Mobility management techniques for the next generation wireless networks

Jun-Zhao Sun[*a], Douglas Howie[**a], and Jaakko Sauvola[***a]
[a]MediaTeam, Machine Vision and Media Processing Unit, Infotech Oulu
University of Oulu, Finland

ABSTRACT

The tremendous demands from social market are pushing the booming development of mobile communications faster than ever before, leading to plenty of new advanced techniques emerging. With the converging of mobile and wireless communications with Internet services, the boundary between mobile personal telecommunications and wireless computer networks is disappearing. Wireless networks of the next generation need the support of all the advances on new architectures, standards, and protocols. Mobility management is an important issue in the area of mobile communications, which can be best solved at the network layer. One of the key features of the next generation wireless networks is all-IP infrastructure. This paper discusses the mobility management schemes for the next generation mobile networks through extending IP's functions with mobility support. A global hierarchical framework model for the mobility management of wireless networks is presented, in which the mobility management is divided into two complementary tasks: macro mobility and micro mobility. As the macro mobility solution, a basic principle of Mobile IP is introduced, together with the optimal schemes and the advances in IPv6. The disadvantages of the Mobile IP on solving the micro mobility problem are analysed, on the basis of which three main proposals are discussed as the micro mobility solutions for mobile communications, including Hierarchical Mobile IP (HMIP), Cellular IP, and Handoff-Aware Wireless Access Internet Infrastructure (HAWAII). A unified model is also described in which the different micro mobility solutions can coexist simultaneously in mobile networks.

Keywords: mobility management, macro mobility, micro mobility, Mobile IP, HMIP, Cellular IP, HAWII

1. INTRODUCTION

With the increasing variety of mobile devices such as Personal Digital Assistants (PDAs), laptop computers, and digital cellular phones, the traditional viewpoint on the Internet has begun to be changed by the new scenarios of the mobile applications. In the near future, more and more Internet services of both conventional and novel types can be smoothly accessed with various mobile devices through the wide deployed wireless networks. The Internet is extending its coverage area into a more spacious and attractive field, which brings service providers and network operators more opportunities of commercial proliferation. To users, this means more benefits and conveniences in their work and daily life.

In the area of telecommunication, mobile communications and wireless networks are continuously one of the hottest topics that are developing at a booming speed. One commonly adopted piece of jargon in industry is to label the evolution of mobile communications with G's. The first-generation (1G) denotes analogue networks such as Nordic Mobile Telephone (NMT) and second-generation (2G) denotes the first digital solutions, such as Global System for Mobile Communications (GSM). Recently introduced General Packet Radio Service (GRPS) and related GSM extensions have been referred to as 2.5G, and finally the Universal Mobile Telecommunication System (UMTS) and International Mobile Telecommunication System 2000 (IMT-2000) [1] efforts as the third-generation (3G). At present, 3G mobile communication systems are just beginning to be deployed , while research on the next generation of mobile communications, the fourth-generation (4G) wireless networks, begins to pave the way for the future.

---

[*] junzhao.sun@ee.oulu.fi;
[**] douglas.howie@ee.oulu.fi;
[***] jaakko.sauvola@ee.oulu.fi;
phone 358 8 553-2532; fax 358 8 553-2534; http://www.mediateam.oulu.fi/; MediaTeam, University of Oulu, P.O.Box 4500, 4SOINF, FIN-90014 University of Oulu, Finland

The boundary between mobile personal telecommunications and wireless computer networks is disappearing, through the converging of mobile and wireless communications with Internet services. With the rapid improvement in both wireless networks and mobile terminals, great increases have emerged in all the fields of mobile communications, including the number of mobile subscribers, the deployment of mobile communication systems, the new advances in mobile techniques, along with the miscellaneous mobile services. The next generation of wireless communications will be based on a global system of both wired fixed and wireless mobile networks and services.

Regardless of the services and networks, one of the most important and challenging problems for the seamless access of wireless networks and mobile services is mobility management. Mobility management is the fundamental technology used to automatically support mobile terminals enjoying their services while simultaneously roaming freely without the disruption of communications. Two main aspects need to be considered in mobility management, i.e. location management (addressing, location tracking and update, locating and paging, etc.) and handoff management (handoff initiation, connection routing, smoothing, etc.). To the coming 4G mobile communications, mobility management becomes more severe a problem that needs to be carefully and perfectly solved with more intensive efforts, in order to catch the trends of the internetworking of Internet with cellular networks, the wider range of moving speed and area, and the requirement of seamless handoff for real-time and multimedia applications.

The protocol of IP plays an important role in the field of mobility management in various types of wireless networks. One of the key features of 4G is that it will be based on an all-IP infrastructure for both fixed and mobile networks. Moreover, positioning on network layer, IP acts as a masking isolator that prevents the protocols, services, and applications of upper layers from the awareness of network interconnecting architecture and possible changes caused. So, as a suitable layer to solve the problem of mobility and provide transparent mobility to applications and higher level protocols like TCP, IP becomes one of the most important research issues in mobility and location management, which results in various techniques and standards based on the extension of fundamental IP protocol proposed.

The IP extensions for mobility solution are mainly carried out at the working group of IP Routing for Wireless/Mobile Hosts (mobileip) in the Internet Engineering Task Force (IETF) [2]. The main goal of the mobileip working group is to develop routing support to permit IP nodes (hosts and routers) using either IPv4 or IPv6 to seamlessly "roam" among IP subnetworks and media types. The proposed Mobile IP method supports transparency above the IP layer, including the maintenance of active TCP connections and UDP port bindings. Several protocols proposed, in form of both Internet Draft and Request For Comments (RFC), were devoted to all the aspects of mobility management. Moreover, great efforts for protocol standardization have also been made in IP-based mobile telecommunications networks. In this movement toward 3G wireless, the two partnership projects that address the issue of standard development, 3rd Generation Partnership (3GPP) [3] and 3rd Generation Partnership Project 2 (3GPP2) [4], together with Mobile Wireless Internet Forum (MWIF) [5], are all moving toward an all-IP mobile network architecture [6].

