Low Latency Handoff Mechanisms and Their Implementation in an IEEE 802.11 Network.

C. Blondia\textsuperscript{a}, O. Casals\textsuperscript{b}, I. Cerdà\textsuperscript{b}, N. Van den Wijngaert\textsuperscript{a}, G. Willems\textsuperscript{a} and P. De Cleyn\textsuperscript{a}

\textsuperscript{a}University of Antwerp, Dept. Mathematics and Computer Science, Universiteitsplein 1, B-2610 Antwerp - Belgium\textsuperscript{*}

\textsuperscript{b}Technical University of Catalonia (UPC), Dept. Computer Architecture, Jordi Girona 1-3, Mdl. D6, E-08034 Barcelona - Spain\textsuperscript{b}

In this paper we compare the performance of two low latency handoff protocols for MIPv4, Pre- and Post-Registration Handoff. These mechanisms proposed by the IETF aim at improving the performance of Hierarchical Mobile IP with respect to handoff latency and packet loss. We propose an analytical model to study the influence of various system parameters on the performance of the two protocols, followed by a comparison of the two schemes. We describe several handoff implementations over a wireless access based on the IEEE 802.11 standard and analyze them by means of an ns simulation.

1. INTRODUCTION

Mobile IP [1], the current support of mobility in IP networks, allows node mobility across media of similar or dissimilar types and delivers packets to a temporary address assigned to the mobile host at its current point of attachment. This temporary address is communicated to a possibly distant Home Agent (HA). This approach applied to an environment with frequent handoffs may lead to high associated signaling load and unacceptable disturbance to ongoing sessions in terms of handoff latency and packet losses. In such an environment, low-latency handoffs are essential to avoid performance degradation and high signaling overhead.

Therefore, a hierarchical mobility management approach has been proposed where Mobile IP supports wide area mobility (e.g., mobility between different operators) while local mobility is handled by more optimized micro-mobility protocols. In that way a home agent does not need to be aware of host movements within an access network, as a Mobile Node (MN) will keep the same address until it moves to another access network, in which case it will have to report it to his home agent.

A number of micro-mobility protocols have been discussed in the IETF [2]. This paper deals with Hierarchical Mobile IP (MIP) [3]. MIP is a hierarchical tunneling technique

\textsuperscript{*}This work was supported by the Ministry of Ed. of Spain under grant CICYT TIC-2001-0956-004-01 and by the Dep. of Universities (Generalitat de Catalunya) under grant CIRIT 2001-SGR-00226, and by the Fund for Scientific Research Flanders G.0315.01 and by DWTC Belgium IAP P5/11 MOTION.
that relies on a tree-like structure of foreign agents. Traffic destined to the MN is encapsulated at the home agent and then delivered to the root foreign agent. Each foreign agent in the tree decapsulates and then re-encapsulates the packets as they are forwarded down the tree of foreign agents towards the MN. Location updates at the adequate points of the tree follow the MN movements between different base stations so that traffic is tunneled to the MN's point of attachment.

The aim of the Pre- and Post-Registration Handoff schemes proposed in [4] is to achieve low latency Mobile IP handoffs by minimizing the period of time that an MN is unable to send or receive IP packets due to the delay in the Mobile IP registration process. These methods can support both the normal Mobile IP model [1] in which the MN is receiving packets from a HA and the Hierarchical Mobile IP model [3] in which the MN receives packets from a Gateway Foreign Agent (GFA).

This paper focuses on the performance evaluation of these handoff schemes, their comparison and possible implementations over IEEE 802.11 as link layer. In section 3 we propose an analytical model that allows computing characteristic performance measures of the handoff schemes. These measures are related to packet loss and experienced delay. The models are not developed for dimensioning purposes, but mainly to investigate the influence of important design parameters and to compare the solutions. For this reason we have assumed Poisson background traffic and exponential processing times. The simplicity of the model also allows the study of more general network topologies than the one considered in this paper. In section 4 we describe possible implementations of the low latency handoff mechanisms over an IEEE 802.11 network and analyze their performance using an ns simulation. Finally section 5 concludes the paper.

