Vol. 45 No. 10 IPSJ Journal Oct. 2004

Regular Paper

Improving MIP Handover Latency Using Location Information

Nik van den Wijngaert† and Chris Blondia†

This paper presents a location-aware improvement over the Pre-Registration low latency handoff proposed by the IETF. Pre-Registration requires layer 2 information that might not be present in the access technology being used. By applying location information available to the mobile node, these obstacles can be overcome, allowing more control over the actual handoff latency and buffering requirements. After describing the protocol, we present an analytical model that provides us with results on packet delay and buffering requirements for a CBR packet stream. We have also developed a detailed OPNET simulation where results are obtained in more realistic circumstances.

1. Introduction

In an emerging wireless world, always-on connectivity becomes increasingly important for a large number of mobile IP devices. One of the most successful solutions for network level roaming is Mobile IP (MIP)[13] as proposed by the IETF. This protocol allows a Mobile Node (MN) to change its network point of attachment across heterogeneous access networks without needing to change its IP address. It does so by making use of a so-called care-of address located at a Foreign Agent (FA). Together with the Home Agent (HA), these entities make up the logic of the MIP infrastructure.

MIP contains a number of built-in delay components. To this end, several enhancements have been proposed to the base specification. Hierarchical MIP (HMIP, also referred to as Regional Tunnel Management)[9] counters the signaling induced latency by introducing a tree-like hierarchy of FAs which are grouped into domains governed by a Gateway Foreign Agent (GFA), the root of the tree. For all mobility purposes within that domain, the GFA acts as the HA for the mobile node, reducing the signaling path significantly. To counter the handoff induced latency, the IETF has introduced low latency handoff mechanisms[7] that rely on layer 2 information to facilitate layer 3 handoffs. Unlike protocols like HMIP that rely solely on network level information, these handoffs break the clear separation between network and link layer. Two different methods are proposed: Pre-Registration and Post-Registration, both relying on the same kind of layer 2 information. However, in access technologies like IEE802.11 (also referred to as WiFi)[17], the required triggers are non-existent, making these low latency schemes problematic to implement. We assume usage of IPv4 throughout this paper.

This paper introduces a location augmented variant of Pre-Registration which relies on certain information to be available to the MN, e.g., trajectory of travel (driving straight on the highway, GPS-generated driving routes) or relative position to the access points (APs) the MN can visit. Some work on the topic of using location information has been done before (e.g., Ref. 8)), but there it is assumed that the FAs are aware of their location and possibly communicate this to each other. In our proposal, only the MN is location-aware. By taking this extra information into account, the link layer dependency of Pre-Registration can be removed and the latency and buffer requirements can be lowered. We present results obtained from both an analytical model and an OPNET simulation model.

2. Low Latency Handoff Mechanisms

The IETF has proposed two different low latency schemes. Pre-Registration attempts to complete the layer 3 handoff before the actual layer 2 handoff takes place, whereas Post-Registration allows for handoffs without the MN's intervention by making use of Bidirectional Edge Tunnels (BETs) and letting the MN conduct its registration at a later point in time at its own discretion.

2.1 Layer 2 Triggers

Both approaches rely on the presence of four layer 2 triggers:

- Anticipation trigger (AT): conveys information to layer 3 that a layer 2 move is

† University of Antwerp, Belgium

about to occur and to which access point.

- Old foreign agent (oFA) layer 2 link down trigger (oFA L2LD): the MN can no longer communicate with the oFA.
- New foreign agent (nFA) layer 2 link up trigger (nFA L2LU): the nFA can send traffic to the MN.
- MN layer 2 link up trigger (MN L2LU): the MN can send traffic to the nFA.

For correct operation of the handoff schemes, the following timing relation should hold: AT < oFA L2LD < nFA L2LU < MN L2LU.

Especially this first requirement proves to be problematic for technologies like WiFi [5]: the AT needs to occur in anticipation of a handoff, i.e., when the handoff about to happen and so before any other layer 2 action, and it needs to convey information about which AP is about to be new one.