This paper discusses the mobility management schemes for the next generation mobile networks. A global framework for the mobility management of wireless networks is presented, in which the mobility management is divided into two complementary tasks: macro mobility and micro mobility. The work presented in the paper is based mainly on, besides others, four of the most outstanding proposals of IETF outlining the layout routing strategy for the IP based wireless networks, including Mobile IP [7], Hierarchical Mobile IP (HMIP) [8], Cellular IP [9], and HAWAII [10]. Mobile IP, together with the optimisation schemes, serves as the macro mobility solution for the inter-domain roaming, while the other three proposals are used as the micro mobility solutions for the intra-domain movement. The salient features of each protocol proposal are clearly described.

The paper is organized as follows: In Section 2, some basic issues on mobility management are discussed, including the concept and operations together with the requirements and solutions of mobility management. In Section 3, we introduce the general mobility model and then present a mobility framework as a hierarchical hybrid model used to globally describe the mobility issue and management tasks, in which the mobility issue is divided into two distinct problems, i.e. macro mobility and micro mobility. Section 4 describes the protocol of Mobile IP as the macro mobility solution, in which the optimal schemes, the differences between IPv4 and IPv6, and the disadvantages of the proposal on performance concerns are discussed. The micro mobility management mechanisms are introduced in Section 5, including the three main protocols proposed, followed by the discussion on the coexistence of the different solutions. Finally, we summarize the paper by conclusions in Section 6.

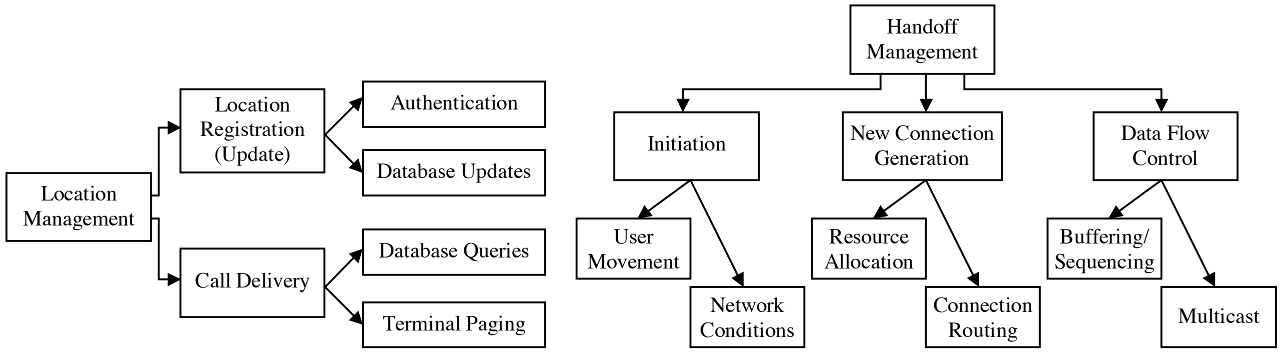
2. BASIC ISSUES OF MOBILITY MANAGEMENT

2.1 Concept and operations

Mobility management is the essential technology that supports roaming users with mobile terminals to enjoy their services in progress through wireless networks. From the viewpoint of functionality, mobility management enables communication networks to:

- Locate roaming terminals in order to deliver data packets, i.e. function for static scenario.
- Maintain connections with terminals moving into new areas, i.e. function for dynamic scenario.

According to the concept above, mobility management contains two distinct but related components: location management and handoff management. The former concerns how to locate a mobile node, track its movement, and update the location information, while the later focuses mostly on the control of the change of a mobile node's access point during active data transmission. The operations of the two components are shown in Figure 1 [11]. Note that many issues in location management are not protocol dependent, while handoff algorithms are much related to the network protocols of e.g. routing and resource management.



(a) Location management operations. (b) Handoff management operations.

Figure 1: Mobility management operations.

Mobility affects the whole protocol stack, from the physical, data link, and network layers up to the transport and application layers. Examples include radio resource reuse at the physical layer, encryption and compression at the data link layer, congestion control at the transport layer, and service discovery at the application layer. Network layer offers routing from one network to another through independent links. Since mobility, modelled as changing node's point of attachment to the network, is essentially an address translation problem, it is therefore naturally best resolved at the network layer by changing the routing of datagrams destined to the mobile node to arrive at the new point of attachment [12]. To implement mobility management at network layer may also shield the upper-level protocols from the nature of the physical medium and make mobility transparent to applications and higher-level protocols such as TCP.

2.2 Requirements and solutions

Besides the basic functions that implement the goal of mobility management, there are many other requirements on performance and scalability that should be carefully taken into account when trying to design or select a mobility management scheme, including:

1) Fast handoff: the handoff operations should be quick enough in order to ensure that the mobile node can receive IP packets at its new location within a reasonable time interval and so reduce the packet delay as much as possible.
2) Seamless handoff: the handoff algorithm should minimize the packet loss rate into zero or near zero which, together with fast handoff, is sometimes referred to as smooth handoff.
3) Signalling traffic overhead: the control data load, e.g. the number of signalling packets or the number of accesses to the related databases, should be lowered to within an acceptable range.

4) Routing efficiency: the routing paths between the communication nodes to the mobile nodes should be optimised to exclude redundant transfer or bypass path as e.g. triangle routing.
5) Quality of Service (QoS): the mobility management scheme should support the establishment of new QoS reservation in order to deliver a variety of traffic, while minimising the disruptive effect during the establishment.
6) Fast security: the mobility scheme should support different levels of security requirements such as data encryption and user authentication, while limiting the traffic and time of security process e.g. key exchange.
7) Special support required: it is better for a new mobility mechanism to require minimal special changes on the components, e.g. mobile node, router, communication media, networks, other communication nodes, etc.