2. DESCRIPTION OF LOW LATENCY HANODF MECHANISMS: PRE AND POST REGISTRATION

Mobile IP was originally designed without assumptions concerning the underlying link layer (L2). This approach implies a clear separation between L2 and L3 functionality but may lead to unacceptable handoff latencies. Indeed, an MN involved in a handoff may only begin the registration process after the L2 handoff to the new Foreign Agent (nFA) has been completed. Moreover, as the messages generated by the registration process need some time to propagate through the network, the MN is unable to send or receive packets during that time. For these reasons, the Mobile IP workgroup of the IETF has proposed so-called low latency handoff schemes [4] based on information on the L2 handoff process received using L2 triggers [5]. Two types of schemes may be distinguished. A first type of scheme allows the MN to communicate with the new FA while still being connected with the old FA. The Pre-Registration scheme belongs to this class. In a second type, the packets can be delivered to the MN at the new FA before the registration process has completed. The Post-Registration scheme belongs to this class. In what follows, we give a description of both schemes. Details can be found in [4].

2.1. The Pre-Registration Handoff scheme

In Pre-Registration, the network assists the MN in performing an L3 handoff before the L2 handoff is completed. Both the MN (mobile-initiated) and the FAs (network-initiated) can initiate a handoff. A mobile-initiated handoff occurs when an L2 trigger is received
at the MN informing it that it will shortly move to the nFA. A network-initiated handoff can be initiated by a source trigger at the oFA (source-initiated handoff) or by a target trigger at the nFA (target-initiated handoff). The L2 trigger contains information such as the nFA’s IP address identifier.

The following messages are involved: (i) First, the oFA should solicit and cache advertisements from the nFA in advance of the Pre-Registration Handoff. (ii) As a consequence of the L2 trigger the MN sends a Proxy Router solicitation to oFA, which in turn sends a Proxy Router Advertisement. (iii) Then the MN should send a registration request to the nFA via the oFA if the L2 handoff is not completed, or directly to the nFA if the L2 handoff is finished. (iv) Upon receiving the registration request the nFA sends a Regional Registration Request to the GFA, which in turn sends a Regional Registration Reply.

Until the MN actually completes the L2 handoff to the new FA and fully establishes the new L2 link, the nFA can receive packets to which it does not have a direct link layer connection. In that case the nFA can decide to drop or to buffer the packets for the MN.

2.2. The Post-Registration Handoff scheme

The Post-Registration Handoff method is based on a network-initiated model of handoff. It does not require any MN involvement until the actual L2 connection with the nFA is completed. The name of this technique finds its origin in the fact that the registration occurs after the L2 handoff has been completed. This approach uses bi-directional edge tunnels (BETs) to perform low latency change in the L2 point of attachment of the MN without requiring any involvement of it.

A handoff occurs when the MN moves from the oFA, where the MN performed a Mobile IP registration, to the nFA. Instead of making a new Mobile IP registration with the nFA, the MN delays it while maintaining connectivity using the BET between the oFA and nFA. In [4], two different Post Registration handoff schemes are defined: Source and Target Trigger Post Registration.

An FA becomes aware that a handoff is about to occur at L2 through the use of an L2 trigger. Two types of triggers can be received: (i) a source trigger at the oFA (L2-ST) and (ii) a target trigger at the nFA. (L2-TT). The FA receiving the trigger sends a Handoff Request (HRqst) to the other FA. The FA receiving the HRqst sends a Handoff Reply (HRply) to the other FA. This establishes a BET. The L2-LD (Link Down) trigger at the oFA and at the MN signals that the MN is not connected anymore with the oFA. When the oFA receives the L2-LD trigger, it begins forwarding the MN packets through the forwarding tunnel to the nFA.

When the nFA receives the L2-LU (Link Up) trigger, it begins delivering packets tunneled from the oFA to the MN and forwards packets from the MN. When the MN receives the L2-LU, it decides to initiate the Mobile IP Registration process with the nFA by soliciting an Agent Advertisement or continues using the BET. Once the Registration process is complete (through the exchange of a Regional Registration Request and a Regional Registration Reply with the GFA), the nFA takes over the role of oFA.