2.2 Pre-Registration

Pre-Registration works as follows. At the AT, Pre-Registration starts normal MIP handoff with the nFA while still connected to the oFA. Instead of the normal Router Solicitation and Advertisement (Sol/Adv), the MN conducts a Proxy Router Sol/Adv exchange with the oFA that has cached the advertisements of its neighboring agents. A normal Registration Request is then sent from the MN to the nFA through the oFA. The request is processed by the GFA and replied to, routing all traffic via the new path to the nFA. The Registration Reply received by the nFA is cached until the MN has moved into its domain (nFA L2LU). Pre-Registration can be both mobile initiated (the MN sends a Proxy Router Solicitation) or network initiated (the oFA unicasts a Proxy Router Advertisement to the mobile). We will not discuss the Post-Registration method as our scheme is based on Pre-Registration only.

3. Location Augmented Approach

While it is possible to perform some form of low latency handovers over a broad range of access technologies, including WiFi [5], the fundamental problem of the trigger timing cannot be solved. Working with multiple interfaces is one possible solution, but then the difficulties of spacing the AT and oFA L2LD triggers remain. In Pre-Registration buffering is required at the nFA to not only hold the registration reply, but also any traffic routed on the new path from the GFA to the nFA while the MN might still be in its old domain. If the AT arrives too early, traffic is routed to the nFA too soon, causing large buffer buildup and packet delay and delay variation but if the trigger is late then the layer 3 handoff might not have been completed before the MN moves out of range causing packet loss at the oFA.

With location systems like GPS (Global Positioning System) becoming more and more popular and integrated in various devices, it is possible to exploit this information to facilitate handoffs. We propose a scheme based on Pre-Registration that addresses two points:

(1) dependence on layer 2 triggers: by using information embedded in the GPS maps that details the location of access points, triggers can be generated from upper layers;

(2) the actual latency is coupled with the spacing of the AT and oFA L2LD triggers, which our proposal tries to minimize.

One first possible solution, if we have location awareness, is to make MIP conduct a handover to the nearest agent or access point. Given the movement vector, we can roughly estimate when the handoff will occur and generate an AT accordingly. Next we force the wireless interface to handover to a new AP.

The protocol we propose goes one step further and is called Location Augmented Bulk Pre-Registration (LABPR). Assume that we are traveling along a trajectory known beforehand. This trajectory can be deducted from the fact that we are moving on a highway, or from the course plotted by the GPS system. Whenever the MN enters a new domain, it can then send out several Pre-Registration Requests in advance for the agents it is going to visit. It can elect to register an arbitrary number of hops in advance. These requests are processed by each new agent and sent up to the GFA which processes these requests and caches the replies. When the MN is about to switch from one network to the other, it sends a Release Cached Request (RCR) to the GFA for the agent it is moving to, which in turn forwards the cached request to the nFA. If the trajectory of agents to visit is not known in advance, it is possible to perform LABPR to a number of eligible FAs and the timer in the pending entries will ensure that unused registrations are not held indefinitely. The timing of the anticipation trigger to send the RCR is now decoupled from the routing path between the FAs, i.e., only depen-

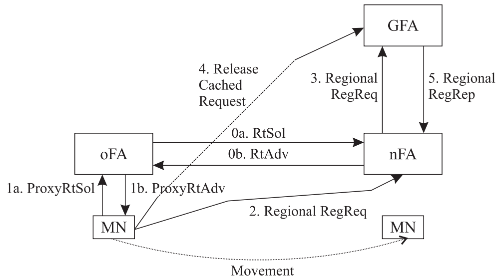

Fig. 1 LABPR.

dent on the path oFA→GFA, and not on the path oFA→nFA→GFA, as is the case in Pre-Registration. This distance can be measured by the MN by pinging the GFA while connected to the oFA. By averaging several of the response times, we can get an estimate of the Round Trip Time (RTT) between the MN and the GFA, which is exactly the time needed between AT and L2LD: the link down can occur when the RCR has reached the GFA and the last packet routed on the old path has reached the MN. The signalization diagram for LABPR is shown in **Fig. 1**.

The oFA here is not necessarily a direct neighbor of nFA, but it is the first agent the MN registers with when entering the new domain at which point messages 0 through 3 are sent. The GFA processes this request (message 3) but does not activate the mobility binding (the current binding stays active) nor does it forward the reply to the nFA. When the AP linked to the nFA becomes the nearest AP, message 4 is sent at the anticipation trigger, causing the GFA to activate its pending binding and forward the Registration Reply and subsequent traffic along the new path. One estimated RTT value later, the link layer is forced to do a handoff to the new access point associated with the new foreign agent. This event is followed by an nFA L2LU trigger at which time the handoff is complete and the buffered packets in the nFA are forwarded to the MN.