There are many distinct but complementary techniques especially for mobility management to achieve its performance and scalability requirements listed above, including:

- Buffering and forwarding, to cache packets by the old attachment point during the MN in handoff procedure, and then forward to the new attachment point after the processing of MN's handoff.
- Movement detection and prediction, to detect and predict the movement of mobile host between difference access points so that the future visited network is able to prepare in advance and packets can be delivered there during handoff.
- Handoff control, to adopt different mechanisms for the handoff control, e.g. layer two or layer three triggered handoff, hard or soft handoff, mobile-controlled or network-controlled handoff.
- Paging area, to support continuously reachable with low overhead on location update registration through location registration limited to the paging area .
- Domain-based mobility management, to divide the mobility into micro mobility and macro mobility according to whether the mobile host's movement is intra-domain or inter-domain. This is the basic architecture studied in this paper, and it is described in more detail in next section.

3. MOBILITY MANAGEMENT MODELS

3.1 General mobility management model

A general mobility model is presented in Figure 2. Typically, we suppose in the paper that the public network is the Internet and so IP is the network layer protocol. The components in the model are introduced as follows:

1. Two mobile network entities. Mobile Node (MN) is a node (terminal device) that can change its point of attachment to the network from one link to another by freely roaming with its user, while still being reachable. Corresponding Node (CN) is either a mobile or a stationary node that can communicate with the concerned mobile node by sending or receiving packets to or from the mobile node.

2. Two networks. Home network is the unique network at which the mobile node is continually reachable to the other corresponding nodes, by an originally assigned address – Home Address (H@). Foreign network is the network to which the mobile node is currently attached instead of its original home network, and is reachable by a new generated address – Care-of Address (CoA).

3. Two addresses. Home Address (H@) is the static unchangeable IP address assigned to the MN that is used to identify the end-to-end connection, also as the IP address when MN is in its home network. Care-of Address (CoA) is the IP address to identify the MN's current point of attachment to the Internet when it is in a foreign network.

4. Two mobility agents. Home Agent (HA) is a router on the home network that makes the mobile node reachable when the mobile node is attached to a foreign network. Foreign Agent (FA) is a router on the foreign network that assists the mobile node to access the Internet by receiving datagrams delivered to the Care-of Address.

The basic assumption within IP is that an IP address always identifies the node's location in the Internet. So if a node moves to another location in the Internet, a new IP address must be generated and assigned to the node in order to route following IP packets to its new attachment point. So the basic function of mobility management in the Internet is just to translate a node's original IP address (Home Address, H@) into the new temporary IP address (Care-of Address, CoA) as promptly and efficiently as possible.

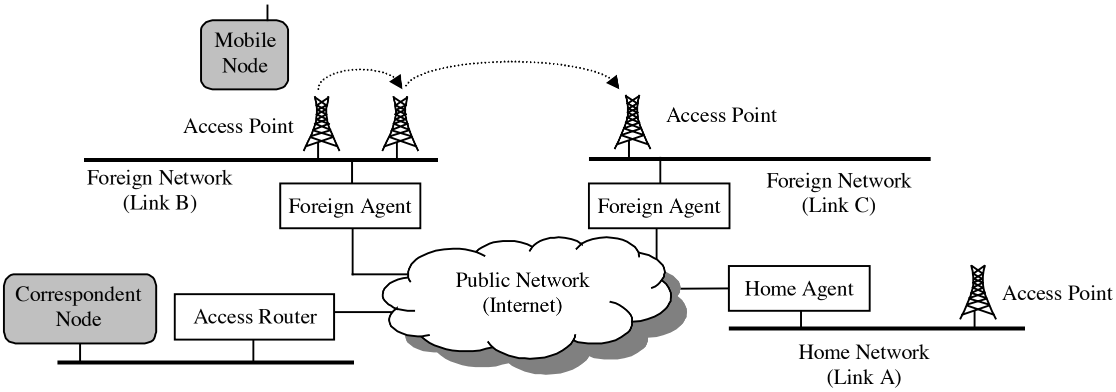


Figure 2: General mobility model.

3.2 Hierarchical mobility management model

The basic idea behind the domain-based mobility management scheme introduced in Section 2.2 is that the mobility management strategy should be based on a hierarchical mobility management scheme that localizes the management of mobility by introducing the concept of domain, in order to achieve the requirements on performance and flexibility especially for frequently moving hosts. With this in mind, two kinds of mobility can be defined as follows, according to the movement span:

1) Micro mobility, i.e. mobile node's movements inside a domain, to which intra-domain mobility management solutions are suitable, focusing mainly on a fast, efficient, seamless mobility support within a restricted coverage.
2) Macro mobility, i.e. mobile node's movements between different domains, to which inter-domain mobility management schemes can be employed, acting as a global mobility solution with the advantages of flexibility, robustness, and scalability.

A domain in the above discussion is defined as a collection of networks sharing a common network administration, which may include one or more foreign (visited) networks. Figure 3 shows the hierarchical mobility management model, as the architecture studied accordingly in this paper. It is worthy to note that the concepts micro and macro mobility based on the definition of domain are possibly recursive: a movement may be micro in one domain whereas macro in another. We introduce the proposed schemes for the implementation of these two types of mobility management in the following two sections respectively.

