Wireless Netw (2007) 13:203–220
DOI 10.1007/s11276-006-6464-y

Impact of the access network topology on the handoff performance

Liesbeth Peters · Ingrid Moerman · Bart Dhoedt · Piet Demeester

Published online: 8 May 2006

Abstract Micromobility protocols such as Cellular IP, Hawaii and Hierarchical Mobile IP are developed to solve problems of high handoff latency and control overhead, which occur when Mobile IP is used in combination with frequent handoffs. Up to now, tree access network topologies are considered to evaluate the protocol performance. However, for reasons of robustness against link failures and load balancing, extra uplinks and mesh links in the topology are desired. This article makes a classification of several topology types and gives a model that points out to which extent the topology influences the protocol performance in terms of handoff latency and handoff packet loss. Simulations confirm the results calculated by the model. Performance metrics such as load balancing, end-to-end delay and robustness against link failures are also evaluated. The study points to several shortcomings of the existing micromobility protocols for different topology types. Several aspects of the studied handoff schemes, their advantages and drawbacks are identified.

Keywords IP mobility management · Micromobility · Access network topology · Protocol performance

L. Peters is a Research Assistant of the Fund for Scientific Research – Flanders (F.W.O.-V., Belgium)

L. Peters ( ) · I. Moerman · B. Dhoedt · P. Demeester
Department of Information Technology (INTEC), Ghent University – IBBT – IMEC, Gaston Crommenlaan 8 bus 201, B-9050 Gent, Belgium
e-mail: Liesbeth.Peters@intec.UGent.be

I. Moerman
e-mail: Ingrid.Moerman@intec.UGent.be

B. Dhoedt
e-mail: Bart.Dhoedt@intec.UGent.be

P. Demeester
e-mail: Piet.Demeester@intec.UGent.be

1. Introduction

For several years, the increasing popularity of the Internet and multimedia applications encourages people to use not only their mobile phones, but also their PDA's and laptops while moving from one place to another. The success of applications like e-mail, ftp, browsing the internet, video conferencing and network gaming will result in huge amounts of packet-based data traffic, exceeding the share of circuit-based voice traffic for which cellular telecommunication networks were originally designed. Although current networks (GSM, GPRS, UMTS [1, 18]) also support data traffic, the data rates at vehicular speed are still limited, and, besides, the 3G-networks are very complex and cost ineffective [7, 10].

Therefore, wireless networks evolve towards IP-based infrastructures to allow a seamless integration between wired and wireless technologies. But in contrast to wired networks, the user's point of attachment to the network changes frequently due to mobility. Since an IP address indicates the location of the user in the network as well as the end point of its connections, user mobility leads to several challenges. During the last years, much research is done in this area and several routing protocols are developed to support IP mobility. Mobile IP (IPv4 [12], IPv6 [9]), which is standardized by the IETF, is the best known routing protocol that supports host mobility. Every time a mobile host moves within the area covered by another access router, it receives, in addition to its fixed home IP address, a second IP address (e.g. by DHCP). This variable second address is called care-of address and gives information about the current point of attachment of the mobile host. The mobile host must register

this care-of address with its home agent in its home domain. This allows the home agent to map the home address to the corresponding care-of address. The home agent tunnels the data packets for the mobile host towards the registered care-of address. Arriving at the care-of address, the endpoint of the tunnel, the data packets are delivered to the mobile host. Frequent registering clearly results in control overhead and a considerable handoff latency.

To solve the weaknesses of Mobile IP, several protocols like Cellular IP [19], Hawaii [15] and Hierarchical Mobile IP [8] are proposed to support the movements within one IP domain. This kind of local mobility is called micromobility. As long as the mobile terminal resides in the same domain, the same care-of address can be used and other mechanisms realize the change of access router. A micromobility protocol restricts the control traffic, needed to update the necessary routing tables after handoff, to this IP domain. However, Mobile IP is still used to support macromobility, i.e. the movements from one IP domain to another. Although all these micromobility protocols are designed to work correctly irrespective of the topology of the IP domains, this topology has an important influence on the performance of those routing protocols, which is studied in this paper.

Existing studies of these micromobility protocols mainly contain detailed descriptions of the protocol mechanisms, classifications of these protocols and generic micromobility models [3, 17]. However, for the development of micromobility protocols, the access network is generally assumed to have a pure tree topology, rooted at a gateway and with branches towards the access routers. Therefore, existing simulation studies are limited to tree topologies [5]. The use of a pure tree topology and a single gateway results in an access network that is very vulnerable to link failures and to the risk that the gateway forms a bottleneck. As indicated in [13] and [14] micromobility protocols can have completely different performance results for more meshed topology types.

This article makes a classification of several topology types and presents a model to evaluate the influence of the topology on the handoff latency and handoff packet loss of the above mentioned micromobility protocols. Simulation results confirm the validity of the proposed formulae. Load balancing, end-to-end delay and robustness against link failures are also investigated through simulations. The study allows to point to the shortcomings of existing micromobility protocols when used in access networks that have not a pure tree topology. An overview of several mechanisms of the studied micromobility protocols, their advantages and drawbacks is presented.

The rest of this article is structured as follows. In Section 2, we shortly describe the topology of an all IP-based cellular network and we make a classification of possible access network topologies. Section 3 analyses the layer 3 handoff process in general and the location of the cross-over node during handoff for the studied micromobility protocols. A model is presented in Section 4, which expresses the influence of the topology of the access network on the performance of the micromobility protocols. Finally, simulation results are presented and discussed in Section 5. The simulation results are also compared with the values calculated by the model. Section 6 contains our concluding remarks.

2. Topology of the network

2.1. All IP-based networks

Wireless networks evolve towards all IP-based infrastructures. Most routing protocols that are developed to support IP mobility assume that the network consists of an IP-based core network and several IP domains. The connection between such an IP domain and the core network is performed by a special router, called the domain gateway. Every mobile host has one home domain, the IP domain where the mobile host normally resides. When a mobile host moves to a domain that is not its home domain, that domain is called its current foreign domain. In order to receive and to send data, a mobile host has to make a connection to the network via a router with a wireless interface, an access router. Figure 1 gives an illustration of this general network topology.

2.2. Access network topologies

The access network topology, i.e. the topology of an IP domain, has an influence on the performance of the micromobility protocols. In order to evaluate this influence, it is necessary to make a classification of possible topologies. To this end, every node of the access network is characterized by a number d, indicating the minimum number of hops needed to reach the domain gateway. Thus, for the domain gateway

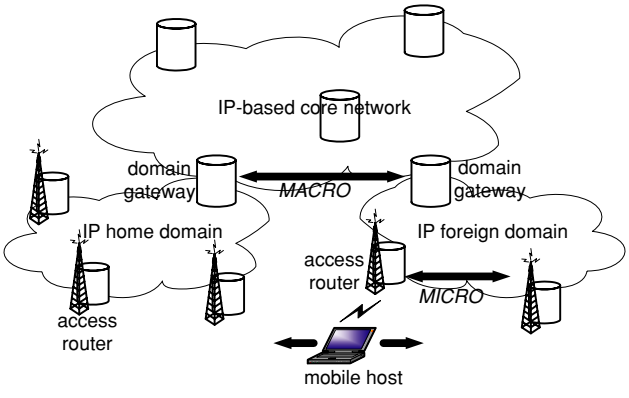

Fig. 1 General topology of an IP-based network. An IP domain is connected to the core network via a domain gateway. Micromobility is the term used to indicate movements within the same IP domain, while macromobility points to a change of IP domain

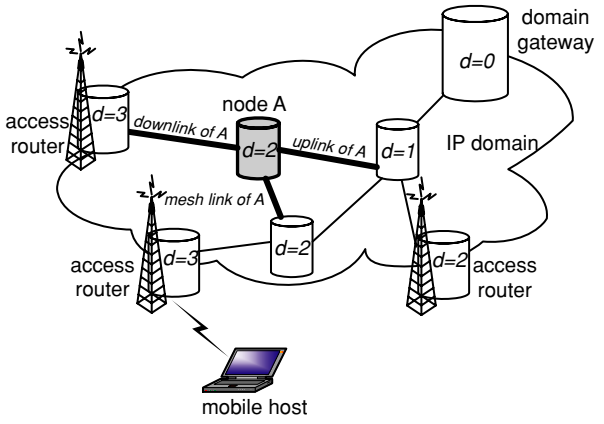

Fig. 2 Illustration of the different link types of node A in an IP domain: downlink, uplink and mesh link

it holds that $d = 0$, while the nodes with a direct link to the domain gateway have $d = 1$, etc. From the viewpoint of a node, several link types can be distinguished. An *uplink* of a node with $d = k$ is defined as a link from this node to another one with $d = (k - 1)$. A *downlink* of a node with $d = k$ ends in a node with $d = (k + 1)$. Finally, links between nodes with the same number d are called *mesh links* of these nodes. The link itself has then a mesh-level d. An example is given in Fig. 2.

These definitions allow the following classification of possible topologies:

- **Tree:** The tree topology is considered as the basic type, because most micromobility protocols implicitly assume that the nodes and links of an IP domain form a tree topology, connected to the core network by a single gateway. Every node of the tree has exactly one uplink and possibly some downlinks. This topology has no mesh links.
- **Mesh:** A mesh topology is defined as a pure tree topology with additional mesh links.
- **Random:** The term random topology is used to indicate a mesh topology with additional uplinks. This means that at least one node of the topology has more than one uplink, while the other nodes have exactly one uplink.

Table 1 gives a summary of the different topology structures and the corresponding per node values for the different link types.

Table 1 Classification of access network topologies. For every topology type, each node in the access network has a number of uplinks, downlinks and mesh links as indicated in the table

Topology	Uplinks	Downlinks	Mesh links
tree	1	$\geqslant 0$	0
mesh	1	$\geqslant 0$	$\geqslant 0$
random	$\geqslant 1$	$\geqslant 0$	$\geqslant 0$

3. Support of micromobility

3.1. Layer 3 handoff process

Every access router sends out beacon messages at fixed *time intervals* T_b in order to announce its presence to the mobile hosts located in its coverage area. Usually, the coverage areas, called cells, of neighbouring access routers overlap. When a mobile host enters the overlap region of two cells, it receives the beacons of both access routers and according to the signal strength of the beacons, the mobile terminal selects one of the base stations to exchange data traffic with the wired network. In what follows, we assume that, although the mobile host can receive the beacons of every access router in which cell it resides, the mobile terminal can exchange data traffic with only one base station at a time.

The different possible scenarios are depicted in Fig. 3. Hereby, three parameters are important for our study: T_n, T_h and T_w. T_n is called the *network time*, i.e. the time between the moment that the new access router (nAR) receives the handoff message (message b in the figure) and the moment that this nAR can (but not necessarily does) receive and forward the first data packet (packet c in the figure) to the mobile host (MH). Analogously, T_h is the *handoff time* and is defined as the time between the moment that the MH sends the handoff message (b) and the moment that this MH can receive its first data packet (c) via the nAR. T_h consists of T_n increased by the time to send the handoff message and the data packet over the wireless link. T_n strongly depends on the used micromobility protocol and the topology of the access network. This is studied and explained in Section 4. T_w, the *wait time*, is the minimum time the MH has to wait to receive the first data packet (c) via the nAR, once it stopped listening to the old access router (oAR). This parameter gives thus an indication of *the handoff latency* or *how fast* the handoff is. For the first situation it obeys $T_w > T_h$, for the second scenario it obeys $T_w = T_h$, while for the third and last situation $T_w < T_h$ applies.