3. PERFORMANCE MODELING OF LOW LATENCY MECHANISMS

In this section we present a mathematical model for the low latency handoff schemes based on a queuing network, similar as in [6-10]. We use the model to compare both
schemes assuming that the L2 triggers are available from a theoretical L2 access protocol. The model has been validated in these contributions.

3.1. The analytical model

Consider the network architecture depicted in Figure 1. For computational tractability reasons all routers are modeled as simple M/M/1 queues. The exponentially distributed service time of a packet in each router is assumed to both include the processing time and the transmission time. If we denote the service rate of any router \( A \) by \( \mu \) and the load by \( \rho \), then its response time \( R_A \) is exponentially distributed with rate \( \mu (1 - \rho) \). Here we will focus on the Pre-Registration case, and refer to [9] for the corresponding model of the Post-Registration Handoff scheme.

Consider an MN moving from the oFA to the nFA, and suppose an overlapping area between the two subnetworks. We assume the L2 handoff starts when the MN enters the overlapping area, and denote this time instant by \( t_0 \). Furthermore a source-initiated handoff is assumed (initiation by an L2-ST trigger), the oFA receives an LD trigger when the connection with the MN is lost and the nFA receives an LU trigger when the L2 connection is fully established. We define the variables \( D_{ST} \), \( D_{LD} \) and \( D_{LU} \) as the time needed, since \( t_0 \), to generate the L2-ST, L2-LD and L2-LU trigger respectively. They are considered constant positive values and we have that \( D_{ST} < D_{LD} < D_{LU} \).

The time instant the GFA starts forwarding packets with destination the MN to the nFA instead of the oFA is denoted by \( t_1 \) and equals \( t_1 = D_{ST} + R_oFA + R_3 + R_nFA + R_2 + R_{GFA} + \) fixed delays. This expression is a sum of exponentially distributed variables and constants. Packets will be routed via the oFA or via the nFA according to whether they arrive at the GFA before or after \( t_1 \).

Now consider a constant bit rate UDP stream of packets originating from a Corresponding Node (CN) destined to the MN. Assume that every \( T \) ms a packet arrives at the GFA. Then each packet of that stream belongs to exactly one of the following classes: Class 0: packets arriving at the oFA before \( t_0 + D_{LD} \); these packets are forwarded directly to the MN. Class 1: packets arriving at the oFA after \( t_0 + D_{LD} \); these packets are lost. Class 2: packets arriving at the nFA before \( t_0 + D_{LU} \); these packets are lost if not buffered at the
nFA. Class 3: packets arriving at the nFA before \( t_0 + D_{LU} \); these packets are forwarded directly to the MN.

In our model the path each packet follows is the sum of some exponential random variables and constants. Hence the delay distribution of each packet can be computed in a straightforward way. Other performance measures such as the loss probability of a packet and the expected total number of lost packets during a handoff follow directly from this.

An interesting parameter now is the size of the buffers that should be installed at the FAs to ensure practically zero packet loss. For Pre-Registration this concerns only the nFA, where packets possibly need to be buffered awaiting the LU trigger. In Post-Registration buffers can be installed at both oFA and nFA, respectively for packets that need to wait for the establishment of the BET and for packets that arrive at the nFA before the LU trigger.

In order to determine the buffer size that is needed to achieve a sufficiently low packet loss probability, we have to compute the distribution of \( N_b \), by which we denote the number of packets that would be lost if there would not be a buffer installed at a specific FA. If we denote by \( R_{loss}(M) \) the probability that at least 1 packet will be lost at the FA with buffer capacity \( M \), then we have \( R_{loss}(M) = P(N_b > M) \). We can approximate this probability by conditioning on the length of the specific time interval in which the arrival of a packet means that this packet needs to be buffered.