4. Analytical Model

4.1 Reference Architecture

Figure 2 shows the reference architecture we will use for the analytical evaluation. Notice the presence of Router 3: it is quite possible that the physical topology is different from the logical foreign agent tree. This has an effect on Pre-Registration, where the distance from the oFA to the nFA is of importance for the latency and buffer requirements: the shorter the distance, the sooner the Registration Request reaches the GFA which then routes traffic along the new path. Similar methods of analytical modeling have been applied to Cellular IP [1),6)], HAWAII [2),16)], Optimized Smooth Handoff [3),14),15)] and Pre- and Post-Registration [4),5)].

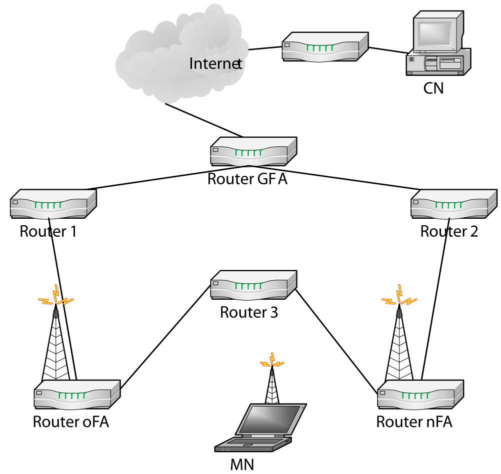

Fig. 2 Reference architecture.

4.2 Model Description

In this section we present an analytical model that allows us to compare several performance measures for Pre-Registration and LABPR, such as the delay characteristics for packets arriving at the MN and buffer requirements in the nFA. In this model, we assume that the router/access point and the agent are located on the same machine. Furthermore, we model all routers in the network as M/M/1 queues, so that every packet going through one of the routers has an exponentially distributed random service time (assumed to be both processing time and transmission time). Consequently, the response time of a packet is also exponentially distributed. Assume that the bulk registrations have happened at some time in the past at an FA governed by the same GFA as the FA the MN is about to move to. The MN performs a handover from oFA to nFA and let t_0 be the time at which the anticipation trigger occurs. We define D_{LD} and D_{LU} as the time between t_0 and the corresponding triggers. It always holds that $0 < D_{LD} < D_{LU}$.

The most important random time instance for both methods is the arrival t_1 of the Registration Request at the GFA for Pre-Registration and the arrival of the RCR message at the GFA for LABPR. Both of these packets are sent at t_0 and their arrival is distributed as the sum of exponential variables (the response time in the routers) and constants (the link propagation delays).

We now consider a CBR UDP packet stream, with packets arriving every T ms at the GFA from a Correspondent Node (CN) destined to the MN. We then observe this stream of packets of which the first arrives at the GFA some time before t_0 and calculate the end-to-end delay distribution of each packet (measured from the GFA to the MN). The delay is also a random variable being the sum of exponential variables and constants. The composition of the delay random variable is determined by the path a packet takes, which is determined by t_1. We will present the model for LABPR; this is based on the model for Pre-Registration detailed in Ref. 4).

Let t^1_{GFA} be the time instant when the first packet arrives at the GFA. It follows then that the arrival of the k-th packet is given by: $t^k_{GFA} = t^1_{GFA} + (k-1) \times T$. We denote the arrival times of the k-th packet at the foreign agents as t^k_{oFA} and t^k_{nFA}. We can then discern four different classes a packet can belong to, depending on the arrival at the GFA and FAs:

(1) $t^k_{GFA} < t_1 \Rightarrow$ k-th packet is sent along the path GFA ↔ oFA
- (a) $t^k_{oFA} < t_0 + D_{LD} \Rightarrow$ packet is delivered directly to the MN through the oFA;
- (b) $t^k_{oFA} > t_0 + D_{LD} \Rightarrow$ packet is lost;

(2) $t^k_{GFA} > t_1 \Rightarrow$ k-th packet is sent along the path GFA ↔ nFA
- (a) $t^k_{nFA} < t_0 + D_{LU} \Rightarrow$ packet needs to be buffered until $t_0 + D_{LU}$;
- (b) $t^k_{nFA} > t_0 + D_{LU} \Rightarrow$ packet is delivered directly to the MN through the nFA.