The time that the MH resides in the overlap region is indicated with the overlap time T_o. At the point in time t_1, the MH enters the cell of the nAR and at t_2, it leaves the cell of the oAR.

In situation 1 of Fig. 3, the MH moves out of the range of the oAR before it detects a nAR and thus before any handoff process is started. This can happen for several reasons: when there is no overlap at all between the neighbouring cells or when T_o is very small and the MH has not yet received a beacon from the nAR before leaving the cell of the oAR. The time between the moment of connection loss with the oAR and the moment of the receipt of a beacon (a) is indicated as T_u, the *unconnected time*. The value T_u is independent of the used micromobility protocol, but is determined by the time T_o (depending on the overlap size and the speed of the MH) and

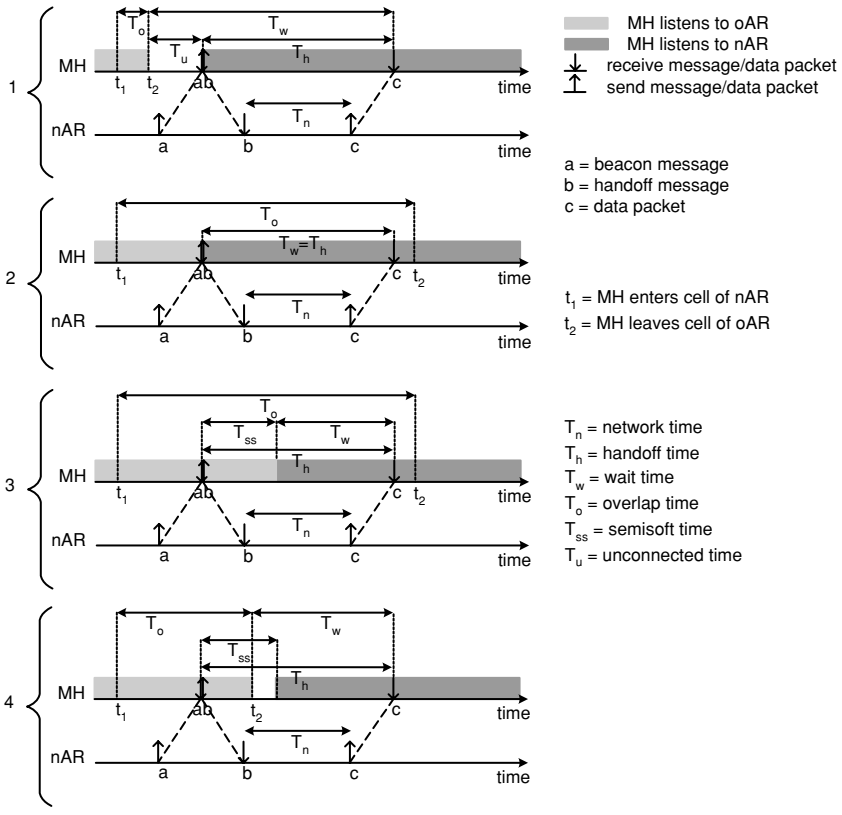

Fig. 3 Scenarios for the Layer 3 handoff process. 1)The MH moves out of the range of the oAR before it receives a beacon from the nAR. 2)The MH detects a nAR and switches listening to this nAR (hard handoff). 3)The MH detects a nAR, starts a handoff but continues listening to the oAR for a certain time (semisoft handoff). 4)During a semisoft handoff, the MH moves out of the range of the oAR

the time interval between two successive beacons. As soon as the MH receives a beacon, it sends a handoff message (*b*) to the nAR. When the nAR receives this message, the used micromobility protocol takes care of the necessary changes in the access network.

A more common case is situation 2: the MH resides long enough in the overlap region to receive a beacon (*a*) from the nAR while it is still connected to the oAR. The MH decides to perform handoff and sends a handoff message (*b*) to the nAR.

In case of situation 3 depicted in Fig. 3, the MH sends a handoff message (*b*) to the nAR but immediately restarts listening to the oAR for a time T_{ss}. After this time the MH switches back to the nAR. This type of handoff is called semisoft handoff and T_{ss} indicates the *semisoft time* or *handoff delay*. This is different from the two previous situations where hard handoff is used: the MH sends a handoff message to the nAR and continues to listen to this new base station.

The last situation 4 is a variation of the previous one: during T_{ss}, the MH leaves the cell of the oAR and can not receive any further packets via this access router. This will not affect the network time T_n, but the wait time T_w will increase.

3.2. Cross-over node in the access network

The description of the handoff process in the previous section is independent of the protocol used to support local IP mobility. However, Cellular IP, Hawaii and Hierarchical Mobile IP update the routing tables in the access network in their own way, resulting in possibly different values for T_n, T_h and T_w. For a detailed description of the micromobility protocols under study, we refer to [4, 8, 16, 20].

During the handoff process, one router in the access network has an important characteristic: as soon as the handoff message updates this router, new data packets that arrive in this router afterwards are correctly delivered to the MH via the nAR. This router is indicated by the term *cross-over node*. The several protocol mechanisms can result in a different cross-over node. This is further explained below and illustrated in Fig. 4. The *cross-over distance* is defined as the minimum number of hops between the new access router and the cross-over node and influences *the packet loss* during handoff or the *handoff smoothness*. Two time parameters are important for the analysis of the packet loss during handoff: $T_{CN\rightarrow oAR}$ and $T_{nAR\rightarrow CN}$. The first one is the time a data packet needs to travel from the cross-over node to the oAR and the latter is the time a handoff message needs to reach the cross-over node from the nAR.

- **Cellular IP:** A route update message is directed from the new access router to the domain gateway, updating the route cache mappings of every router on the path. The cross-over node is found as the node closest to the new access router, that is situated on both the path between the domain gateway and the old access router and the path

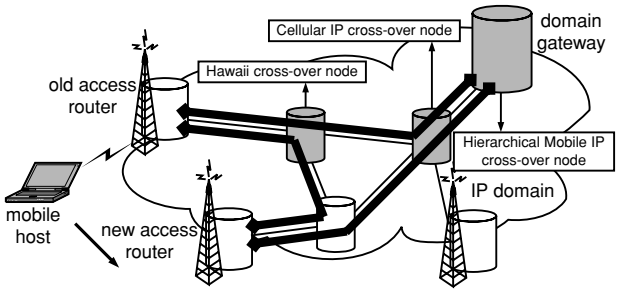

Fig. 4 Location of the cross-over node in the access network. The Cellular IP cross-over node is situated at the intersection of the path between the GW and the oAR and the path between the GW and the nAR. The Hawaii cross-over node is located on both the path between the GW and the oAR and the shortest path between both ARs. For Hierarchical Mobile IP, the GW can be considered as the cross-over node

between the domain gateway and the new access router. As soon as the route update message reaches the cross-over node, the host routes in the access network are updated to take into account the new access point of the MH.

- **Hawaii:** In this case, path setup messages are exchanged between the two access routers and the routing tables of the intermediate routers are updated. The node that is referred to as the cross-over node is now defined as the node closest to the new access router that lies at the intersection of the path between the gateway and the old access router, and the shortest path between the new and old access router. Note that for the same topology, Cellular IP and Hawaii can have different cross-over nodes.
- **Hierarchical Mobile IP:** When moving within the same IP domain, the change of access router must be notified to the domain gateway. In this case, there is no real cross-over node for handoff: a regional registration request is sent via the new access router towards the gateway and must reach this gateway in order to realize an effective change of routes. Routing tables of intermediate nodes are not changed: data packets are tunneled by the domain gateway towards the current access router.

4. Influence of the topology on the performance

The value of T_u is independent of the used micromobility protocol and access network topology and only depends on the time that a mobile host resides in the overlap region, i.e. T_o, and the time between two successive beacons from the same access router, i.e. T_b. T_u is zero as soon as $T_o \geqslant T_b$. In what follows, we assume that T_u is zero and situation 1 of Fig. 3 does not occur. In addition, T_o is assumed to be large enough, so that also situation 4 does not happen. For, when the MH moves out of the range of the oAR during T_{ss}, it should start listening to the nAR immediately, so that the applied time T_{ss} is shorter than initially intended.

The MH is the receiver of data packets, sent by a fixed terminal in the core network at a rate r. The opposite scenario, in which the MH sends data packets towards the core network, depends much less on the access topology, because the data packets are routed towards the gateway via the shortest path, and is therefore not described. The bandwidth of the links of the access network is such that the network is free from congestion.

In order to make the formulae more understandable, the time needed to process the packets in the nodes is assumed to be fixed and included in the link delay, and all the wired links of the access network have the same delay of d_{link} seconds. The exchange of protocol messages and data packets between the MH and the access routers, happens via a wireless link. The delay of this wireless link is influenced by the MAC layer of the IEEE802.11 protocol family, namely the CSMA/CA and Virtual Carrier Sense (RTS/CTS/data/ACK) mechanism [2]. Roughly, we can divide this delay into four parts: d_{RTS} (the sender sends a RTS control packet), d_{CTS} (the receiver responds with a CTS control packet), d_{data} (the effective transmission of the handoff message or data packet) and the last part d_{ACK} (the time needed to send the ACK control message).

4.1. General

Study of the handoff mechanism in combination with the access network topology results in a model for the handoff latency and the handoff packet loss. While the following sections will focus on every protocol separately, we now give some general considerations.

4.1.1. Handoff latency

Concerning the *handoff latency* of situations 2 and 3 of Fig. 3, the following relations between T_h, T_n and T_w (see Section 3.1) are valid:

$$\text{hard handoff: } T_h = T_w = T_n + d_{handoff} + d_{packet} \quad (1)$$

$$\text{semisoft handoff: } T_h = T_w + T_{ss} = T_n + d_{handoff} + d_{packet} \quad (2)$$

Hereby $d_{handoff}$ and d_{packet} are the delays needed to send a handoff message and data packet over the wireless link and are independent of the used micromobility protocol and topology. So in the following sections, only T_n is used.

4.1.2. Handoff packet loss

For the calculation of the *packet loss* during handoff, the position of the cross-over node is very important (see Section 3.2). Before handoff, every $1/r$ seconds a data packet arrives in the oAR and passes through the RTS-CTS-data-ACK process in order to arrive in the MH. We consider the period of

$1/r$ seconds in which the MH performs handoff, starting at the moment that a data packet arrives in the oAR. This data packet is the last one that the oAR tries to send before handoff. This data packet needed a time $T_{\mathrm{CN}\to\mathrm{oAR}}$ to move from the cross-over node to the oAR. Depending on the moment that the MH receives a beacon from the nAR and decides to perform handoff, this data packet may or may not be successfully received by the MH. Therefore d_{b} is used to indicate the time that passes from the start of this period of $1/r$ seconds to the receipt of the beacon by the MH. Furthermore, $d_{\mathrm{MH}\to\mathrm{nAR}}$ refers to the sum of the first, second and third part of the RTS-CTS-data-ACK process and is the time needed to send a handoff message from the MH to the nAR. As soon as this message arrives, the access router can take the appropriate actions, so the sending of the ACK (the last part) must not be taken into account for the calculation of the time needed to perform a complete handoff. The handoff message needs an additional time of $T_{\mathrm{nAR}\to\mathrm{CN}}$ seconds to travel from the nAR to the cross-over node.