As an example, consider the packets that need to be buffered at the nFA in case of the Pre-Registration scheme. The interval in which they arrive is given by \([t_1 + R_{GFA} + R_2 + R_{nFA} + \text{fixed delays}; D_{LU}]\). The length of the interval is denoted by \( I_b \) and has a distribution composed of sums and differences of exponential variables and constants. It is straightforward to compute an approximation to the distribution of \( N_b \) conditioned on \( I_b \), so we can use \( R_{loss}(M) \approx \sum_{i=1}^{N} P(N_b > M|I_{b,i}) P(I_{b,i}) \), where \( I_{b,i}, i=1,\cdots,N \) is some discretization for \( I_b \). We can then determine the required buffer size by \( \min M : R_{loss} < 10^{-\alpha} \), where \( \alpha = 5 \).

### 3.2. Performance evaluation using the analytical model

We assume a network topology as depicted in Figure 1 and we consider a CN that transmits packets every \( T = 10 \) ms. The propagation delays on the links connecting the GFA and the oFA, as well as the links connecting the nFA are all set to \( \tau_1 = 5 \) ms, while the links connecting the oFA and the nFA have a propagation delay of \( \tau_2 = 3 \) ms. The service rate \( \mu \) in each router is set to 1 packet/ms. All routers have a load of 0.8.

First we present some results for the delay distribution of a stream of packets involved in a handoff. Here we assume there is no capacity to buffer the incoming packets at the FAs. This means that if packets arrive at the oFA before the BET (Post-Registration) or at the nFA before the LU trigger (Post- and Pre-Registration), they are lost. The playout time is the maximum allowed end-to-end delay: if a packet’s end-to-end delay exceeds this playout time, it will be dropped and thus will also be lost. The start of the handoff is set to \( t_0 = 0 \) and we consider a stream of 30 packets, the first of which is transmitted by the GFA at \( t = -80 \) ms.

Figure 2 shows the expected number of packets from this stream that are dropped due to expiration of the playout time or the absence of a buffer, as a function of this playout
time. Both the results for Pre-Registration and Post-Registration are shown, for two different values of the time between the LD and the LU trigger. The timing of the LD trigger is set \( t_0 = 60 \text{ ms} \). Note that these curves tend to the expected number of lost packets due to the absence of buffer capacity when the playout time tends to infinity. It can be seen that Pre-Registration implies more losses than Post-Registration, while the average delay for packets that are not lost is slightly larger for the Post-Registration scheme. The latter follows from the fact that packets using the BET have a longer delay. When the time between the LD and the LU trigger increases, more packets are lost, so more buffer capacity would be needed to avoid losses.

Now we proceed with results on the required buffer capacity as a function of the timing of the L2 triggers. Figure 3 and Figure 4 each depict both the expected number of packets that need to be buffered at the oFA or nFA (thick line), as well as the minimum buffer size \( M \) to ensure that \( P_{\text{loss}}(M) < 10^{-5} \) (thin line). Each figure compares Pre-Registration (solid line) to Post-Registration (dotted line).

Figure 3 concerns packet loss and buffering at the oFA, and here the comparison does not entirely hold, because there is no benefit at installing a buffer for Pre-Registration since packets arriving after the LD trigger are simply lost and will not be forwarded. Therefore, for Pre-Registration, \( M \) is to be interpreted as the minimum number for which it holds that the probability that there are more packets lost is smaller than \( 10^{-5} \). Since the timing of the LU trigger does not influence losses at the oFA, \( D_{LU} \) was fixed at 50 ms, while \( D_{LD} \) varies. As \( D_{LD} \) increases, packet loss at the oFA of course diminishes. However, assuming exponential service rates still brings about that a relatively large buffer size is needed to ensure practically zero losses.