Since the above four packet classes are disjoint and form a partition of the sample space, we can apply the total probability theorem to calculate $P(arr_k > t)$, with arr_k the point in time when the k-th packet arrives at the MN.

In order to determine the size of the buffer to be installed in the nFA, if we want to minimize losses due to packet arrivals before the D_{LU} trigger, we can compute the distribution of N_b which is the number of packets that would get lost if no buffers were present.

4.3 Numerical Results

The network from Fig. 2 is configured with service rate μ in all routers set to 1 pk/ms and the load ρ in all the routers is 0.8. It follows from the M/M/1 assumption that the response time is exponentially distributed with rate $\mu(1-\rho) = 0.2$ pk/ms. Let T equal 10 ms for the UDP packet stream. The propagation delay is set to 5 ms for all the links.

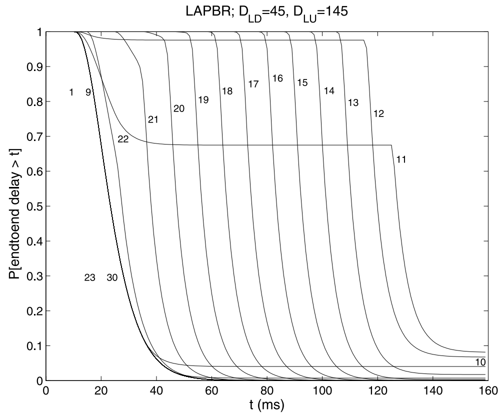

Fig. 3 Individual delays — LABPR.

Let us first look at the individual delays for a stream of 30 packets, starting at 80 ms before t_0. **Figure 3** shows the probability that the end-to-end delay of each packet measured from the GFA to the MN exceeds a certain time t for LABPR when we assume an infinite buffer to be available at the nFA. The RTT in this model is estimated to be 45 ms if we consider the link delays and average router response times. This estimator is taken to be exactly the average RTT value. Due to the high variability in response times throughout the routers, some packets may be lost in the oFA if they arrive too late. This effect can be countered somewhat by adding corrections to the RTT estimator (see Section 5.2). The traffic involved in the handoff experiences low end-to-end delays overall: the signalization route taken by LABPR is shorter than in Pre-Registration and hence subject to lower variability.

Packet loss in the oFA is one of the problems in Pre-Registration, as we have no idea how long the registration message takes to travel from the oFA to the GFA via the nFA. **Figure 4** illustrates this by starting D_{LD} at $t_0 + 0$ ms and gradually increasing its value by 10 ms up to 100 ms. As this value becomes larger, we can see the drop probability of individual packets becoming smaller, but bear in mind that choosing a large trigger value will have an impact on buffer requirements in the nFA (large trigger values imply that traffic will

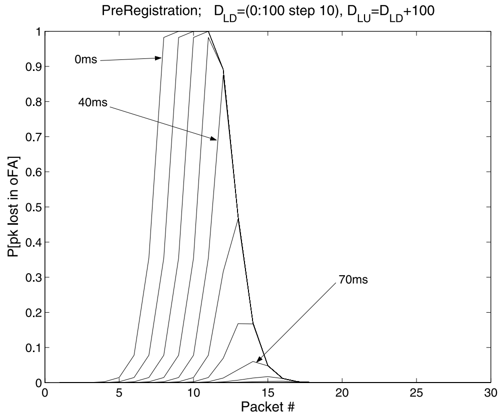

Fig. 4 Loss probability in the oFA — Pre-Registration.

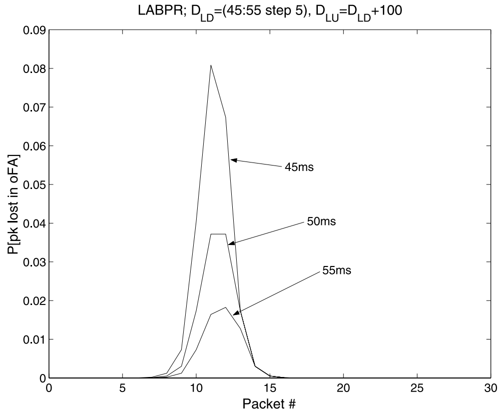

Fig. 5 Loss probability in the oFA — LABPR.

be routed along the new path sooner, resulting in more packets occupying the buffer). In **Fig. 5** we start off with an estimated RTT of 45 ms and gradually increase its value by 5 ms up to 55 ms. Even on the lowest value, only a few packets have a drop probability peaking at 8%.