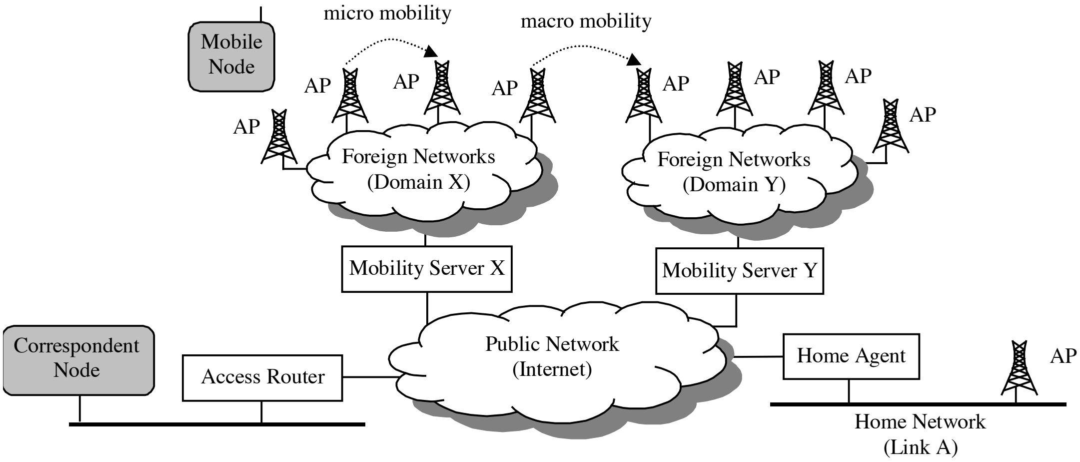


Figure 3: Hierarchical mobility management model.

4. MACRO MOBILITY SOLUTION: MOBILE IP

As the macro mobility solution, we introduce Mobile IP proposal [7] in this section, together with some optimisations on it. An emphasis is also given to Mobile IPv6 that presents new advanced features and represents the future trend.

4.1 Basic principles of Mobile IP

As shown in Figure 2, the basic Mobile IP components include, aside from the Mobile Node (MN), the Home Agent (HA) in the home network, the Foreign Agent (FA) in the foreign network, the Correspondent Node (CN) and the Access Router (AR). The MN has a Home Address (H@), typically registered in a Domain Name Server (DNS). The upper 8 bytes of the H@ match the home subnet prefix of the MN's home link. The home link has at least one AR that can offer HA services to the MN. When the MN moves out of the home network into any other foreign network, it can be reached through obtaining a Care-of Address (CoA). The general working process of Mobile IP is as follows:

a. The mobility agents (HA and any FA) continuously send agent advertisement messages to the local link, to announce their existences.
b. The MN receives the advertisement messages, according to which it can determine whether its current attachment point is in the home network or any foreign network.
c. If the MN finds itself still in its home network (the advertisement message comes from the HA), then it will not start the mobility management functions and the process is terminated; if it comes back from a foreign network, it should cancel the registration on the HA by sending a corresponding message to the HA.
d. If the MN finds itself in a foreign network, then it will obtain a CoA that is either the IP address of the corresponding FA or a new temporary IP address.
e. The MN should then register to its HA to inform it of the new CoA.
f. After successful registration, data packets from any CN originating to the MN's H@ are intercepted and sent by the HA to the MN through "IP tunneling".
g. The packets sent by the MN to any CN will be routed as normal, and do not need the assistance from the HA.

According to the general working procedure, four main cooperating mechanisms are the keys to understand Mobile IP, including movement detection and agent discovery, CoA and obtaining it, CoA registration, and tunneling and routing.

1. Movement detection and agent discovery. Mobile agents make themselves known by periodically sending mobility agent advertisement messages – the extension of Router Advertisement specified in Internet Control Message Protocol (ICMP) [13]. The message may contain local link number, one or more available CoAs, and information on some special features. If an MN cannot receive the agent advertisement from its original agent for a specified time, it will assume to have moved into a new area. In this case the MN may passively wait for another agent advertisement. When an MN needs to get a new CoA but does not want to wait for the periodic agent advertisement, it can also actively broadcast an agent solicitation to solicit a new mobility agent in order to get a new CoA.

2. CoA and obtaining it. A CoA can be of two different types: FA's IP address or Co-located CoA (CCoA). A CoA can just be the FA's IP address, in which multiple MNs share a same CoA. By this mode the limited IP address space can be saved. Another type of CoA (CCoA) is a temporary IP address that is assigned to one of the MN's network interfaces by some external assignment mechanism such as Dynamic Host Control Protocol (DHCP). For a CCoA, all the natural FA functions have to be performed by the MN itself.

3. CoA registration. Once an MN has a new CoA, a registration process will be initialised to inform the HA of the new IP address. First the MN (when CCoA is used) or the FA sends to the HA a registration request which is known as Binding Update (BU). A BU can be treated as a triplet containing the MN's HA, CoA, and registration lifetime. Once the HA receives the BU, it updates the corresponding entries in its routing table and approves the request by sending a registration reply back to the MN. On the other hand, the HA may send a binding request to the MN to get the MN's current binding information if the latest BU expires. The MN then responds to the binding request with its new BU and waits for a binding acknowledgment from the HA.

4. Tunneling and routing. After the registration of a new CoA, the further packets destined to the MN's H@ will be redirected to the MN's CoA by constructing a new IP header – tunnel header – that contains the MA's CoA as the destination IP address. The new datagram then encapsulates the original packet with the new header, i.e. IP-within-IP.

Then an IP tunnel is built, with the HA as the tunnel source and the CoA as the tunnel destination. When the tunnel destination – either the FA or the MN – receives the new packet, it merely eliminates the tunnel header to recover the original packet and deliver the rest to the MN.

Detailed description of Mobile IP can be found in the protocol specification in [7].

4.2 Mobile IP optimizations

There are some problems in the basic Mobile IP specification that are to be improved by optimal schemes. Two main problems and corresponding optimal protocols are discussed below.