Finally Figure 4 presents results for buffering at the nFA, now as a function of \( D_{LU} \). The value of \( D_{LD} \) is fixed at \( D_{LU} - 40 \text{ ms} \). As can be seen from Figure 4 the required buffer capacity for Post-Registration mainly depends on the time between the triggers. The fact that for smaller values of \( D_{LU} \) less packets need to be buffered is explained by noting that in these cases less packets are transmitted through the BET because the BET might not yet be established at the time of \( D_{LU} \).
4. HANDOFF IMPLEMENTATIONS OVER IEEE 802.11

In the IEEE 802.11 standard [11] the base stations used by the MN to access the fixed part of the network are referred to as Access Points (AP). Before an MN is allowed to transfer data packets to an AP, it has to be associated with it. The MN initiates the association by sending an Association Request frame, which, in turn, is answered by an Association Response frame by the AP. The MN can only be associated with one AP. If the MN decides to handoff to another AP, then it sends a Re-association Request to the new AP. The MNs can use L2-beacons sent by the APs to determine which AP would make the best connection, and thus, the Association and Re-association Requests are sent to this AP. In the following we describe how these L2 packets could be used as L2-triggers in order to implement the protocols described in previous sections.

We have implemented these protocols using the network simulator, ns [12]. In the ns version given in [12] there is a MIP implementation where the handoffs are completely managed at layer 3. The implementation consists of the FA sending Router Advertisements that are used by the MNs to decide when to handoff to a new FA. In order to have L2-trigger support, we added the beacon and association mechanisms to the 802.11 protocol implemented in ns. This allowed us to extend ns by adding MIP with Pre-Registration and MIP with Post-Registration. In the next sections we describe our implementation of these protocols giving some simulation results. In our simulations we have assumed the following:

- The Foreign Agent (FA) is an embedded entity in the Access Point (AP). We shall refer to this system indistinctively as AP or FA.

- The MN is in coverage with both FA during an overlapping time. Note that this time depends on the mobile speed and the AP range.

- We assume that the MN uses the same channel with both AP (thus, the MN can listen both AP during the overlapping region) and movement at L2 layer is detected upon receiving the first beacon from the nAP. 802.11 beacons are sent every 100 ms.

- Unsolicited router advertisements are sent every 1 s.

- 802.11 beacons and unsolicited router advertisements are sent uniformly distributed regarding the overlapping time.

- We use the network topology shown in Figure 1. All links have a transmission rate of 4 Mbps. All wired links have exponential background traffic in both directions with $\rho = 0.8$.

- The CN sends CBR packets (of size 100 bytes) with a period of 10 ms.

4.1. Pre-Registration Handoff

As all 802.11 handoffs are mobile initiated, the network-initiated Pre-Registration handoff is not applicable to 802.11 [13]. The MN sends a Proxy Router Solicitation to the oFA once it has chosen a new AP. The oFA maps the nFA link layer address into the IP address of the nFA (assuming the oFA maintains a mapping table) and returns a Proxy Router
Advertisement to the MN. The MN sends a Registration Request message to the nFA through the oFA since the MN is not yet connected to the nFA. The nFA will forward the Registration Request message to the GFA. At this point the MN should complete the L2 handoff; this means exchanging Re-association messages. The Registration Reply message will be unicast by the nFA to the MN on-link as soon as the MN connects to the nFA [4]. If the Registration is successful then packets for the MN will be tunneled from the GFA to the nFA where the MN has moved.

The events during the handoff in our implementation are the following: (i) The 802.11 beacon from nFA triggers the MN sending the Proxy Router Solicitation to oFA. (ii) The oFA sends the Proxy Router Advertisement. (iii) Upon receiving the Proxy Router Advertisement the MN sends a Registration Request to oFA, with destination address the nFA. Then the MN sends the Re-association to nAP. (iv) Upon receiving the Re-association reply from the nFA, the MN also sends a Registration Request along nFA, just in case the Proxy Router Solicitation is lost.

Note that losses can occur when the Pre-Registration fails due to loss of coverage with the oFA. Now the MN sends the Proxy Router Solicitation to oFA when it is already out of coverage. In our ns implementation the MN Re-associates with the nFA if no answer (the Proxy Router Advertisement) is received from the oFA. However, this Re-association is delayed approximately 50 ms because 802.11 retransmits several times the Proxy Router Solicitation to the oAP waiting for the acknowledgment. Upon reception of the Re-association Reply, the MN uses the standard MIPv4 registration with the nFA.