We can also look at the buffer requirements as a function of the layer 2 trigger timing. We show curves for the expected buffer sizes (= expected number of packets lost in the absence of buffers) and for the minimum buffer sizes required to achieve a probability of loosing a packet smaller than 10-5. On first glance at **Fig. 6**, it seems that Pre-Registration has lower buffer requirements. This is true when comparing both protocols using the same trigger values in both cases: the RCR packet of LABPR arrives much sooner in the GFA than the Registration Request, routing packets to the nFA much earlier. Varying the DLD trigger has no impact on the buffer, which is only

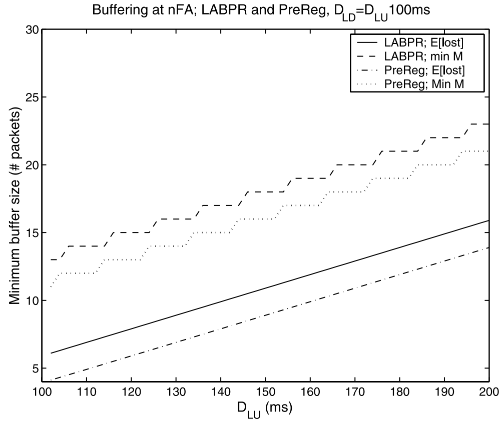

Fig. 6 Buffering in nFA — LABPR and PreReg.

influenced by AT and L2LU spacing.

We can see the advantage of using LABPR if we interpret Fig. 6 as follows: suppose we take the 45 ms RTT estimate for delivering a trigger. DLU would subsequently occur at 145 ms, showing a buffer requirement of 10 packets, which is exactly what is to be expected given the UDP stream specifications. In the figure, the expected number of lost packets in the nFA drops even below this value as packets are also lost in the oFA when L2LD occurs too soon after AT. For Pre-Registration to function correctly, the handoff will most likely be initiated at a very safe margin, causing a higher traffic load to be sent to the nFA, yielding larger buffers.

5. Performance Evaluation: OPNET Simulation

5.1 Model Description

OPNET Modeler is a discrete event simulator developed by OPNET Technologies, Inc [12]. It has a functionally very rich model library which we have modified and extended [18] to provide support for: MIP, HMIP, Smooth Handoff, Optimized Smooth Handoff, Pre- and Post-Registration and now LABPR as well [19]. For a detailed explanation of the OPNET models and more results, we refer to Ref. 19).

5.2 Simulation Results

First we present simulation results for a tagged UDP traffic stream. Using a network similar to the one depicted Fig. 2, we send a flow of UDP packets to the MN, who is performing an LABPR handover between the two networks. The re-association delay with the new access point is uniformly distributed between 200 and 400 ms, values assumed to be reason-

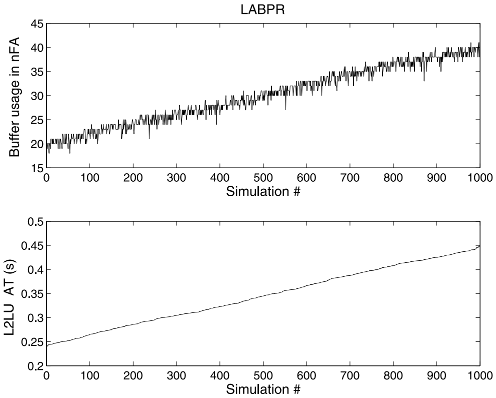

Fig. 7 LABPR buffering.

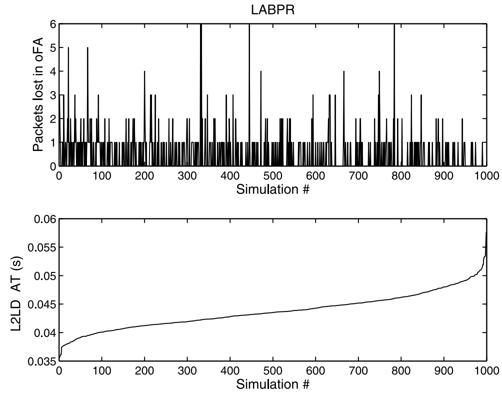

Fig. 8 LABPR oFA packet loss.