During a *hard handoff*, the MH sends a handoff message to the nAR and stops listening to the previous one. At the moment that the MH switches to the nAR, say moment t, data packets are situated on the path between the cross-over node and the MH. As the MH switches listening to the nAR, these packets never reach the MH and are lost. In addition, all the data packets that pass through the cross-over node before the appropriate entry in the routing table of this router is updated by the handoff message, are still routed towards the oAR and are lost as well. The data packets that get lost are the packets that pass the cross-over node during a total time $T_{\mathrm{loss}}(t)$. Again, consider the period of $1/r$ seconds in which the MH performs handoff, starting at the moment that a data packet arrives in the oAR. Then, t is an arbitrary epoch within this period, obeying $0 \leqslant t \leqslant 1/r$, and $T_{\mathrm{loss}}(t)$ is given by:

$$T_{\mathrm{loss}}(t) = T_{\mathrm{CN}\to\mathrm{oAR}} + t + d_{\mathrm{MH}\to\mathrm{nAR}} + T_{\mathrm{nAR}\to\mathrm{CN}} \tag{3}$$

Depending on the value of d_{b}, the packet loss is given by:

$$\text{packet loss} = \lfloor T_{\mathrm{loss}}(d_{\mathrm{b}}) \cdot r \rfloor + 1 \quad \text{for } 0 < d_{\mathrm{b}} < d_{\mathrm{RTS}} \tag{4}$$

$$\text{packet loss} = \lfloor T_{\mathrm{loss}}(d_{\mathrm{RTS}} + d_{\mathrm{CTS}} + d_{\mathrm{data}}) \cdot r \rfloor \quad \text{for } d_{\mathrm{RTS}} < d_{\mathrm{b}} < d_{\mathrm{RTS}} + d_{\mathrm{CTS}} + d_{\mathrm{data}} \tag{5}$$

$$\text{packet loss} = \lfloor T_{\mathrm{loss}}(d_{\mathrm{b}}) \cdot r \rfloor \quad \text{for } d_{\mathrm{RTS}} + d_{\mathrm{CTS}} + d_{\mathrm{data}} < d_{\mathrm{b}} < 1/r \tag{6}$$

If the MH receives a beacon from the nAR before the RTS message from the oAR (4), it switches immediately to the nAR and the last data packet is not successfully sent. After receipt of the RTS message, the MH waits until it receives the entire data packet before switching to the nAR (5). In the last case (6), the data packet is successfully received before the MH detects a beacon and handoff is performed immediately. Note that for $T_{\mathrm{loss}}(t) \cdot r \gg 1$ the difference between the formulae can be neglected. When the optional RTS/CTS exchange is not used, e.g. to obtain a higher throughput, the number of possible situations is reduced to two: if the MH receives a beacon before the data packet sent by the oAR, it switches to the nAR and the data packet is not received. If the beacon is received after the data packet, the MH sends an ACK and switches then to the nAR.

As the mobile host moves independently of the sending of beacons by the access routers, d_{b} is uniformly distributed in $[0, 1/r[$. The expected value of the packet loss is given by:

$$\begin{aligned}\text{expected packet loss} &= \int_0^{d_{\mathrm{RTS}}} (\lfloor T_{\mathrm{loss}}(d_{\mathrm{b}}) \cdot r \rfloor + 1) r \,\mathrm{d}d_{\mathrm{b}} \\ &+ \int_{d_{\mathrm{RTS}}}^{d_{\mathrm{RTS}}+d_{\mathrm{CTS}}+d_{\mathrm{data}}} \lfloor T_{\mathrm{loss}}(d_{\mathrm{RTS}} + d_{\mathrm{CTS}} + d_{\mathrm{data}}) \cdot r \rfloor r \,\mathrm{d}d_{\mathrm{b}} \\ &+ \int_{d_{\mathrm{RTS}}+d_{\mathrm{CTS}}+d_{\mathrm{data}}}^{1/r} \lfloor T_{\mathrm{loss}}(d_{\mathrm{b}}) \cdot r \rfloor r \,\mathrm{d}d_{\mathrm{b}}\end{aligned} \tag{7}$$

However, some protocols have a mechanism that aims to reduce the packet loss during handoff. By *a semisoft handoff*, the MH restarts listening to the oAR for a certain time. Otherwise, *a buffer* in the oAR can be used to forward data packets from the old to the new access router. So, the expected value of the packet loss during a handoff in general is given by:

$$\text{packet loss(semisoft)} = \max\{\text{packet loss(hard)} - N, 0\} \tag{8}$$

Here, N is the expected value of the amount of data packets that are correctly received by the MH thanks to the use of a semisoft handoff or a buffer. In case of a hard handoff, N is zero.

Table 2 gives an overview of the used abbreviations in the formulae of the following sections.

4.2. Cellular IP

When the MH decides to perform a handoff after receiving a beacon from a nAR, it sends a *route update message* to the nAR. This route update is then forwarded towards the domain gateway (GW) in a hop-by-hop way, i.e. updating the route cache mappings in every router on the path. Although this route update finally reaches the GW, all the necessary changes in the routers of the access network are completed as soon as the route cache mappings of the

Table 2 Clarification of the used symbols in the formulae

Abbreviation	Significance
r	data rate (fixed number of data packets per second)
d_{link}	delay (seconds) of a wired link of the access network
d_{b}	delay (seconds) between arrival of data packet in oAR and receipt of beacon
$d_{\text{MH}\rightarrow\text{nAR}}$	delay (seconds) for RTS/CTS exchange and sending handoff message
$n_{\text{nAR}\rightarrow\text{oAR}}$	minimum number of hops on shortest path between new and old access router
$n_{\text{oAR}\rightarrow\text{CN}}$	minimum number of hops to reach cross-over node from old access router
$n_{\text{nAR}\rightarrow\text{CN}}$	minimum number of hops to reach cross-over node from new access router
$n_{\text{oAR}\rightarrow\text{GW}}$	minimum number of hops between domain gateway and old access router
$n_{\text{nAR}\rightarrow\text{GW}}$	minimum number of hops between domain gateway and new access router

cross-over node (CN) are updated. If $n_{\text{nAR}\rightarrow\text{CN}}$ indicates the number of hops between the nAR and the cross-over node, the route update will arrive in the cross-over node after a time $d_{\text{MH}\rightarrow\text{nAR}} + n_{\text{nAR}\rightarrow\text{CN}} \cdot d_{\text{link}}$. Two handoff schemes are possible: hard handoff and semisoft handoff.

- **Cellular IP hard handoff:** This is illustrated by situation 2 of Fig. 3. After sending a route update message (message *b*), the MH continues listening to the nAR. With $n_{\text{oAR}\rightarrow\text{CN}}$ the number of hops between the cross-over node and the oAR, $T_{\text{CN}\rightarrow\text{oAR}}$ is given by $n_{\text{oAR}\rightarrow\text{CN}} \cdot d_{\text{link}}$. As soon as the cross-over node is updated, the first data packet can be routed towards the nAR. A more detailed expression for $T_{\text{loss}}(t)$ and an expression for the network time T_{n} during a hard handoff are given by:

$$\begin{aligned} T_{\text{loss}}(t) &= n_{\text{oAR}\rightarrow\text{CN}} \cdot d_{\text{link}} + t + d_{\text{MH}\rightarrow\text{nAR}} \\ &\quad + n_{\text{nAR}\rightarrow\text{CN}} \cdot d_{\text{link}} \\ T_{\text{n}} &= 2 \cdot n_{\text{nAR}\rightarrow\text{CN}} \cdot d_{\text{link}} \end{aligned} \tag{9}$$

- **Cellular IP semisoft handoff:** In case of a semisoft handoff, as illustrated in situation 3 of Fig. 3, the MH sends a route update message to the nAR but immediately restarts listening to the oAR for an additional time T_{ss}, which is a protocol parameter. While the appropriate route cache mappings along the path between the nAR and the GW are created, the MH can still receive a number of data packets via the oAR during a time T_{ss}, so that the expected value N is given by $r \cdot T_{\text{ss}}$. As a result, when the MH finally switches listening to the nAR, the wait time T_{w} is much smaller compared to a hard handoff. We assume that during T_{ss} the MH stays within the range of the oAR and the route cache mappings do not timeout. The MH receives all the packets that are still sent to the oAR and not to the nAR as soon as T_{ss} equals T_{loss} in (9). When the MH waits longer before switching to the nAR, it receives duplicated packets (assuming that no packets get lost due to a buffer overflow in the nAR). The use of T_{ss} has no influence on the network time. The resulting expected value of the number of packets received during T_{ss} and the network time are now given by:

$$\begin{aligned} N &= r \cdot T_{\text{ss}} \\ T_{\text{n}} &= 2 \cdot n_{\text{nAR}\rightarrow\text{CN}} \cdot d_{\text{link}} \end{aligned} \tag{10}$$

4.3. Hawaii

In case of Hawaii, the MH is not aware of the use of Hawaii in the access network and uses Mobile IP, a macromobility protocol. To perform a handoff, the MH sends a *Mobile IP registration request* to the nAR. This access router sends a *path setup message* towards the oAR, which replies with another *path setup message*. The way these messages update the routing tables of the routers in the access network, depends on the used path setup scheme. The two studied path setup schemes are the Multiple Stream Forwarding (MSF) and the Unicast Non Forwarding (UNF) scheme. Finally, when the nAR receives the path setup message from the oAR, it sends a *Mobile IP registration reply* to report the successful handoff to the MH. For both handoff schemes, situation 2 of Fig. 3 is applicable (since $T_{\text{o}} \geqslant T_{\text{b}}$ is assumed).

- **Hawaii Multiple Stream Forwarding:** Using the MSF path setup scheme, the nAR sends the path setup message directly towards the oAR, without updating the routing table of any router on this path. When the oAR receives this message, its routing table is updated and it sends a path setup message back to the nAR, this time in a hop-by-hop way, updating every intermediate router.
  When we assume that the oAR has no buffer, all the data packets that pass the cross-over node before the first path setup message passes this node, are lost. After updating the oAR, data packets are forwarded from the oAR to the nAR and as soon as the second path setup message updates the routing table of the cross-over node, data packets are directly sent via the cross-over node to the nAR. Some data packets may arrive out of order in the

nAR. This mechanism results in the same expected value of the packet loss as for the hard handoff mechanism of Cellular IP.

