown in figures 7 and 8, we observe that the paging gain is less than the case of more frequently moving mobile nodes. The reason for this behavior is similar to that observed with fast and slow moving mobile nodes. For mobile nodes that move less, the longer the mobile node remains in the same cell. As a result, there are fewer cell boundary crossings and registrations observed. Under these conditions, there is less need for paging. Paging becomes the sole overhead for nodes that do not move at all. However, such stationary conditions are an anomaly in mobile communication networks.

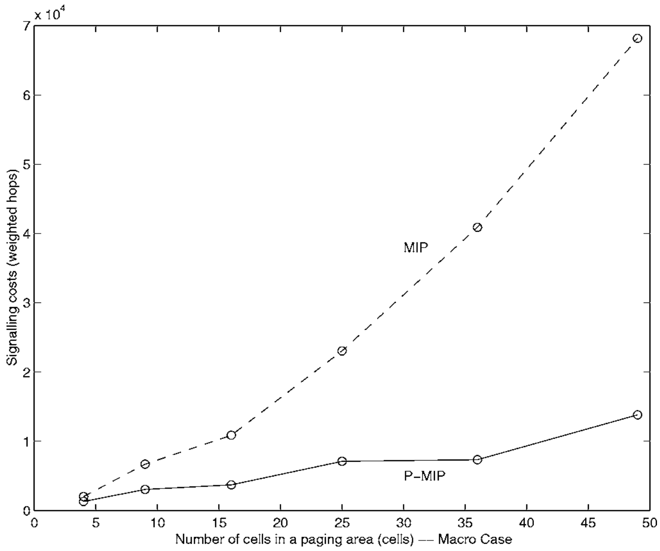

Figure 7. Effect of paging area size on signalling cost for the macro cellular configuration.

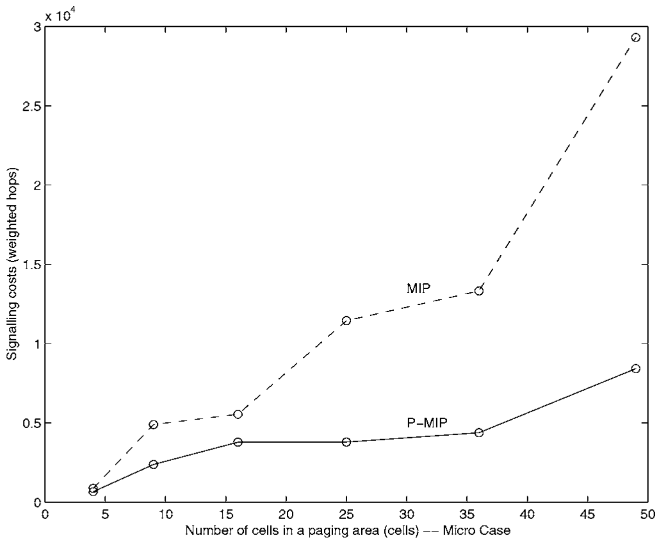

Figure 8. Effect of paging area size on signalling cost (micro case).

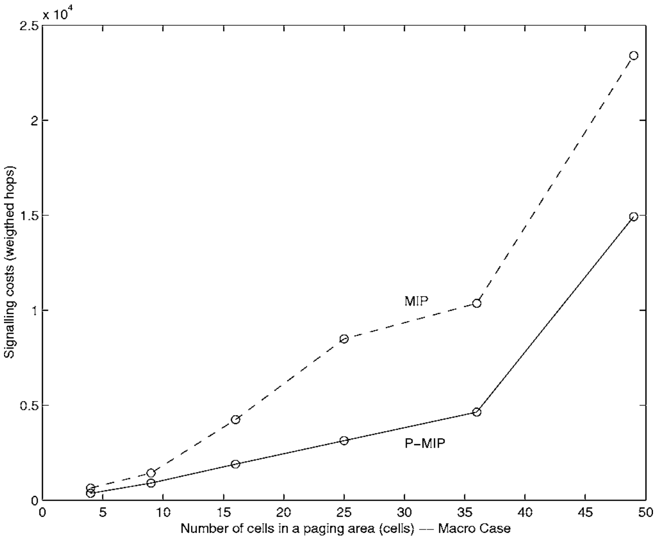

Figure 9. Effect of paging area size on signalling cost with 60 s pause for the macro cellular configuration.

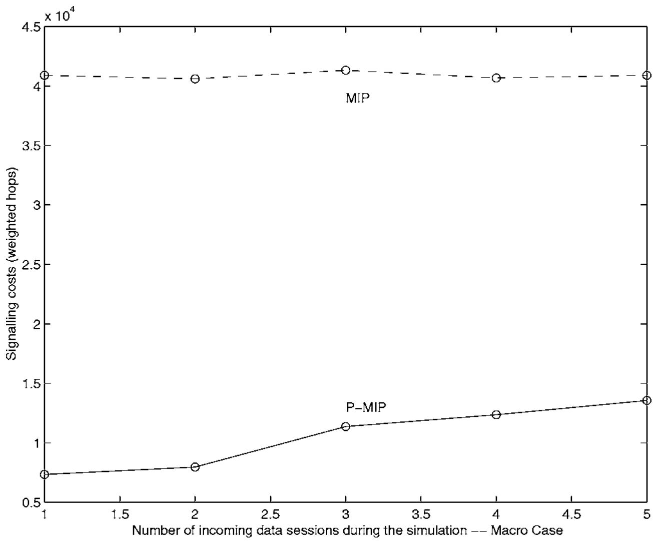

Figure 10. Effect of data session rate on signalling cost.