4.2. Post-Registration Handoff

In this scheme the MN re-associates with the nFA once it has chosen a new AP. This re-association message will act as an L2 target trigger [13] at the nFA. The nFA and oFA will establish a bi-directional edge tunnel (BET) after the exchange of a handoff request and a handoff reply message. At this point the BET is established and traffic is tunneled between the two FAs so that the MN continues to receive service through the BET without being registered with the nFA. At some future time instant the MN will register with the nFA.

The events during the handoff in our implementation are the following: (i) Upon reception of the 802.11 beacon from nFA the MN sends the Re-association Request to nAP. (ii) The Re-association Request triggers the nFA to send a Handoff Request to oFA. (iii) When the oFA receives the Handoff Request it sends back a Handoff Reply. At this point, the oFA sends the packets addressed to the MN through the BET. (iv) When the MN receives an unsolicited Router Advertisement from the nFA, the MN initiates the normal MIP registration with nFA.

Losses can occur if the handoff is initiated when the connection with the oAP has been lost.

4.3. Comparison using CBR sources

For sake of comparison, Figure 5 shows the expected number of packets lost during a handoff for the different handoff protocols varying the overlapping period. The average shown in this figure has been obtained repeating the handoff for each point between 100 and 200 times until obtaining a reasonable confidence interval. The confidence intervals are also shown in the figure.
This figure illustrates the benefits of using L2-triggers: the losses without L2-triggers are much higher than in the schemes based on L2 triggers. This is because of the difference between the time period between L2-beacons (100 ms) and routers advertisements (1 s).

Confronting the schemes based in L2-triggers, Figure 5 shows that the best performance is obtained using the Post-Registration scheme. This is because the BET established between the oFA and nFA is able to save some packets arriving to oFA when the MN has already moved out of coverage. However, the Pre-Registration scheme achieves worse performance than the simpler MIP based on triggers. This is because in those handoffs were the MN was moved out of coverage with the oFA, trying to pre-register through the oFA is useless and increases the handoff latency.

4.4. Comparison using TCP sources

In this section we have used the same simulation environment as in the previous sections, but using TCP sources. The simulation experiment consists of analyzing the impact of a handoff over a TCP connection that generates 80 different segments. These segments are transmitted from the CN to the MN using IP packets of 500 bytes each. During the TCP transmission a handoff occurs. As in the CBR evaluation of section 4.3, the simulation was repeated for each point between 100 and 200 times until obtaining a reasonable confidence interval.

Figure 6 shows the expected normalized goodput. This measure is obtained as follows: the time interval between the first and the last segment received at the MN (tctime) is measured. Then, the goodput in bps is estimated as goodput = 79-500·8/tctime. Finally, since the links are 4 Mbps and have an 80 % load, the normalized goodput is computed as goodput·100/(8·105).

We observed that the Post-Registration always gives the lowest number of packets lost for the same reason as explained in section 4.3. However, this loss reduction does not always translate into a goodput increase. This occurs because of the negative influence on TCP of the delay variations introduced on the TCP flow when the tunnel is established and removed.
5. CONCLUSIONS

In this paper we have analyzed the Low Latency Handoff methods based on L2-triggers proposed by the IETF by means of an analytical model. Furthermore, we have described possible implementations of these protocols in an 802.11 wireless network.

The analytical model has been used to evaluate the performance of Post and Pre-Registration Handoff methods. The evaluation done for constant bit rate real-time (UDP) traffic is characterized by two measures: the expected number of tunneled packets that are dropped due to the expiration of the playout time together with the expected number of packets lost in the oFA and/or nFA depending on the triggers timing. From this analysis it follows that these losses could be avoided by using appropriately dimensioned buffers in both the oFA and the nFA.

The simulation results indicate that: (i) the timing of the triggers has a major impact on the packet loss rate, (ii) there is a clear benefit to using the Low Latency Handoff schemes over MIP, (iii) although Post-Registration achieves the highest loss reduction, this does not always translate to a TCP goodput increase.

REFERENCES

12. ns with Micromobility Support.
   URL http://comet.ctr.columbia.edu/micromobility