A buffer in the oAR, characterized by a buffer size of B_{size} packets and a buffer time B_{time}, can be used to achieve a lower packet loss. When the path setup message arrives in the oAR, all the packets that are buffered within the last B_{time} seconds, i.e. $\min(r \cdot B_{\text{time}}, B_{\text{size}})$ packets are also forwarded towards the nAR. At the time of handoff, the buffer of the oAR may contain some data packets that are not yet sent to the MH. In addition, new data packets may arrive in the buffer before the update message. In order to forward all these data packets, B_{time} must be at least the sum of d_{b} and the time the Mobile IP registration request and the path setup message need to travel from the MH to the nAR and from the nAR towards the oAR. Older packets were sent successfully via the oAR and should not be forwarded to avoid duplicated packets. With $n_{\text{nAR}\to\text{oAR}}$ the number of hops between both access routers, this results in $B_{\text{time}} = d_{\text{b}} + d_{\text{MH}\to\text{nAR}} + n_{\text{nAR}\to\text{oAR}} \cdot d_{\text{link}}$.

$$\begin{aligned} \text{no buffer: } T_{\text{loss}}(t) &= n_{\text{oAR}\to\text{CN}} \cdot d_{\text{link}} + t + d_{\text{MH}\to\text{nAR}} \\ &\quad + n_{\text{nAR}\to\text{CN}} \cdot d_{\text{link}} \\ \text{buffer: } N &= \min(r \cdot B_{\text{time}}, B_{\text{size}}) \\ T_n &= 2 \cdot n_{\text{nAR}\to\text{oAR}} \cdot d_{\text{link}} \end{aligned} \tag{11}$$

- **Hawaii Unicast Non Forwarding:** The UNF scheme acts a little differently. The first path setup message is sent hop-by-hop from the nAR to the oAR. Every router on this path is updated and as soon as the routing table of the cross-over node is updated, packets are sent directly to the nAR. No packets are forwarded from the oAR to the nAR and no packets will arrive out of order. Although the mechanism is different, this does not change the expected value of the packet loss during handoff compared to the MSF mechanism without buffering. The network time on the other hand differs from the MSF scheme, as the first data packet that arrives in the nAR is routed via the cross-over node and not via the oAR. This mechanism is designed for networks where the MH is able to listen to both the old and new access router at the same time (e.g. a CDMA network). For such networks, which are not considered in this article, the UNF mechanism would have better results in terms of the packet loss.

$$\begin{aligned} T_{\text{loss}}(t) &= n_{\text{oAR}\to\text{CN}} \cdot d_{\text{link}} + t + d_{\text{MH}\to\text{nAR}} \\ &\quad + n_{\text{nAR}\to\text{CN}} \cdot d_{\text{link}} \\ T_{\text{n}} &= 2 \cdot n_{\text{nAR}\to\text{CN}} \cdot d_{\text{link}} \end{aligned} \tag{12}$$

4.4. Hierarchical Mobile IP

The proposed regional registration protocol (see [8]) supports one hierarchical level. The domain gateway fulfills the function of a gateway foreign agent (GFA) and one or more regional foreign agents (RFA) are located at the access routers. To perform a handoff using Hierarchical Mobile IP, the MH sends a *regional registration request* directly to the domain gateway (GW) via the nAR and starts listening to the nAR, like situation 2 of Fig. 3. The request is directly sent to the GW and only the routing table of the GW is updated. Then, the GW sends a *regional registration reply* to the MH to inform it about the successful handoff operation. In contrast to the previous protocols, the function of the cross-over node is fulfilled by the GW and $n_{\text{oAR}\to\text{GW}}$ and $n_{\text{nAR}\to\text{GW}}$, the number of hops between the oAR and the GW and between the nAR and the GW respectively, have an influence on the packet loss. Therefore, the formulae for Hierarchical Mobile IP (HMIP) are given by:

$$\begin{aligned} T_{\text{loss}}(t) &= n_{\text{oAR}\to\text{GW}} \cdot d_{\text{link}} + t + d_{\text{MH}\to\text{nAR}} \\ &\quad + n_{\text{nAR}\to\text{GW}} \cdot d_{\text{link}} \\ T_{\text{n}} &= 2 \cdot n_{\text{nAR}\to\text{GW}} \cdot d_{\text{link}} \end{aligned} \tag{13}$$

4.5. Remarks

From the formulae in the previous sections, it is clear that the topology of the access network has an important influence on the performance of the micromobility protocols. If we assume that every wired link of the access network has the same delay, the distance between two nodes in the topology is directly related to the hop count of the path between the nodes. For Cellular IP, the number of hops between an access router (old and new) and the cross-over node is crucial. Also the choice of an appropriate value of the T_{ss} parameter in case of semisoft handoff is determined by these distances. For Hawaii, the number of hops between the two access routers is important and influences the choice of a good buffer size for the MSF scheme. Moreover, Cellular IP and Hawaii can have a different location of the cross-over node for the same topology. In case of Hierarchical Mobile IP, the number of hops between an access router and the domain gateway becomes the most important factor. All three protocols react on the access network topology and what is more important, they react in a different way.

5. Simulations

The simulations in this section are performed with the network simulator ns-2 [11]. The Columbia IP Micromobility

Suite (CIMS) [6], which includes an implementation of the studied micromobility protocols for the ns-version 2.1b6, is used. The obtained simulation results are also compared with the formulae in Section 4.

The following parameter values are used in the simulations and calculations. For many parameters such as the link delay, link bandwidth, cell size, cell overlap, MH speed and traffic bit rate, typical values are chosen (such as in [5]):

- **Wired and wireless links:** The wired links of the access network have a delay of 2 ms and a capacity of 10 Mbit/s. The processing time of the packets in the routers is assumed to be included in the link delay, but a node also waits until it receives the entire packet before sending it to the next hop router. As a result, d_{link} is the sum of the fixed 2 ms and an additional delay depending on the packet size (different for data packet, update message, ...). We assume that the wireless link is idle at the moment an access router starts sending. As we will consider only one traffic flow and no network congestion, this assumption is fulfilled.
- **Access routers and mobile host:** The distance between two adjacent access routers is 200 m, with a cell overlap of 30 m. All the base stations are placed on a straight line. During the simulations, the mobile host travels from one access router to another at a speed of 20 m/s, maximizing the overlap time. As every access router broadcasts beacons at fixed time intervals T_{b} of 1.0 s and the mobile host resides in the overlap region during a time T_{o} of 1.5 s, the assumption of Section 4 that T_{u} is zero, is valid.
- **Traffic:** A CBR data traffic pattern is used, with a packet inter arrival time of 10 ms and a data packet size of 210 bytes. This results in a bitrate of 0.168 Mbit/s. Therefore, the routing of the traffic is not limited by the capacity of the wired and wireless links. One UDP connection is set up between the sender (a fixed host in the core network directly connected to the domain gateway) and the receiver (the mobile terminal).
- **Access network topology:** As explained in Section 2.2, the access network can have a tree, mesh or random topology. The topologies that are used for the simulations, are given in Fig. 5. The presented mesh topology consists of the tree structure (full lines) with the indicated additional mesh links (dashed lines), while the random topology is formed by adding extra uplinks (dotted lines) to the mesh topology.

The simulation results are average values of a set of more than 100 independent simulations, i.e. there is no correlation between the sending of beacons by the ARs, the movements of the MH and the arrival of data packets in the ARs. In order to avoid any correlation, the moments that the access routers start sending beacons, the mobile host starts moving and the data traffic starts, are randomly chosen in [0.0 s, 1.0 s[. As a result, these moments vary across the simulations.

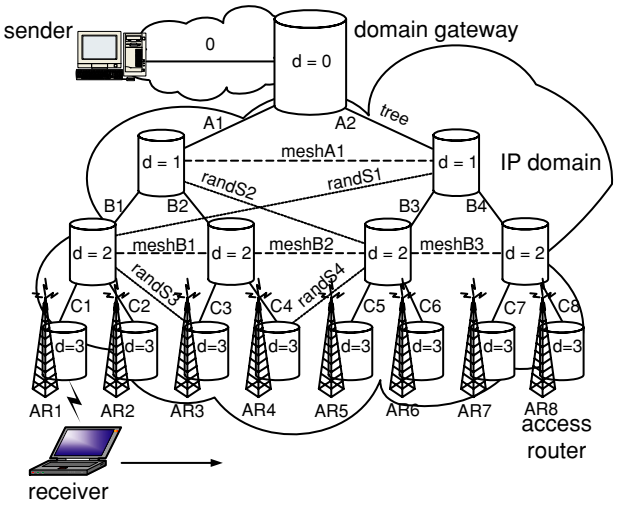

Fig. 5 Example of a tree (full lines), mesh (tree plus dashed lines) and random (mesh plus dotted lines) topology. Every link has a link name that will be used to present the simulation results for the load balancing in Section 5.3

5.1. Cross-over distances and meshes

When the mobile host moves from the leftmost to the rightmost access router in Fig. 5, it performs seven handoffs to maintain its connection with the sending terminal. Considering e.g. the first (AR1 → AR2) and the fourth (AR4 → AR5) handoff, the number d of the cross-over node and the cross-over distance are significantly different, depending on the used protocol and the used topology. Therefore this subsection investigates the handoff performance of a single handoff as a function of cross-over distance and mesh-level. The results are shown in Fig. 6. The diagrams illustrate the average packet losses during one handoff and a comparison with the calculated values can be found in Table 3. The values of the simulation parameters in ns-2 that are used for the calculations are given in Table 4. For the simulations of the Hawaii MSF handoff scheme, no buffer is used while for the semisoft handoff of Cellular IP, a value of 12 ms is chosen for the parameter T_{ss}. This value of T_{ss} is sufficient to eliminate every packet loss during handoff as long as the cross-over distance is small enough. For higher distances a small packet loss is observed.

In order to obtain diagram *A* of Fig. 6, a situation is considered where Cellular IP (CIP) and Hawaii (HAW) have the same cross-over node, namely the domain gateway. The cross-over distance is then the number of hops from the new access router to the domain gateway. An example of such a situation is the handoff from AR4 → AR5 in the tree topology of Fig. 5, where the cross-over distance equals 3. The update messages of CIP, HAW and Hierarchical Mobile IP (HMIP) have to reach or pass through the gateway to update the routing tables after a handoff, so the packet losses during handoff increase with the distance to the domain gateway.

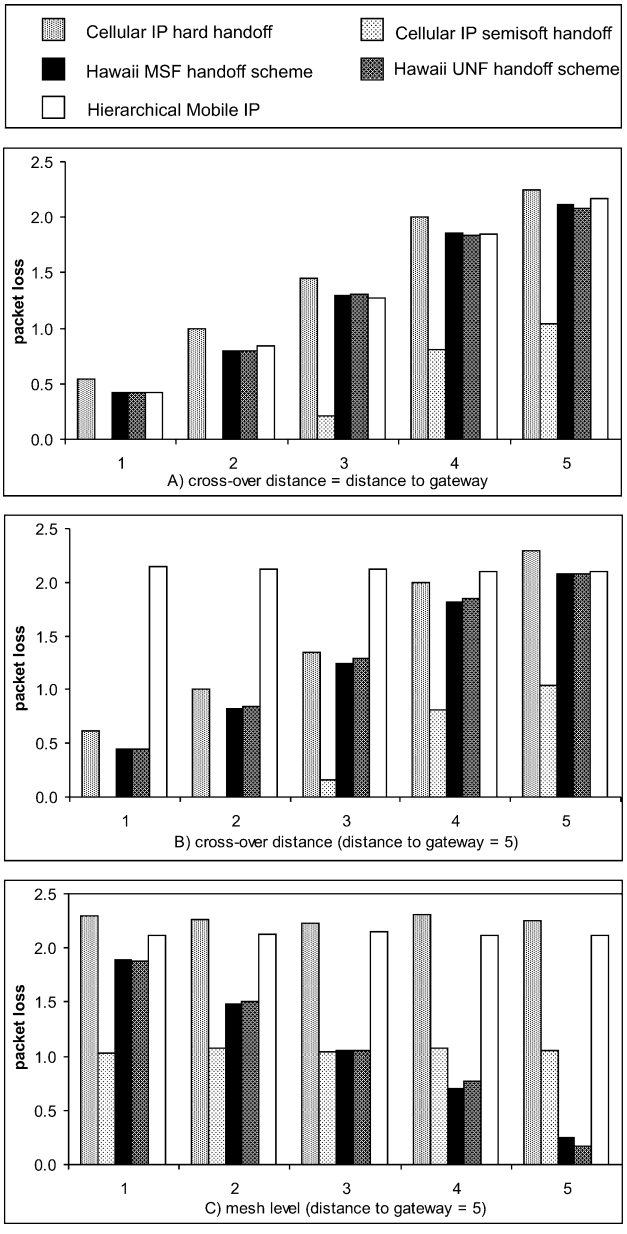

Fig. 6 Simulation results, showing the influence of the cross-over distance and the presence of mesh links on the average number of packet losses during 1 handoff

In this case, the formula for the packet loss is the same for CIP hard handoff, HAW and HMIP (see Section 4). Nevertheless, it should be taken into account that the data packets of HMIP are encapsulated and that the size of the update messages of HAW and HMIP slightly differs from CIP. The simulation results for HAW and HMIP show a systematically lower packet loss than for CIP. This is caused by a specific artefact in the ns-2 CIMS implementation [6]: using HAW or HMIP, the MH only stops listening to the oAR at the moment that the handoff message arrives in the nAR and not when it sends the handoff message as by CIP.