Data session rate. In this experiment, we consider the effect of the data session rate on the total signalling cost. Simulation and analysis consider an active mobile node percentage to be approximately 5% [3]. This is configured as follows. For analysis, we explicitly specify the parameter α. Under simulation, we obtain this implicitly by setting the data session holding time. For this experiment, the pause time is set to 0 and paging area size to 49.

We observe that the simulation results, shown by figure 10, follow the same trends as the analysis results discussed in section 4.

From simulation, we conclude that paging reduces the signalling cost significantly when the data session rate is low. In this case, paging reduces unnecessary registration without introducing too much paging overhead. More paging is required at higher data rates, and as a result, paging increases the signalling cost. Paging eventually degrades the system performance at very high data session rates. Signalling cost in Mobile IP does not change with the data session rate, as indicated in section 5.3 and figure 10.

6. Conclusion

In this paper, we have introduced a set of simple paging extensions for Mobile IP protocol called P-MIP. While P-MIP extends the base protocol it is also backward compatible with Mobile IP. Therefore, Internet Service Providers have the freedom to configure the P-MIP paging capability on an as needed basis without affecting the operation of the entire Mobile IP enabled Internet.

Both the analysis and simulation results presented in this paper show that by carefully selecting suitable system configuration parameters (e.g., paging area size, active timer, etc.), paging can reduce unnecessary registration, thereby reducing the overall signalling cost with the result of improved system performance. Paging also minimizes the impact of mobility rate, cell size and mobile node density on the signalling cost.

While the analysis and results presented in the paper highlight signalling savings, P-MIP is designed to provide power savings at mobile nodes through reduction of registration signalling. In the next phase of our work, we will study how to best interface P-MIP to mobile devices in a simple manner in support of power efficient operations. Finally, we are developing a testbed implementation of the protocol and plan to publish the code in the public domain in due course. Until then the source code for *ns* P-MIP extensions used to evaluate the protocol in this paper is available on the Web (`comet.columbia.edu/pmip`).

Acknowledgements

The work is supported in part by the NSF Wireless Technology under Award AWI-9979439. Aspects of the work discussed in this paper were part of a keynote by Andrew T. Campbell given at the Third International Workshop on Modeling Analysis and Simulation of Wireless and Mobile Systems (MSWiM), August, Boston, 2000. Finally, the authors would like to thank Prof. Mischa Schwartz for his input.

Appendix. Messages

A.1. New ICMP router discovery message extension

- PAI extension

```
0                   1                   2                   3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Type       |    Length     |        PAI                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

This PAI extension is used in non-overlapping paging area construction scheme. It should be added in the agent advertisement message, and periodically broadcast on the network.

A.2. The new control messages

- Paging Request

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      | Length        |      Sequence number          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    the home address of paged mobile node 1                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    the home address of paged mobile node 2                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                              ...                              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Extensions ...
+-+-+-+-+-+-+-+-+
```

Length $= 2 + 4 \cdot N$ bytes, where N is the number of mobile nodes paged. Length does not cover the type, length, and extensions fields. Sequence number is the count number of Paging Request messages sent to a mobile node. It is used to distinguish the Paging Request messages. If the same Paging Request message is sent to the same mobile node more than once, the sequence number is the same.

- Paging Reply

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      | Length        |           reserved            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    the home address of paged mobile node 1                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    the care-of address of paged mobile node 1                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Extensions...
+-+-+-+-+-+-+-+-
```

Length $= 2 + 8 \cdot N$ bytes, where N is the number of responding mobile nodes. Length does not cover the type, length, and extensions fields.

A.3. New extension in Mobile IP control message

- Foreign–Foreign Authentication extension

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      |     Length    |     SPI...
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   ... SPI (cont.)              |  Authenticator ...
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- Paging Area ID extension

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      | length        |           reserved            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    IP address of FA1                                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    IP address of FA2                                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    IP address of FA3                                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|
+-+-+-+-+-+-+-+-
```

Paging Area ID extension is used in overlapping paging area construction scheme. This extension represents a list of FA IP addresses in the same paging area. It should be inserted in the Registration Reply message.

A.4. Modifications to existing Mobile IP messages

- Registration Request message

```
                    the modified bit
                           |
 0                   1     V             2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      |S|B|D|M|G|V|P|r|          Lifetime             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                          Home Address                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                           Home Agent                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                        Care-of Address                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
+                         Identification                        +
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Extensions ...
+-+-+-+-+-+-+-+-
```

Change one of the reserved bits in the Registration Request message used in Mobile IP to the "P bit" which indicates if the mobile node supports the paging function or not. A P bit of '1' means that the mobile node supports paging and a P bit of '0' means that the mobile node does not support paging.