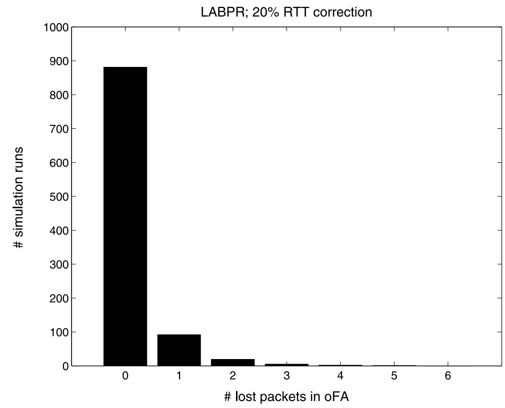

Fig. 9 LABPR oFA losses, 20% correction.

able according to Ref. 10). For LABPR, the timing between AT and L2LD is determined by the RTT measurements, whereas for Pre-Registration it is taken from a uniform distribution between 10 and 150 ms (as we do not know beforehand how big it should be). If we do a thousand simulation runs, we can look at the produced trigger values and the corresponding effect on nFA buffer usage and oFA packet loss. The following figures are to be interpreted as follows: the lower half shows the observed lengths of the time intervals defined by 2 triggers, sorted on increasing value. The upper half then shows the produced results for these trigger occurrences.

The effect of the length of the interval [AT, L2LU] on buffer requirements is displayed in **Fig. 7**. The behavior is very similar to Pre-Registration. While the spacing of the AT and L2LD has an impact (the further they are spread apart, the larger the number of packets arriving in the buffer), the dominating factor here is the length of [oFA L2LD, nFA L2LU] (it is the largest component). The relation here is fairly linear, and the jitter observed on the buffer usage graph is due to the background load present in the system.

Packet loss in the oFA for Pre-Registration is very unpredictable, as it again suffers from the lack of trigger scheduling. **Figure 8** shows the observed packet loss for LABPR, and the lower half of this figure in fact displays the measured values of the RTT estimator. The losses measured here are fairly low and consistent, but they are still there due to the unexpected delay a packet can encounter.

On first sight, it seems that LABPR's performance here is fairly bad, considering we can get to nearly zero losses in Pre-Registration if there would be a way to deliver the anticipation trigger in time. When using the average of ping measurements for the AT value, the performance of LABPR is less than optimal. About 700 runs (out of 1,000) have lossless performance and about 25% of the handovers loose 1 packet in the execution. An obvious improvement to this protocol would be to add a correction factor to this RTT estimate, shown in **Fig. 9** where 20% of its value is added to deliver a more robust scheme. Now we can see that slightly less than 90% of the runs proceed without any losses at all and the single packet drops are now below 10%.

6. Conclusions

In this paper we discussed the problems related to the current low latency handoff schemes. These rely on the presence and timing of layer 2 triggers, an obstacle in many cases. We proposed an extension to Pre-Registration that breaks the strict dependency on lower layers if certain location information is available.

We presented an analytical model that shows an improvement for the delay components involved in MIP. Not only are there fewer packets involved in the handoff, but packet loss and buffer usage are more predictable. The probability of packet loss in the oFA is lower and so is the number of packets lost. The OPNET simulations confirmed these results in a more complex environment and showed that increasing the RTT estimator by a certain amount can have a big impact on lowering packet loss. We now have a way to influence packet losses in the oFA: if we want very low latency, the chances of losing packets will be higher and conversely, if lossless operation is required then the triggers will need to be spaced more resulting in slightly higher buffering requirements and overall latency.