Diagram *B* considers situations in which the cross-over node is the same for CIP and HAW, but this node is not necessarily the domain gateway. This is the case for the handoff from AR2 → AR3 in the tree topology of Fig. 5. The cross-over distance equals 2, while the distance to the domain gateway counts 3 hops. For the simulations, the distance to the gateway is always 5 and the cross-over distance for CIP and HAW varies from 1 to 5. As shown in the diagram, the cross-over distance becomes the determining factor for the packet loss of CIP and HAW, in contrast to the gateway distance for HMIP. These results show that the packet loss for CIP and HAW can be much lower than for HMIP, due to the fact that the gateway distance is often much higher than the cross-over distance. The formulae for CIP and HAW are still the same, but differ from the formula for HMIP.

The fact that the cross-over node can be different for CIP and HAW is illustrated in diagram *C*. In a situation like the handoff from AR4 → AR5 in the mesh topology of Fig. 5, the cross-over node for CIP is the domain gateway, resulting in a cross-over distance 3. In contrast, HAW uses the mesh link with mesh-level 2 to find a shorter route to the old access router. For the resulting cross-over node, the cross-over distance has value 2. For the simulations, the distance to the gateway for HMIP as well as the cross-over distance for CIP is always 5. Only HAW takes advantage of the mesh links to reduce the cross-over distance and as a result also the packet losses. For higher mesh-levels, i.e. mesh links closer to the access routers, the packet loss of HAW decreases drastically compared to CIP and HMIP. The formula for HMIP is the same as for CIP in this situation. As HAW MSF uses no buffer, the formula is the same as for HAW UNF.

5.2. Handoff delay and buffer time parameters

The top diagram of Fig. 7 investigates the average amount of lost or duplicated packets during a Cellular IP semisoft handoff and compares these results with Cellular IP hard handoff. For the simulations, a topology is considered in which both $n_{\mathrm{oAR}\rightarrow\mathrm{CN}}$ and $n_{\mathrm{nAR}\rightarrow\mathrm{CN}}$ (i.e. the cross-over distance) equal 3. The buffer size of the access routers is 200 packets, which is large enough to receive data packets during 2 s. As the handoff delay T_{ss} varies from 0 to 50 ms, no packets are deleted due to buffer overflow in the new access router.

The parameter T_{ss} is highly dependent on the topology of the access network. For the considered topology, the calculated minimum value is pointed to in the diagram. For a smaller value, the MH switches to the nAR too soon and some packets are lost. When T_{ss} further decreases, the packet loss increases and finally reaches the value of the hard handoff packet loss, indicated by the horizontal line in the figure. However, when T_{ss} is higher than the above calculated minimum value, the update message reaches the cross-over node and updates all the necessary route cache mappings within

Table 3 Comparison of the simulation results and the values calculated by the formulae for the packet loss

A) Cross-over dist	1		2		3		4		5	
	sim	calc	sim	calc	sim	calc	sim	calc	sim	calc
CIP hard handoff	0.545	0.545	1.000	1.000	1.445	1.390	2.000	1.812	2.250	2.235
CIP semisoft handoff	0.000	0.000	0.000	0.000	0.210	0.190	0.810	0.612	1.040	1.035
Hawaii MSF	0.415	0.535	0.800	1.000	1.295	1.376	1.855	1.796	2.115	2.217
Hawaii UNF	0.415	0.535	0.800	1.000	1.305	1.376	1.840	1.796	2.085	2.217
HMIP	0.415	0.548	0.840	1.000	1.275	1.396	1.850	1.820	2.165	2.244
B) Cross-over dist	1		2		3		4		5	
	sim	calc	sim	calc	sim	calc	sim	calc	sim	calc
CIP hard handoff	0.615	0.545	1.000	1.000	1.350	1.390	2.000	1.812	2.300	2.235
CIP semisoft handoff	0.000	0.000	0.000	0.000	0.160	0.190	0.810	0.612	1.040	1.035
Hawaii MSF	0.440	0.534	0.825	1.000	1.250	1.376	1.815	1.796	2.080	2.217
Hawaii UNF	0.450	0.534	0.845	1.000	1.285	1.376	1.845	1.796	2.080	2.217
HMIP	2.150	2.244	2.120	2.244	2.125	2.244	2.095	2.244	2.095	2.244
C) Mesh level	1		2		3		4		5	
	sim	calc	sim	calc	sim	calc	sim	calc	sim	calc
CIP hard handoff	2.300	2.235	2.265	2.235	2.230	2.235	2.310	2.235	2.250	2.235
CIP semisoft handoff	1.035	1.035	1.070	1.035	1.045	1.035	1.080	1.035	1.055	1.035
Hawaii MSF	1.890	2.000	1.480	1.580	1.055	1.159	0.700	0.738	0.250	0.318
Hawaii UNF	1.880	2.000	1.505	1.580	1.050	1.159	0.770	0.738	0.175	0.318
HMIP	2.120	2.244	2.130	2.244	2.145	2.244	2.115	2.244	2.110	2.244

T_{ss}. The mobile host receives data packets via the oAR that are also sent towards the nAR. After switching to the nAR, all the packets in the buffer of the nAR are forwarded to the MH. For an increasing T_{ss} value, the number of packets in the buffer and thus the number of duplicated packets increases. The amount of duplicated packets can be approached by $(T_{ss} - T_{ss\_min})r$ (a negative value indicates packet loss).

In the bottom diagram of Fig. 7, the MSF handoff scheme of Hawaii is used to perform handoff. The reference line indicates the packet loss for the UNF handoff scheme. The results are obtained for the same topology as the Cellular IP semisoft handoff. As there are no mesh links, the number of hops between the access routers equals 6.

The buffer size B_{size} is set to 5 packets. The topology determines the minimum B_{time}, which is also indicated in the diagram. When MSF uses no buffer in the oAR, the average packet loss is the same as for UNF: all the data packets that pass the cross-over node before the first path setup message, are lost. If a buffer is present, all data packets that arrived in the buffer within the last B_{time} seconds before the arrival of the path setup message are sent to the nAR. For a value smaller than the minimum B_{time}, some packets are lost. For a higher value, some data packets in the buffer were successfully sent via the oAR and are still forwarded to the nAR, resulting in duplicated packets. Analogously, $(B_{time} - B_{time\_min})r$ is an approximation of the amount of duplicated packets.

Note that the results in the two diagrams of Fig. 7 are very similar. This is because the considered topology and the goal of the used mechanism are the same for the two cases. Both the handoff delay T_{ss} and the buffer time B_{time} are used to reduce the packet loss during handoff.

5.3. Load balancing

In order to study the load balancing in the access network, we consider a situation where the MH visits several cells and we observe how the data and control traffic are spread over the wired links. The simulation results shown in Fig. 8 and Fig. 9 are obtained for the random topology of Fig. 5, which consists of a tree topology with 4 mesh links and 4 additional uplinks. For the exact topology and the significance of the used link names, we also refer to Fig. 5. During the simulation, the MH moves from the leftmost to the rightmost access router, performing 7 handoffs. The Hawaii MSF scheme uses no buffer, while in the case of Cellular IP semisoft handoff, a T_{ss} of 12 ms is used. The figures present the average number of control and data packets that pass on the different links during 1 simulation. For example, *B2down* monitors the number of packets routed from the node with $d = 1$ to the node with $d = 2$, while *B2up* counts the packets routed in the opposite direction. The same explanation is valid for *randS2down* and

Table 4 Values of several simulation parameters in ns-2. In the access routers, the packets have an additional delay of 25 μs

Simulation parameter	CIP	HAW	HMIP
data packet size (bytes)	210	210	230
protocol update message size (bytes)	70	48	72
d_{link} for data packet (ms)	2.168	2.168	2.184
d_{link} for update message (ms)	2.056	2.038	2.058
$d_{MH\rightarrow nAR}$ for update message (μs)	951	863	959
d_{RTS} (μs)		226	
d_{CTS} (μs)		162	
d_{data} for data packet (μs)		1108	

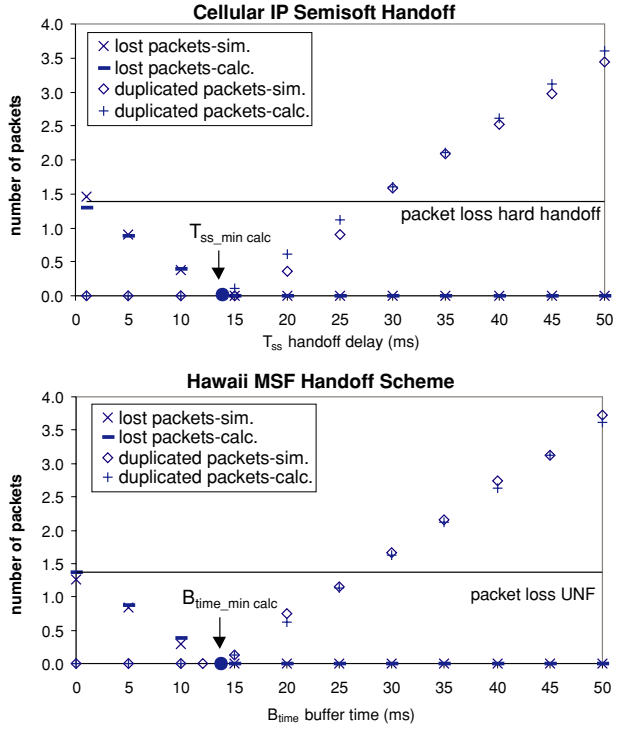

Fig. 7 The average number of lost or duplicated packets during one handoff for a specific topology with cross-over distance 3. Also the optimal value of the Cellular IP semisoft handoff delay T_{ss} and of the buffer time B_{time} for the Hawaii MSF handoff scheme, are indicated