- Mobility Agent Advertisement extension

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      |    Length     |        Sequence Number        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Registration Lifetime     |R|B|H|F|M|G|v|P|   reserved    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 Zero or More Care-of Addresses ^              |
|                              ...               |              |
                                                 |
                                          the modified bit
```

Change one of the reserved bits in the Agent Advertisement used in MIP to the "P bit" which indicates if the foreign agent supports the paging function or not. A P bit of '1' means that the foreign agent supports paging and a P bit of '0' means the foreign agent does not support paging.

References

[1] J. Broch, D. Maltz, D. Johnson, Y.-C. Hu and J. Jetcheva, A performance comparison of multi-hop ad hoc network routing protocols, in: *Proceedings of the Fourth Annual ACM/IEEE International Conference on Mobile Computing and Networking,* ACM, Dallas, TX (October 1998).

[2] T. Brown and S. Mohan, Mobility management for personal communications systems, IEEE Transactions on Vehicular Technology 46(2) (May 1997) 269–278.

[3] A. Campbell, J. Gomez, C.-Y. Wan, Z. Turanyi and A. Valko, Cellular IP, Internet Draft, IETF (1999).

[4] C. Castelluccia, Extending Mobile IP with adaptive individual paging: A performance analysis, INRIA (November 1999) `http://www.inrialpes.fr/planete/people/ccastel/Welcome.html\#PUB`
[5] K. Fall and K. Varadhan (eds.), *ns* notes and documentation, `http://www-mash.cs.berkeley.edu/ns`
[6] M.D. Gallagher and R.A. Snyder, *Mobile Telecommunications Networking with IS-41* (McGraw-Hill, New York, 1997).
[7] A. Campbell and J. Gomez, IP micro-mobility protocols in the IETF, in: *ACM SIGMOBILE MC2R*.
[8] E. Gustafsson, A. Jonsson and C.E. Perkins, Mobile IP regional registration, Internet Draft, IETF (March 2000).
[9] G. Heine, *GSM Networks: Protocols, Terminology, and Implementation* (Artech House, 1998).
[10] S. Mohan and R. Jain, Two user location strategies for personal communications services, IEEE Personal Communications (First Quarter 1994) 42–50.
[11] C. Perkins, IP mobility support, RFC 2002, IETF (October 1996).
[12] R. Ramjee, T. La Porta and L. Li, Paging support for IP mobility using HAWAII, Internet Draft, IETF (December 1999).
[13] R. Ramjee, T. La Porta, S. Thuel, K. Varadhan and L. Salgarelli, IP micro-mobility support using HAWAII, Internet Draft, IETF (December 1999).
[14] R. Ramjee, T. La Porta, S. Thuel, K. Varadhan and S.Y. Wang, HAWAII: A domain-based approach for supporting mobility in wide-area wireless networks, in: *International Conference on Network Protocols, ICNP'99* (1999).
[15] S.H. Redl, M.K. Weber and M.W. Oliphant, *An Introduction to GSM* (Artech House, 1995).
[16] A.G. Valko, Personal communication (September 2000).
[17] J. Widmer, Network simulations for a mobile network architecture of vehicles, `http://www.icsi.berkeley.edu/~widmer/mnav/ns-extension`
[18] P-MIP *ns*-2 source code `http://comet.columbia.edu/pmip`

Xiaowei Zhang received her B.S. degree from Hebei Institute of Technology in Tianjin, China, and Master's degree from University of Mississippi, US, both are in electrical engineering. Currently, she is a Ph.D. student in the COMET group at the Center for Telecommunications Research, Columbia University, New York. Her research areas cover pervasive networking and Mobile IP.
E-mail: xzhang@comet.columbia.edu

Javier Gomez Castellanos obtained a B.S. degree with honors in electrical engineering in 1993 from National Autonomous University of Mexico (UNAM), Mexico, and M.S. degree in electrical engineering in 1996 from Columbia University, New York. Since 1996 he has been a Ph.D. student in the COMET Group at the Center for Telecommunications Research, Columbia University, New York. His research interests cover routing, QoS, and power-aware design for cellular and ad hoc wireless networks. He is a student member of the IEEE.
E-mail: javierg@comet.columbia.edu

Andrew T. Campbell is an Associate Professor in the Department of Electrical Engineering and a member of the COMET Group at the Center for Telecommunications Research, Columbia University, New York. His research interests include open programmable networks, mobile networking, and quality of service. He is active in the IETF and the international working group on Open Signaling (OPENSIG). Andrew currently serves as a technical program co-chair for 8th ACM Conference on Mobile Computing and Networking (ACM MobiCom 2002). He is a past co-chair of the 5th IFIP/IEEE International Workshop on Quality of Service (IWQOS'97) and the 6th IEEE International Workshop on Mobile Multimedia Communications (MOMUC99). He received his Ph.D. in computer science in 1996 and the NSF CAREER Award for his research in programmable mobile networking in 1999.
E-mail: campbell@comet.columbia.edu