References

1) Blondia, C., Casals, O., De Cleyn, P. and Willems, G.: Performance analysis of IP micromobility handoff protocols, *Proc. Protocols for High Speed Networks 2002* (*PfHSN 2002*), Berlin, pp.211–226 (2002).
2) Blondia, C., Casals, O., Cerda, Ll. and Willems, G.: Performance analysis of a forwarding scheme for handoff in HAWAII, *Proc. Networking 2002*, Pisa, pp.504–514 (2002).
3) Blondia, C., Casals, O., Van den Wijngaert, N. and Willems, G.: Performance analysis of smooth handoff in Mobile IP, *Proc. Fifth ACM International Workshop on Modeling, Analysis and Simulation of Wireless and Mobile Systems* (*MSWiM2002*), Atlanta, USA, pp.22–29 (2002).
4) Blondia, C., Casals, O., Cerda, Ll., Van den Wijngaert, N., Willems, G. and De Cleyn, P.: Performance comparison of low latency Mobile IP schemes, *Proc. WiOpt03*, INRIA Sophia Antipolis, France, pp.115–124 (2003).
5) Blondia, C., Casals, O., Cerda, Ll., Van den Wijngaert, N., Willems, G. and De Cleyn, P.: Low latency handoff mechanisms and their implementation in an IEEE 802.11 network, *Proc. ITC 2003* (to appear).
6) Campbell, A.T., Gomez, J., Wan, C.-Y., Kim, S., Turanyi, Z. and Valko, A.: Cellular IP, draft-ietfmobileip-cellular-00-txt (1999).
7) El Malki, K., et al.: Low latency handoffs in Mobile IPv4, draft-ietf-mobileip-lowlatency-handoffs-v4-04.txt (2002).
8) Ergen, M., et al.: Application of GPS to Mobile IP and routing in wireless, *IEEE VTC*, Vancouver, Canada (2002).
9) Gustafsson, E., Jonsson, A. and Perkins, C.: Mobile IP regional registration, draft-ietf-mobileip-regtunnel-06.txt (2002).
10) Mishra, A., Shin, M. and Arbaugh, W.: An empirical analysis of the IEEE 802.11 MAC layer handoff process, Technical Report No.CS-TR-4395, Department of Computer Science, University of Maryland (2002).
11) Ohta, M.: Smooth handover over IEEE 802.11 wireless LAN, draft-ohta-smooth-hand-overwlan-00.txt (2002).
12) OPNET Modeler, www.opnet.com.
13) Perkins, C.: IP mobility support, *rfc 2002* (1996).
14) Perkins, C. and Johnson, D.: Route optimization in Mobile IP, draft-ietf-mobileipoptim-11.txt.
15) Perkins, C. and Wang K.-Y.: Optimized smooth handoffs in Mobile IP, *Proc.IEEE Symposium on Computers and Communications*, Egypt (1999).
16) Ramjee, R., La Porta, T., Thuel, S., Varadhan, K. and Salgarelli, L.: IP micro-mobility support using HAWAII, draft-ietf-mobileip-hawaii-00.txt (1999).
17) ANSI/IEEE Std 802.11, 1999 Edition.
18) Van den Wijngaert, N. and Blondia, C.: Integration of IP mobility in OPNET: modeling and simulation, *Proc. OPNETWORK 2002*, Washington D.C., USA (2002).
19) Van den Wijngaert, N. and Blondia, C.: A location augmented low latency handoff mechanism for Mobile IP: modeling and simulation, *Proc. OPNETWORK 2003*, Washington D.C., USA (2003).

(Received April 10, 2004)
(Accepted July 1, 2004)

Nik van den Wijngaert obtained his Masters in Mathematics and Computer Science from the University of Antwerp in 2001, where he since then has been pursuing his Ph.D. in the PATS research group. His main interests are protocol modeling methodologies, seamless mobility and performance analysis. He has published a number of papers in international journals and conference proceedings.

Chris Blondia obtained his Ph.D. in Mathematics from the University of Ghent (Belgium) in 1982. Between 1986 and 1991 he was a researcher at the Philips Research Laboratory and from 1991 till 1994 he was an associate professor at the Computer Science Department of the University of Nijmegen (The Netherlands). In 1995 he joined the Department of Mathematics and Computer Science of the University of Antwerp, where he is currently a professor, head of the research group "Performance Analysis of Telecommunication Systems" (PATS) and head of the Department. His main research interests are related to both theoretical methods for stochastic modelling (in particular queueing systems) and to the design, the modelling and performance evaluation of telecommunication systems, in particular related to traffic management in broadband networks, IP mobility management, medium access control for wireless and wired access networks, etc. He has published a substantial number of papers in international journals and conference proceedings. He is editor of the "Journal of Network and Computer Applications" and member of IFIP W.G. 6.3 on "Performance of Computer Networks". He is and has been involved in several European research programs (RACE, ACTS, IST and COST).