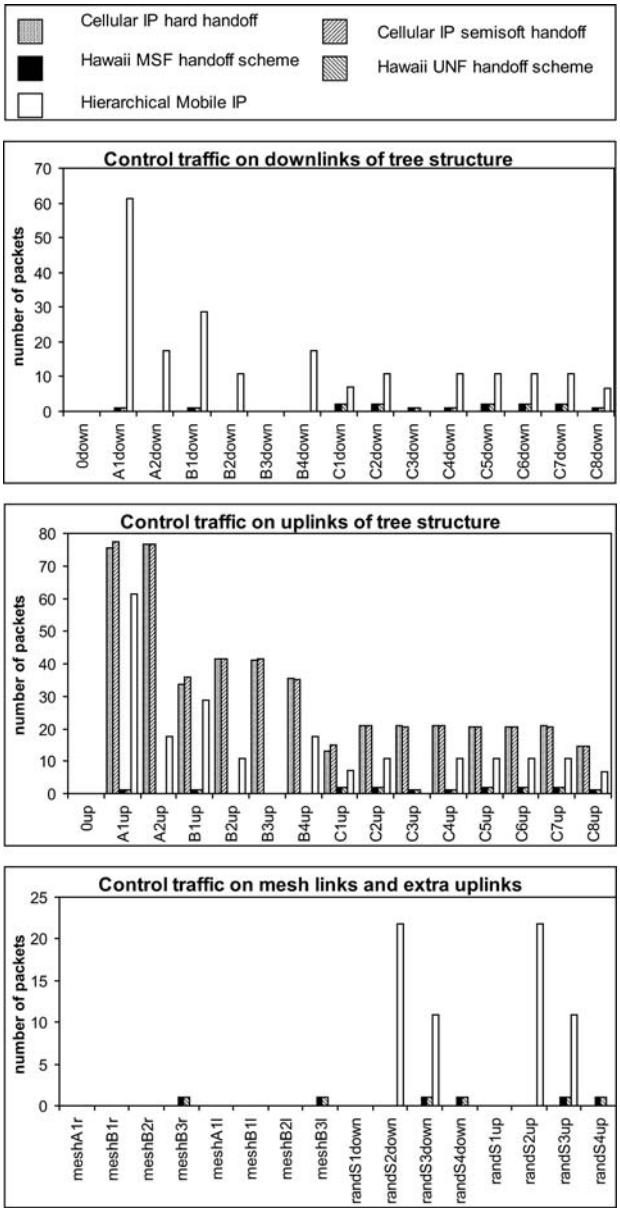

Fig. 8 Average number of control packets that pass on the links of the access network, which has the random topology of Fig. 5. During the simulation, the MH is the receiver of a CBR traffic stream (see Fig. 9) and performs 7 handoffs to maintain its connection with the sender

randS2up. For a mesh link like *meshA1*, *meshA1r* monitors packets travelling from the left to the right router with $d = 1$ and *meshA1l* is used for packets routed from the right to the left router.

Figure 8 gives the results for the control load. CIP uses only the uplinks of the tree structure for its control traffic: page update and route update messages sent by the MH, are directed towards the domain gateway. There are much more control packets on the links than the number of handoffs, due to the use of a registration interval of 0.5 s to update the soft-state routes. The regional registration requests by HMIP are also directed to the domain gateway, now with a registration interval of 1 s, but every request is answered by a registration reply, resulting in the same amount of control traffic on the uplinks and corresponding downlinks. HAW sends a path setup message from the new to the old access router, which replies with a setup message from the old to the new access router. This protocol also uses the presence of mesh links to find the shortest path between these access routers, resulting in control traffic especially on the mesh links and the links closest to the access routers. HAW uses soft-state routes, but the use of refresh messages is not implemented in the CIMS suite, so no periodical route updates are sent and the amount of control traffic on the links is much lower in comparison with the two other protocols. Otherwise, the control traffic load would be higher, depending on the registration interval.

As a connection is set up between a fixed sender in the core network and the receiving mobile host, all data packets have to be routed from the core network, via the domain gateway and the access network, towards the mobile terminal. Figure 9 shows that CIP uses only the downlinks of the tree structure. In addition, links at the same distance of the domain gateway are equally loaded, resulting in a good load balancing for the links of the tree. The link load decreases with increasing distance to the gateway. HAW shows completely different results: the number of data packets monitored on *A1down* equals the number of packets on *0down*, while *A2down* is

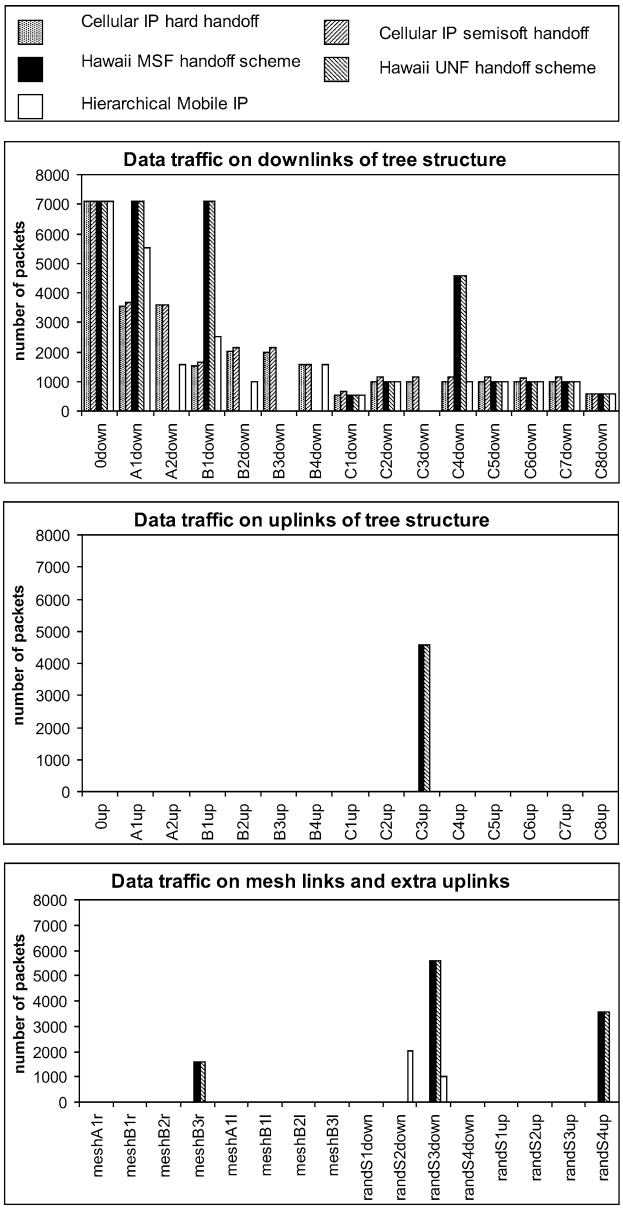

Fig. 9 Average number of data packets that pass on the links of the access network during 1 simulation. The data traffic is sent by a fixed terminal in the core network to the MH. Hence all data traffic is routed via the domain gateway. The load caused by the control traffic to maintain the routing tables of the routers during the handoffs is pictured in Fig. 8

not used during the entire simulation. This results in a bad load balancing for HAW. Also other links than the downlinks of the tree topology are used, namely *randS3down*, *C3up*, *randS4up* and *meshB3r*. The results for HMIP are similar to those of CIP. Also the additional links *randS2* and *randS3* are used to route the data traffic, resulting in an inferior load balancing for the tree links. The mesh links are not used.

The different results for the three protocols can be explained by the differences between the handoff mechanisms. These differences have important implications for the routing within the access network. The handoff mechanism in HAW will result in the use of suboptimal routes after several handovers. When the mobile host, initiating its connection while being in the area of the leftmost access router, arrives in the area of the rightmost access router, the data packets are routed via the links *A1-B1-randS3-C3-C4-randS4-meshB3-C8*. This path counts 8 hops, while the shortest path between an access router and the domain gateway has only 3 hops. In addition, the paths depend on the moving pattern of the mobile host and the location of the previously visited access points, which is an undesirable characteristic. Even for several mobile hosts, the links close to the access routers are more loaded than by the use of CIP or HMIP and the network using HAW is more sensitive to a concentration of users setting up a connection.

5.4. Ring topology

An interesting structure for the access network is a ring topology, as illustrated in Fig. 10. This is rather a simple topology, but completely different from the tree structure which is generally assumed for the study of micromobility protocols. During one simulation, the mobile host travels from the leftmost to the rightmost access router, meanwhile performing 7 handoffs.

Figure 11 shows the average packet loss during 1 simulation. The results are quite different for the three protocols. Hawaii (HAW) is the protocol with the lowest packet loss, due to the fact that path setup messages are exchanged between the access routers. For every handoff, the oAR is only 1 hop away from the new one because of the presence of a direct mesh link. The HAW MSF handoff scheme shows no packet loss, due to the use of buffers with a B_{size} of 5 packets and a B_{time} of 12 ms, which is large enough to reduce the packet loss to zero. In contrast, Hierarchical Mobile IP (HMIP) sends a regional registration request towards the domain gateway, which is minimum 3 (last handover) and maximum 6 (third and fourth handoff) hops away, resulting in a much higher packet loss. This high packet loss is not desired in case of a handoff between two access routers with a direct connection. Cellular IP (CIP) sends route updates via

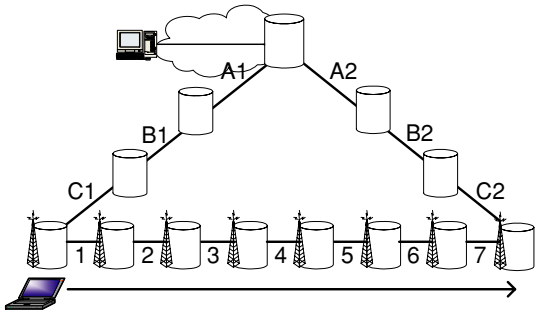

Fig. 10 Access network with a ring topology

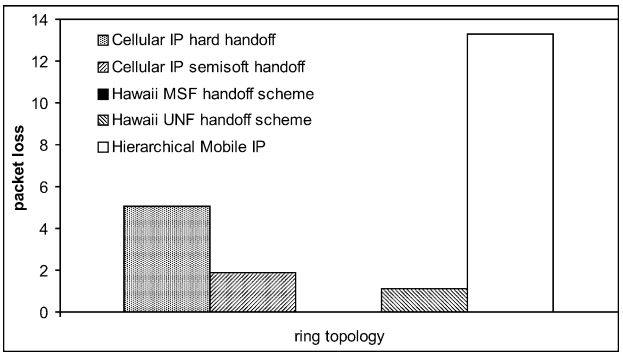

Fig. 11 Average packet loss during one simulation, which implies that the MH moves from the leftmost to the rightmost access router of Fig. 10 and performs 7 handoffs

the shortest path to the domain gateway. When the oAR is situated on this path, the cross-over node is only 1 hop away and the packet loss is comparable to HAW. However, during the fourth handoff, route updates are sent via the right side of the ring and the gateway has the function of the cross-over node. This explains why the packets loss is higher for CIP than for HAW. For the next handoffs, the nAR is the cross-over node and only very few packets get lost. In case of CIP semisoft handoff, using a handoff delay T_{ss} of 12 ms, only during the fourth handoff a small packet loss occurs. The domain gateway is the cross-over node for only one handoff, so the packet loss is much lower for CIP than for HMIP.