1. Triangle Routing problem. According to the basic Mobile IP protocol, while the MN can send out packets (may be through the FA) along an optimal path that directly route to the CN, the incoming packets from the CN to the MN have to firstly arrive at the HA in order to use IP tunnelling. This is called the Triangle Routing problem. When the current location of the MN is quite close to the CN but the HA is very far away, datagrams need to take a long way.

Mobile IP route optimisation [14] proposes a new mechanism to alleviate the Triangle Routing problem. Any CN maintains a binding cache. A BU message will be sent to the source CN to inform the MN's new CoA by the FA or the MN itself. The CN then updates the binding cache and tunnels any later packets directly to the MN's current CoA without bypassing the HA anymore.

2. Smooth Handoff. During the MN's handoff, many operations should be implemented together with messages to be sent, e.g. movement detection and FA discovery, registration and BUs. Before the HA (and the CNs) is informed of the MN's new CoA by BU, the packets within the handoff interval will be lost.

The process of smooth handoff [14] tries to overcome this disadvantage by optimising the basic Mobile IP standard. As an extension of the registration process, the MN's new FA may send a Previous FA Notification message, including a BU, to the previous FA. The previous FA then updates its binding cache and re-tunnels any packets destined to the MN to its new CoA. Moreover, the previous FA may also buffer the packets destined to the MN that arrive before the previous FA notification finishes, and re-tunnels the buffered packets to the MN's new CoA later after receiving the Previous FA Notification message.

4.3 Advances in Mobile IPv6

IPv6 is defined in the IETF working group of IP Next Generation (ipngwg) [15], by providing enhancements over the capabilities of existing IPv4 service. Basic improvements to IPv4 include optimal header format [16], reasonable addressing architecture [17], neighbour discovery mechanism [18], stateless auto-configuration [19], and security and QoS support. Mobility support in IPv6 [20] takes full advantage of these enhancements. The basic ideas behind the Mobile IPv6 protocol are the same as in IPv4. Besides, there are some main changes in Mobile IPv6 standard, described as follows:

1. In Mobile IPv6, since the requirement in IPv4 of sharing IP address in order to save the limited address space has disappeared, there is no mandatory need for FA anymore. In this case the MN should process most of the FA's function instead.

2. Support for mobility is a built-in feature as a fundamental part of the IPv6. A new CoA can be generated by the IPv6 mechanism of stateless auto-configuration [19], and IPv6-within-IPv6 tunnelling is also already specified in [21] to tunnel any packet arriving at the HA to the MN at its auto-configured CoA.

3. In Mobile IPv4, it is not always possible for an MN to send packets directly to CN without being routed by the HA due the router's effect of ingress filtering [22]. In solving this, reverse tunnelling has to be used in Mobile IPv4 [23], with the sacrifice of routing efficiency optimisation. In Mobile IPv6, since the CN can cache MN's BU, the problem of reverse tunnelling is solved without affecting the operation of ingress filtering.

4. Since it is supposed that all IPv6 nodes are to implement strong authentication and encryption features in order to improve Internet security [24 25], the Mobile IPv6 protocol is then simplified so that it need not specify the security procedures by itself.

4.4 Mobile IP disadvantages

Mobile IP has been widely recognized as the suitable global solution to macro mobility by all the proposals. The IETF is currently on the way to considering Mobile IP as a proposed standard. However, there are some essential disadvantages in Mobile IP solution that prevent it from being the unique solution to both of the macro and micro mobility scenarios. These include the bad satisfaction of some of the requirements on performance and scalability listed in Section 2.2, being analysed in detail below.

1. Fast handoff. Due to the long message delay, it will take a BU a very long way to the HA to announce the new CoA, e.g. several seconds when the MN is on the other side of the world.

2. Seamless handoff. The packet loss rate will be quite high due to the long location registration process during handoff.

3. Signalling traffic overhead. This will be too high, especially to frequently moving MNs, since the new CoA of the MN is registered all the way with the HA whenever the MN has moved into a new area. In the optimised version of Mobile IP and Mobile IPv6, the location information (BU) is also transferred to the CNs.

4. Quality of Service (QoS). Acquiring a new CoA on every handoff would trigger the establishment of new QoS reservations between the HA and the FA, although most of the path from MN to a CN would be the same before and after the handoff.

These are the main reasons why possible solutions for micro mobility are needed as the necessary complement to the global solution of Mobile IP. The basic ideas behind the micro mobility management is that the HA need not be aware of every handoff that the MN performs. Instead, the foreign networks in a limited range can take care of the MN's small movements locally. Micro mobility solution proposals are discussed in the next section.

5. MICRO MOBILITY SOLUTIONS: THREE PROPOSED PROTOCOLS

Micro mobility solutions are presented for the intra-domain mobility management to implement a fast and seamless handoff and a minimized control traffic overhead. As shown in Figure 3, movement within a foreign network domain need not inform the MN's HA of the new attachment. The micro mobility protocols ensure that the packets arriving at the mobility server (gateway) can be correctly forwarded to the appropriate access point that the MN currently attaches.

Three main proposals are discussed in this section, i.e. HMIP, Cellular IP, and HAWAII. Table 1 shows a simple comparison of the three proposals. Note that none of these suggestions are trying to replace the Mobile IP. Instead they are enhancements to the basic Mobile IP with the micro mobility management capability. There are still many other micro mobility proposals existing, e.g. the Intra-Domain Mobility Management Protocol (IDMP) [26], Edge Mobility Architecture [27], Hierarchical Mobile IPv6 (HMIPv6) [28], etc.