Figure 12 represents the average number of data packets that pass on the different links during 1 simulation and Fig. 13 shows the average number of control packets. The links that are not used by the data traffic, are not shown in the diagram. Both CIP and HMIP use both sides of the ring. Their route updates and regional registration requests are directed to the domain gateway via the shortest route. This means that for the first three handoffs, the left side of the ring is used, while the update packets for the last four handoffs result in data routes using the right side of the ring. HAW uses only one side of the ring, in this case the left-hand side because the mobile host initiates its connection via the leftmost access router. The paths used should however not be influenced by the original location of the mobile terminal, since it makes the protocol sensitive to concentrations of mobile users at the start of the simulation. The amount of control packets is much lower for HAW than for CIP and HMIP, because of the same reason as in Section 5.3

5.5. End-to-end delay

Besides the fact that one desires a low packet loss during handoff and a good load balancing on the links of the access network, it is also important that a micromobility protocol results in the use of optimal routes within the access network. This means that the used path between the domain gateway

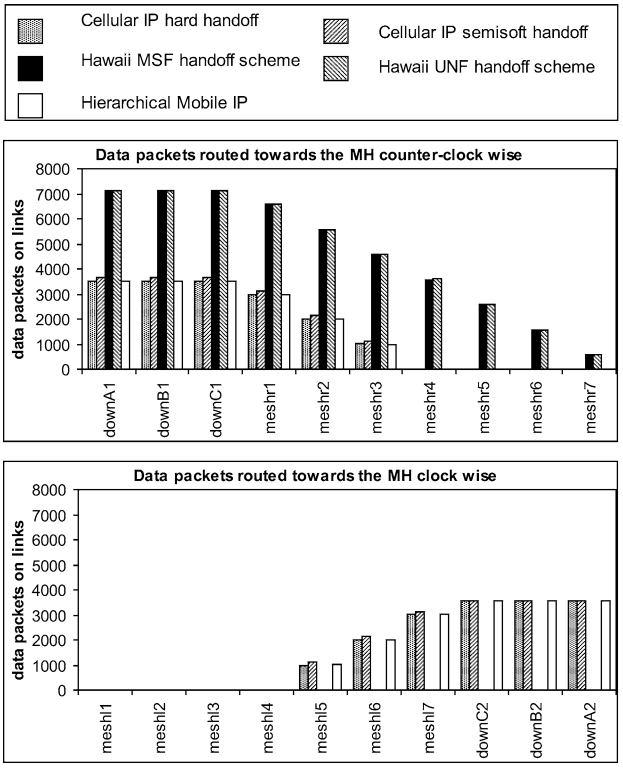

Fig. 12 Average number of data packets on the links of the ring topology during 1 simulation. The data traffic is routed via the domain gateway towards the MH

and the current access router should have a minimum end-to-end delay or should consist of a minimum number of hops if the delay of all the wired links is the same (which is 2 ms for the simulations). The use of paths with only a few hops, decreases the total load on the access network and the chance that the path goes down due to a broken link.

The results of Fig. 14 are obtained for a tree, mesh, random and ring topology (see Fig. 5 and Fig. 11). During one simulation, the mobile host moves again from the leftmost access router to the rightmost access router and performs 7 handoffs to maintain its connection. The results shown in Fig. 14 represent the average number of hops of the used path between the domain gateway and the access routers. Starting from the total end-to-end delay $d_{\text{end-to-end}}$, i.e. the time between the sending and the arrival of a data packet, the average number of hops in the access network can be calculated, using:

$$\text{hop count} = \frac{d_{\text{end-to-end}} - d_{\text{wireless}} - d_{\text{core}}}{d_{\text{link}}} \tag{14}$$

In this formula, d_{wireless} is the time needed to send the data packet over the wireles link and d_{core} is the time the packet spends in the core network.

For Cellular IP (CIP), both hard and semisoft handoff result in the same paths. Analogously, the path setup scheme used by Hawaii (HAW), MSF or UNF, has no influence on the resulting paths. So only the results for CIP hard handoff and

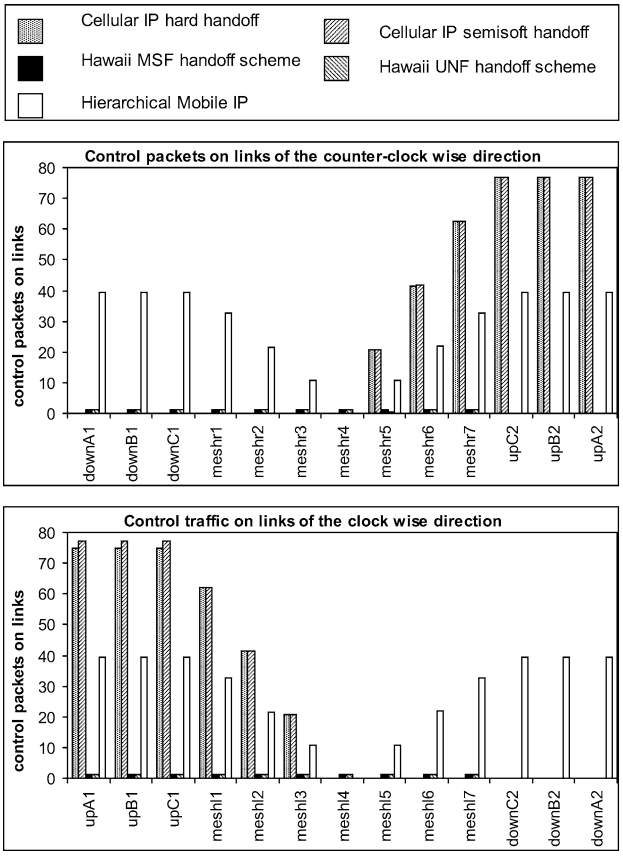

Fig. 13 Average number of control packets on the links of the ring topology during 1 simulation, necessary to maintain and update the routing tables of the routers in the access network

HAW MSF are shown. The shortest path from the domain gateway to an arbitrary access router always counts 3 hops, except for the ring topology. CIP and Hierarchical Mobile IP always use paths with this minimum number of hops. In case of the ring topology, both protocols start routing the data packets via the left side of the ring and switch after the fourth handoff to the right side. In contrast, HAW uses suboptimal routes: additional meshes and uplinks are used to find a shorter route between the access routers to exchange the path setup messages, but this results in longer data paths.

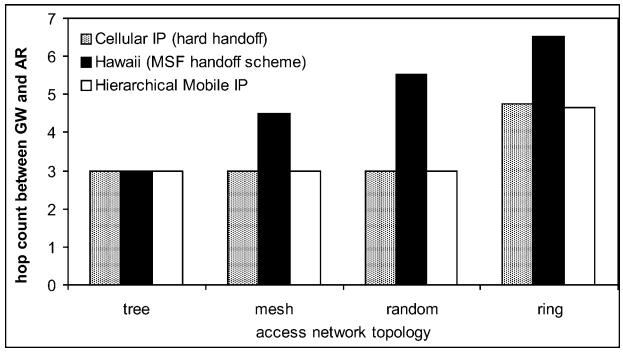

Fig. 14 Average hop count between the domain gateway and the access routers for several topologies

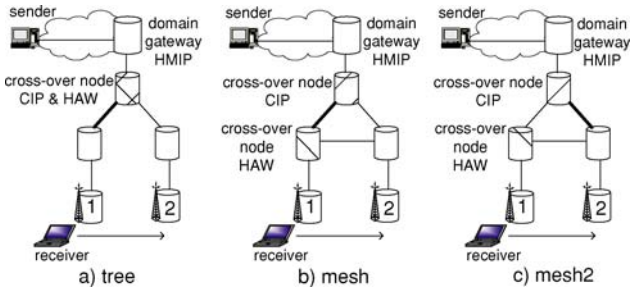

Fig. 15 Access network topologies used to study the robustness against link failures

5.6. Robustness against link failures

The three protocols react differently on link failures and depending on the topology of the access network, Hawaii (HAW) has either a better, or a worse performance than Cellular IP (CIP) and Hierarchical Mobile IP (HMIP). Figure 15 shows the basic topologies and scenarios under study. During the simulations, the link indicated in bold goes down for a period of 15 s. After the link goes down, the mobile host starts moving from the left to the right access router, with a speed of 20 m/s. The results are shown in Fig. 16.

Using the scenario of Fig. 15(a), the mobile host sets up a connection via access router 1. When the indicated link goes down, packets get lost. Since the mobile terminal moves towards access router 2, the mobile host tries to restore its connection via this new access router, as soon as it receives a beacon from this base station. In case of CIP, a route update is routed via base station 2 to the domain gateway. Similarly, HMIP sends a registration request towards the gateway. As the path between the domain gateway and access router 2 has no broken links, a new connection is set up and no more packets get lost. HAW tries to send a path setup message from the new to the old access router, but fails because, as long as the indicated link is down, there is no path available between the two base stations and the handover process fails. Packets get lost during the whole period that the link is down. When the link is up again, HAW still has to perform a handoff and a few additional packets may get lost.

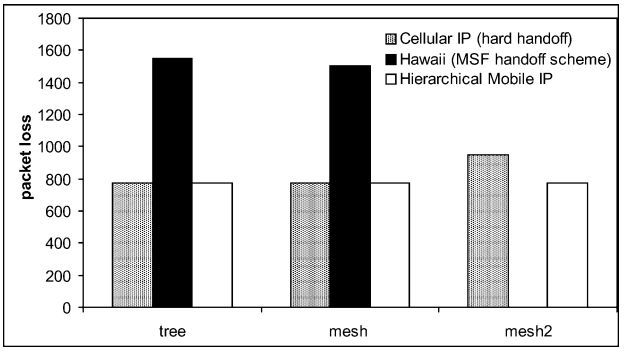

Fig. 16 Average packet loss during 1 simulation caused by link failure

Table 5 Overview of the handoff mechanisms of the studied micromobility protocols

Handoff mechanism	Advantages	Disadvantages	CIP	HAW	HMIP
Send RU towards ...					
GW.	Short paths between GW and AR.	RU traverses whole access network.	X	-	X
oAR.	Load of RU concentrated near oAR and nAR.	Suboptimal paths, influenced by mobility pattern.	-	X	-
RU updates ...					
every router on path.	Fast update.	Every router processes RU.	X	X	-
selected routers.	More scalable.	Slower update, determined by distance between selected routers.	-	-	X
Selected routers use ...					
host routes, no standard IP routing.	Simplicity.	Router is unaware of access network topology.	X	-	-
host routes and standard IP routing.	Router knows access network topology.	Larger routing tables.	-	X	X
State of routes is ...					
soft-state.	Old, unused routes time-out.	More control traffic due to periodical updates.	X	X	X
hard-state.	Less control traffic.	Old and invalid routes have to be deleted.	-	-	-
Report successful handoff.	MH retries untill handoff succeeds.	Extra control traffic.	-	X	X
MH runs ...					
micromobility protocol.	MH can use features of protocol.	MH must be aware of used protocol.	X	-	X
macromobility protocol.	MH can use Mobile IP, independently of used protocol.	MH can not use features of local protocol.	-	X	-
Semisoft handoff.	Reduction of handoff packet loss.	Bicasting of data packets gives more traffic, appropriate handoff delay depends on topology.	X	-	-
Forward packets from oAR to nAR	Reduction of handoff packet loss.	Data packets may arrive out of order, appropriate buffer size depends on topology.	-	X	-

The topology of Fig. 15(b) has an extra mesh link, so that the path setup messages of HAW can reach the old base station and the handoff process can complete without any problems. Unfortunately, the broken link is also part of the (suboptimal) path between the gateway and the new base station, and the mobile host is still not able to receive packets. Due to the fact that the handover process has already finished, the routing tables of the routers are already updated and the mobile host can receive data as soon as the link is up again (the packet loss is a little smaller compared to scenario *a*).

The scenario of Fig. 15(c) is slightly different. The link going down is part of the shortest path between the gateway and the new access router. The route updates of CIP and the regional registration requests of HMIP can not reach the gateway and the handover fails. In case of CIP, the mobile host does not know that the handover process failed and the use of periodical page updates (sent every 3 s) finally results in a successful route update. HMIP uses registration replies, so for every new beacon (sent with a period of 1 s), the host sends a new registration request until it receives a reply to report a successful registration. This explains the lower packet loss of HMIP compared to CIP. In contrast, HAW does not use this link in its path to the new access router and does not notice that the link is down. The only packet losses occur during the handoff process.

Thus when a link in the access network goes down, several situations are possible depending on the topology and the used protocol. It is obvious that an access network with a random topology can solve many of the problems: additional uplinks result in the presence of alternative paths. This requires however a protocol that provides dynamic routing, in contrast to the studied micromobility protocols that do not recalculate their routing tables after a link failure has occcured.

6. Conclusion

This article studies the relationship between access network topology (namely the tree, mesh and random type) and micromobility protocol performance. A ring topology is considered as a special case of the mesh topology type.

This study is conducted for the popular Cellular IP (CIP), Hawaii (HAW) and Hierarchical Mobile IP (HMIP) protocols, supporting local mobility. Formulae for the packet loss during handoff and for the handoff network time, indicate that

these metrics highly depend on the topology. The simulation results confirm this statement.

Table 5 gives an overview of several aspects of the handoff mechanism and indicates their advantages and drawbacks (RU stands for Route Update). Due to the differences in handoff mechanism, the cross-over distance is a very important handoff parameter for CIP and HAW, while the number of hops to the gateway is the key factor for the amount of handoff packet loss by HMIP. Depending on the topology, CIP and HAW can have a different location of the cross-over node. For a pure tree topology, the cross-over node is the same and both protocols have a similar performance.

Another important aspect is how mesh links are used by the protocols. In contrast to CIP and HMIP, HAW takes advantage of extra mesh links to reduce the handoff latency and packet loss drastically. However, the routing mechanism of HAW also results in the use of a suboptimal route after several handoffs, which gives a bad load balancing in the access network. In addition, the used path depends on the mobility pattern of the mobile host and its location at the time the connection was set up, as illustrated for a random and ring topology.

The CIP semisoft mechanism and the HAW MSF handoff mechanism are mechanisms that allow us to reduce the packet loss during handoff. However the handoff delay T_{ss} and the buffer time B_{time} respectively have to be carefully chosen. The optimal values of these parameters, resulting in a number of lost and duplicated packets as low as possible, are also strongly determined by the topology and the cross-over distances. Thus the use of these mechanisms for an arbitrary topology is not straight forward.

The investigation of the end-to-end delay, confirms the fact that HAW, although very successful in reducing the packet loss during handover, results in suboptimal routes inside the access network. In case of a pure tree topology, the topology that was initially assumed for the development of these micromobility protocols, all micromobility protocols use the same routes.

The study about the robustness against link failures revealed that in case of CIP and HMIP, the reachability of the domain gateway by the new access router is necessary to perform a successful handoff. In case of HAW, the old access router must be reachable from the new one. Due to the lack of registration replies in case of CIP, the mobile host is not aware of failed handoffs and reacts much slower when the broken link is up again.

Acknowledgments

Liesbeth Peters is a Research Assistant of the Fund for Scientific Research - Flanders (F.W.O.-V., Belgium). Part of this research is funded by the Belgian Science Policy Office (BelSPO, Belgium) through the IAP (phase V) Contract No. IAPV/11, and by the Institute for the promotion of Innovation by Science and Technology in Flanders (IWT, Flanders) through the GBOU Contract 20152 “End-to-End QoS in an IP Based Mobile Network”.

References

1. 3GPP A Global Initiative, http://www.3gpp.org
2. ANSI/IEEE Std 802.11, Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications, IEEE (1999).
3. A. Campbell and J. Gomez, IP micro-mobility protocols, ACM SIGMOBILE Mobile Computing and Communications Review 4 (October 2000) pp. 45–53.
4. A. Campbell, J. Gomez, S. Kim, A. Valkó, C. Wan and Z. Turanyi, Design, implementation and evaluation of Cellular IP, IEEE Personal Communications (August 2000) pp. 42–49.
5. A. Campbell, J. Gomez, S. Kim, C. Wan, Z. Turanyi and A. Valkó, Comparison of IP micromobility protocols, IEEE Wireless Communications (February 2002) pp. 72–82.
6. Columbia IP micro-mobility suite (CIMS), http://www.comet.columbia.edu/micromobility.
7. M. Frodigh, S. Parkvall, C. Roobol, P. Johansson and P. Larsson, Future-generation wireless networks, IEEE Personal Communications (October 2001) pp. 10–17.
8. E. Gustafsson, A. Jonsson and C. Perkins, Mobile IPv4 regional registration, ftp://ftp.ietf.org/internet-drafts/draft-ietf-mobileip-reg-tunnel-07.txt (October 2002, work in progress).
9. D. B. Johnson, C. Perkins and J. Arkko, Mobility support in IPv6, ftp://ftp.ietf.org/internet-drafts/draft-ietf-mobileip-ipv6-21.txt (February 2003, work in progress).
10. W. Mohr and W. Konhäuser, Access network evolution beyond third generation mobile communications, IEEE Communications Magazine (December 2000) 122–133.
11. NS-2 home page, http://www.isi.edu/nsnam/ns.
12. C. Perkins, Ed., IP mobility support for IPv4, *IETF RFC 3344* (August 2002).
13. L. Peters, I. Moerman, B. Dhoedt and P. Demeester, Influence of the topology on the performance of micromobility protocols. in: *Proceedings of the workshop WiOpt'03 “Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks”* (Sophia Antipolis, France, March 2003) pp. 287–292.
14. L. Peters, I. Moerman, B. Dhoedt and P. Demeester, Performance of micromobility protocols in an access network with a tree, mesh, random and ring topology. in: *Proceedings of the IST Mobile & Wireless Communications Summit 2003 “Enabling a Pervasive Wireless World”*, (Aveiro, Portugal, June 2003) pp. 63–67.
15. R. Ramjee, T. La Porta, L. Salgarelli, S. Thuel and K. Varadhan, IP-based access network infrastructure for next-generation wireless data networks, IEEE Personal Communications (August 2000) pp. 34–41.
16. R. Ramjee, K. Varadhan, L. Salgarelli, S. Thuel, S. Wang and T. La Porta, HAWAII: a domain-based approach for supporting mobility in wide-area wireless neworks, IEEE/ACM Transactions on Networking 10(3) (June 2002) 396–410.
17. P. Reinbold and O. Bonaventure, A comparison of IP mobility protocols, IEEE SCVT 2001 Proceedings (2001).
18. UMTS Forum, http://www.umts-forum.org
19. A. Valkó, Cellular IP: a new approach to internet host mobility, ACM Computer Communication Review (January 1999).
20. A. Valkó, J. Gomez, S. Kim and A. Campbell, On the analysis of Cellular IP access networks, *IFIP Sixth International Workshop of Protocols for High Speed Networks (PfHSN'99)*, (Salem Massachussetts, August 1999).

Liesbeth Peters was born in Temse, Belgium, in 1978. She received her Master of Science degree in Electrotechnical Engineering from Ghent University, Gent, Belgium in 2001. Since August 2001, she has been working as a doctoral researcher with the Department of Information Technology (INTEC) of the Faculty of Applied Sciences, Ghent University, where she joined the Broadband Communications Networks Group. Since October 2002, she works there as a research assistant of the Fund for Scientific Research—Flanders (F.W.O.-V., Belgium). Her current research interests are in broadband wireless communication and the support of IP mobility in wired cum wireless networks.

Ingrid Moerman was born in Gent, Belgium, in 1965. She received the degree in Electro-technical Engineering and the Ph.D degree from the Ghent University, Gent, Belgium in 1987 and 1992, respectively. Since 1987, she has been with the Interuniversity Micro-Electronics Centre (IMEC) at the Department of Information Technology (INTEC) of the Ghent University, where she conducted research in the field of optoelectronics. In 1997, she became a permanent member of the Research Staff at IMEC. Since 2000 she is part-time professor at the Ghent University. Since 2001 she has switched her research domain to broadband communication networks. She is currently involved in the research and education on broadband mobile & wireless communication networks and on multimedia over IP. Her main research interests related to mobile & wireless communication networks are: adaptive QoS routing in wireless ad hoc networks, personal networks, body area networks, wireless access to vehicles (high bandwidth & driving speed), protocol boosting on wireless links, design of fixed access/metro part, traffic engineering and QoS support in the wireless access network. Ingrid Moerman is author or co-author of more than 300 publications in the field of optoelectronics and communication networks.

Bart Dhoedt received a degree in Engineering from the Ghent University in 1990. In September 1990, he joined the Department of Information Technology of the Faculty of Applied Sciences, University of Ghent. His research, addressing the use of micro-optics to realize parallel free space optical interconnects, resulted in a PhD degree in 1995. After a 2 year post-doc in opto-electronics, he became professor at the Faculty of Applied Sciences, Department of Information Technology. Since then, he is responsible for several courses on algorithms, programming and software development. His research interests are software engineering and mobile & wireless communications. Bart Dhoedt is author or co-author of more than 100 papers published in international journals or in the proceedings of international conferences. His current research addresses software technologies for communication networks, peer-to-peer networks, mobile networks and active networks.

Piet Demeester finished his PhD thesis at the Department of Information Technology (INTEC) at the Ghent University in 1988. At the same department he became group leader of the activities on Metal Organic Vapour Phase Epitaxial growth for optoelectronic components. In 1992 he started a new research group on Broadband Communication Networks. The research in this field resulted in already more than 300 publications. In this research domain he was and is a member of several programme committees of international conferences, such as: ICCCN, the International Conference on Telecommunication Systems, OFC, ICC, and ECOC. He was Chairman of DRCN'98. In 2001 he was chairman of the Technical Programme Committee ECOC'01. He was Guest Editor of three special issues of the IEEE Communications Magazine. He is also a member of the Editorial Board of the Journals "Optical Networks Magazine" and "Photonic Network Communications". He was a member of several national and international PhD thesis commissions. Piet Demeester is a member of IEEE (Senior Member), ACM and KVIV. His current research interests include: multilayer networks, Quality of Service (QoS) in IP-networks, mobile networks, access networks, grid computing, distributed software, network and service management and applications (supported by FWO-Vlaanderen, the BOF of the Ghent University, the IWT and the European Commission). Piet Demeester is currently full-time professor at the Ghent University, where he is teaching courses in Communication Networks. He has also been teaching in different international courses.