Table 1: Simple Comparison of Cellular IP, Hawaii and Hierarchical Mobile IP [29]

	Hierarchical MIP	Cellular IP	HAWAII
OSI Layer	"L3.5"	L3	L3
Nodes Involved	FAs	all CIP nodes	all routers
Mobile Host ID	home addr	home addr	c/o addr
Intermediate Nodes	L3 routers	L2 switches	L2 switches
Means of Update	signalling msg	data pkt	signalling msg
Paging	explicit	implicit	explicit
Tunnelling	yes	no	no
L2 Triggered Handoff	no	optional	optional
MIP Messaging	yes	no	yes

5.1 Hierarchical Mobile IP

The same basic idea of hierarchical structure of visited networks is employed by several proposed protocols. In all these protocol suggestions, the MN's HA needs not to be informed of every movement that the MN performs inside the foreign network domain. In this section, we introduce the proposal from Ericsson and Nokia that employs a hierarchy of FAs to handle the MN's local registration [8].

Consider the illustration in Figure 4, where the FAs in a domain are organized into a hierarchical tree-like structure. The root of the hierarchy (FA_1) is a special kind of foreign agent called Gateway Foreign Agent (GFA). An FA's agent advertisement is extended to include in the CoA field the IP addresses of FAs from the FA itself through all the ancestor FAs until the GFA (in the figure FA_4, FA_3, FA_1). The MN's registration is then processed by all the FAs (updating the maintained visitor list entry) on the uplink path ended by the GFA and finally the HA stores the GFA's IP address as the current CoA of the MN. Through this mechanism, the location information is managed in a distributed mode.

When an incoming packet (from the HA or any CN) arrives at the GFA, the downlink path is formed by searching the visitor list for a corresponding entry by each FA on the path. Each of the FA then re-tunnels (decapsulating and reencapsulating) the packet to its next lower-level FA, and the packet is forwarded down the tree of FAs toward the MN's point of attachment.

As a handoff occurs, the MN first finds the Closest Common Ancestor (CCA) by comparing the new CoA vector with the old one and choosing the lowest-level FA that appears in both CoA vectors. The MN then regionally registers this movement to the CCA and leaves all the higher-level FAs unaffected. In Figure 4, when the MN moves from FA_4 to FA_5, the FA_2 is the CCA. When moving from FA_5 to FA_6, the FA_1 is the CCA.

Typically only two levels of the hierarchy are considered: at the top level there is one or several GFAs, and at the lower level all FAs are connected to the corresponding GFA. If there is a hierarchy of FAs between the GFA and the MN's current FA, the FA must support the smooth handover routing optimisation [14] described in Section 4.2. In [30] a paging extension for HMIP is presented, allowing idle mobile nodes to operate in a power saving mode while located within a paging area.

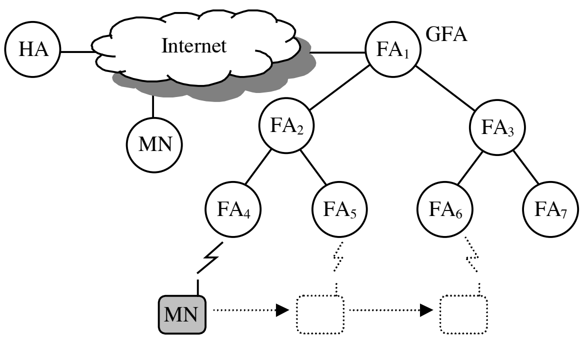


Figure 4: Hierarchical FAs.

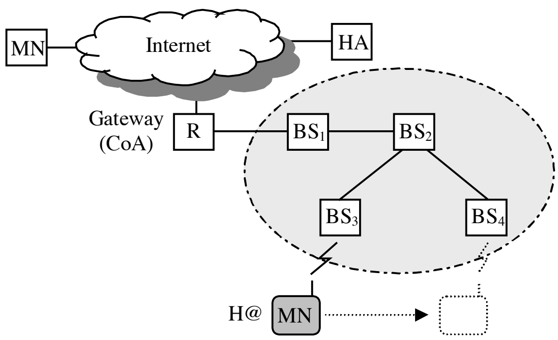


Figure 5: Cellular IP access network.

5.2 Cellular IP

Columbia University and Ericsson propose the Cellular IP [9] for very frequently moving hosts as well as rarely moving and totally static hosts. The Cellular IP combines the capability of cellular networks in providing smooth fast handoff and efficient location management for active and idle mobile hosts with the inherent flexibility, robustness, and scalability found in IP networks. Location management and handoff support are integrated with routing in Cellular IP access networks. The Cellular IP is intended for use in local or metropolitan area networks. It is an extension to basic Mobile IP protocol instead of a replacement for it.

Figure 5 illustrates the basic structure of Internet containing Cellular IP access networks – networks implementing the cellular IP protocol. The universal component of a Cellular IP network is the Base Station (BS) serving as an access point. The boundary between a Cellular IP access network and the Internet is the Gateway router: mobility between Gateways is macro, while within the access network it is micro and handled by Cellular IP. The IP address of the

Gateway is the CoA of all the MNs attached to the access network. Packets destined to the MN are first routed to the Gateway. Within the Cellular IP access networks, the MN is identified by its H@ and packets are forwarded to the MN through BSs hop-by-hop instead of using tunnelling or address conversion.

The Gateway periodically broadcasts a beacon packet, depending on which each BS can form the uplink to the Gateway and then route the uplink packets from the MN to the Gateway hop-by-hop accordingly. Each BS is also responsible for maintaining a routing cache. An entry in the routing cache binds the MN's H@ with the interface through which the MN can be arrived at. The routing information of the entry is generated and refreshed by monitoring regular packets sent by an MN in order to minimize control messaging. The downlink path can then be formed by reversing the path. Since the bindings in the routing caches have timeout values, an MN can keep the entry valid by sending empty packets (route-update packets) to the Gateway at regular intervals.

Cellular IP supports two types of handoff scheme: hard handoff and semisoft handoff. Hard handoff is based on a simplistic approach that supports fast and simple handoff by minimizing handoff signalling with the potential packet loss. Handoff is initiated by the MN on the basis of signal strength measurements, and performed by sending a route-update packet. Semisoft handoff takes advantage of the feature that there is a period in which both the old and the new downlink routes are valid and packets are delivered through both BSs. The MN may first initiate a handoff by sending a semisoft packet to the new BS, and then perform the regular handoff after a delay during which the new downlink path has been configured. Through this mechanism, the handoff performance is improved by providing a probabilistic guarantee instead of fully eliminating packet loss.

An explicit paging scheme is employed by Cellular IP to reduce power consumption at the terminal end and minimize the signalling overhead in the access network. An idle state is introduced for the MN that has not received packets for a specified interval but still wants to remain reachable. Cellular IP tracks the location of idle hosts only approximately. A paging cache may be maintained by some BSs and it has a longer timeout period than routing cache. As soon as incoming packets need to be sent to an idle MN, the MN can be paged through a limited scope broadcast if the entry in the routing cache has expired.

5.3 HAWAII

The protocol of Handoff-Aware Wireless Access Internet Infrastructure (HAWAII) [10] is proposed by Lucent Technologies Bell Labs as a separate routing protocol to take care of the micro mobility inside the visited domain. Still, HAWAII relies on Mobile IP to provide wide-area inter-domain macro mobility management. HAWAII is now transparent to MNs that are compatible with Mobile IP with route optimisation, challenge/response, and Network Access Identifier (NAI) extensions. The main goals of HAWAII include achieving good performance, providing intrinsic support for QoS, and enhancing reliability.

The basic network architecture is illustrated in Figure 6. The gateway in each domain is called the Domain Root Router (DRR). No HA is involved when an MN's movement is within the home domain, where the MN is identified by its IP address. When an MN is moving within a foreign domain, the MN is assigned a CCoA. Packets can then be tunnelled to the MN by the HA in its home domain. This CCoA remains unchanged as long as the MN is moving within the foreign domain, thus the HA needs not be notified of these movements unless the mobile host moves to a new domain. This is achieved by enabling any mobile host to register with a BS while using a CCOA and then locally handling the registration by the corresponding BS.

The processing and generation of the Mobile IP registration messages are split into two parts: between the mobile host and the base station, and between the base station and the home agent. Nodes in a HAWAII network execute a generic IP routing protocol and maintain mobility-specific routing information. Location information (i.e. mobile-specific routing entries) is created, updated, and modified by explicit signalling messages sent by MNs.

HAWAII defines different path setup schemes to update the routers in a domain so that the connectivity to the MN is maintained across handoffs. Two scenarios are considered by HAWAII: power up and following handoff. Then, two variants of the path setup schemes are described for the operations of handoff, motivated by two types of wireless networks. The Forwarding scheme is optimised for networks where the MN is able to listen/transmit to only one BS (e.g. TDMA network). The Non-Forwarding scheme is optimised for networks where the MN is able to listen/transmit to two or more BSs simultaneously (e.g. CDMA network).

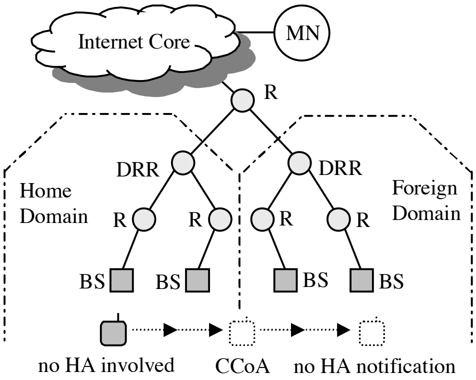


Figure 6: HAWAII network architecture.

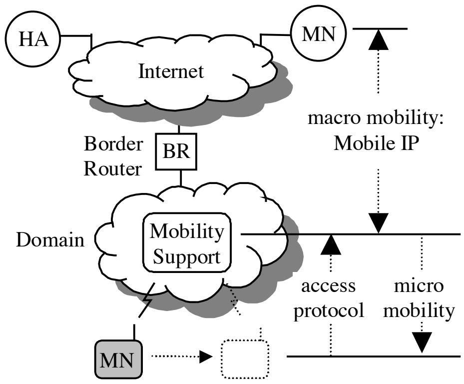


Figure 7: Unified mobility framework.

5.4 Unified hierarchical mobility

It is widely agreed that Mobile IP is suitable to handle the macro mobility between networks, whereas the micro mobility solutions described in the previous sections, together with all the other proposals, define various micro mobility support protocols to be used inside certain subnetworks. It is difficult to find such a micro mobility solution that can be optimal for any kind of network. This situation leads to a need to make it possible that different micro mobility protocols can coexist in the Internet so that the CN and the HA would not need to be aware of the difference.

The unified hierarchical mobility model proposed by INRIA [31] presents a framework for interoperability between different types of micro-mobility protocols (e.g., Cellular IP and HAWAII), through defining in a model how different micro mobility solutions can coexist simultaneously in the Internet. In the framework, the mobility management is composed of three protocol components [31]:

1) The access mobility management protocol, specifies the registration procedures between the MN and the domain that it is attached to. It is standard and independent of the micro and macro mobility management protocols used in the core of the network.
2) The micro mobility management protocol, handles the MN's local mobility within the domain and can vary from one domain to another depending on which protocol is supported (e.g., HAWAII, HMP, Cellular IP, etc).
3) The macro mobility management protocol, takes care of the global mobility management between domains. Mobile IP is proposed to be used probably in all cases.

The mobility support is then defined as a router or a set of routers that maintains a binding of MNs currently visiting the domain. The operations of mobility support include general actions, e.g. sending BUs for and redirecting packets to the MN. Mobile node registration is independent of the micro mobility protocol operating within a specific domain. The nature of the mobility support is therefore very much dependent on which micro mobility protocols are deployed. Figure 7 shows the proposed framework.

6. CONCLUSIONS

This paper discusses the mobility management schemes for the next generation mobile networks. A global hierarchical framework model for the mobility management of wireless networks is presented, in which the mobility management is divided into two complementary tasks: macro mobility and micro mobility. As the macro mobility solution, basic principle of Mobile IP is introduced, together with optimal schemes and the advances in IPv6. The disadvantages of the Mobile IP on solving the micro mobility problem are analysed, based on which three main proposals are discussed as the micro mobility solutions, including HMIP, Cellular IP, and HAWAII. A unified model is also described in which the different micro mobility solutions can coexist simultaneously in mobile networks. The presented framework and content discussed in the paper can serve as an effective guide to the overall solution and systematic research on the problem of mobility management for the next generation wireless communications. Further research topics focus mainly on the evaluation of the proposals in a testbed with actual traffic load and a careful consideration of security.

ACKNOWLEDGMENT

Financial support by the National Technology Agency of Finland is gratefully acknowledged.

REFERENCES

1. IMT-2000, http://www.itu.int/imt/.
2. IETF Mobile IP Working Group (mobileip), http://www.ietf.org/html.charters/mobileip-charter.html.
3. 3rd Generation Partnership Project (3GPP), http://www.3gpp.ort.
4. 3rd Generation Partnership Project 2 (3GPP2), http://www.3gpp2.ort.
5. Mobile Wireless Internet Forum (MWIF), http://www.mwif.org/.
6. 3GPP TR 23.922, "Architecture for an All IP Network," Dec. 1999.
7. C. Perkins, ed., "IP mobility support," RFC 2002, Oct. 1996.
8. E. Gustafsson, A. Jonsson, and C. Perkins, "Mobile IP regional registration," Internet Draft, draft-ietf-mobileip-reg-tunnel-04, Work in Progress, Mar. 2001.
9. A. Campbell, J. Gomez, C-Y. Wan, S. Kim, Z. Turanyi, and A. Valko, "Cellular IP," Internet Draft, draft-ietf-mobileip-cellularip-00, Work in Progress, Dec. 1999.
10. R. Ramjee, T. La Porta, S. Thuel, K. Varadhan, and L. Salgarelli, "IP micro-mobility support using HAWAII," Internet Draft, draft-ietf-mobileip-hawaii-00, Work in Progress, June 1999.
11. I. Akyildiz, J. Ho, and W. Wang, "Mobility management in next-generation wireless systems," *Proceedings of the IEEE,* Vol. 87, No. 8, pp. 1347-1384, Aug. 1999,
12. P. Bhagwat, C. Perkins, and S. Tripathi. "Network layer mobility: an architecture and survey," *IEEE Personal Communications,* Vol. 3, No. 3, pp. 54-64, June 1996.
13. S. Deering, ed., "ICMP router discovery messages," RFC 1256, Sep. 1991.
14. C. Perkins and D. Johnson, "Route optimization in Mobile IP," Internet Draft, draft-ietf-mobileip-optim-10, Work in Progress, Nov. 2000.
15. IETF IP Next Generation Working Group (ipngwg), http://www.ietf.org/html.charters/ipngwg-charter.html.
16. S. Deering and R. Hinden, "Internet Protocol, version 6 (IPv6) specification," RFC 2460, Dec. 1998.
17. R. Hinden and S. Deering "IP version 6 addressing architecture," RFC 2373, July 1998.
18. T. Narten, E. Nordmark, and W. Simpson, "Neighbor discovery for IP version 6 (IPv6)," RFC 2461, Dec. 1998.
19. S. Thomson and T. Narten, "IPv6 stateless address autoconfiguration," RFC2462, Dec. 1998.
20. D. Johnson and C. Perkins, "Mobility support in IPv6," Internet Draft, draft-ietf-mobileip-ipv6-14, Work in Progress, July 2000.
21. A. Conta and S. Deering, "Generic packet tunneling in IPv6 specification," RFC 2473, Dec. 1998.
22. P. Ferguson and D. Senie, "Network ingress filtering: defeating denial of service attacks which employ IP source address spoofing," RFC 2267, Jan. 1998.
23. G. Montenegro, ed., "Reverse tunneling for Mobile IP, revised" RFC 3024, Jan. 2001.
24. S. Kent and R. Atkinson, "IP authentication header," RFC 2402, Nov. 1998.
25. S. Kent and R. Atkinson, "IP encapsulating security payload (ESP)," RFC 2406, Work in Progress, Nov. 1998.
26. A. Misra, S. Das, A. Mcauley, A. Dutta, and S. K. Das, "IDMP: an intra-domain mobility management protocol using mobility agents," Internet Draft, draft-misra-mobileip-idmp-00, Work in Progress, Jan. 2000.
27. A. O'Neill, G. Tsirtsis, and S. Corson, "Edge mobility architecture," Internet Draft, draft-oneill-ema-02, Work in Progress, July 2000.
28. H. Soliman, C. Castelluccia, and L. Bellier, "Hierarchical Mobile IPv6 (HMIPv6)," Internet Draft, draft-ietf-mobileip-hmipv6-04, Work in Progress, July 2001.
29. A.T. Campbell, J.Castellanos, "IP micro-mobility protocols," *ACM Mobile Computing and Communications Review,* Vol. 4, No. 4, pp. 45-53, Oct. 2000.
30. H. Haverinen and J. Malinen, "Mobile IP regional paging," Internet Draft, draft-haverinen-mobileip-reg-paging-00, Work in Progress, June 2000.
31. C. Castelluccia and L. Bellier, "Toward a unified hierarchical mobility management framework," Internet Draft, draft-castelluccia-uhmm-framework-00, Work in Progress, June 1999.