

Performance Evaluation of Layer 3 Low Latency Handoff Mechanisms¹

C. Blondia^a, O. Casals^b, Ll. Cerdà^b, N. Van den Wijngaert^a, G. Willems^a

^aUniversity of Antwerp, Dept. Mathematics and Computer Science, Middelheimlaan 1,
B-2020 Antwerp – Belgium, Tel: +32 3 2180903, email: chris.blondia@ua.ac.be

^bTechnical University of Catalonia (UPC), Dept. Computer Architecture, Jordi Girona 1-3,
D6, E-08034 Barcelona–Spain, Tel: + 34 934016985, email: olga@ac.upc.es

Abstract

This paper investigates the performance of two Layer 3 low latency handoff protocols proposed by the IETF, namely Pre- and Post-Registration. These protocols use Layer 2 triggers to reduce the built-in delay components of Mobile IP. We propose a simple analytical model that allows assessing the packet loss and the delay characteristics of these protocols. We describe several handoff implementations over a wireless access based on the IEEE 802.11 standard and analyze several implementation issues. Finally we study the scalability of the protocols using an OPNET simulation.

Keywords Mobile IP, low latency handoff, IEEE 802.11, performance evaluation.

1. Introduction

The growing number of portable computing devices and the requirement to provide seamless connectivity to the global Internet using end-to-end IP solutions for mobile users have stimulated the research into IP mobility protocols. Mobile IP ([17]) is the current standard solution for mobility management in IP networks. It allows a Mobile Node (MN) to change its point of attachment from one access router to another across media of similar or dissimilar types with the help of Mobility Agents. Establishment of new tunnels can introduce considerable delays in the handoff process due to the round-trip time between the Foreign Agent (FA) and the Home Agent (HA) during the registration process. Applied in an environment with frequent handoffs, this may lead to unacceptable disturbance to ongoing sessions in terms of handoff latency and packet loss.

¹ This work was supported by the Ministry of Education of Spain under grant CICYT TIC-2001-0956-C04-01 and by the Department of Universities, Research and Information Society (Generalitat de Catalunya) under grant CIRIT 2001-SGR-00226, by the Fund for Scientific Research Flanders under project G.0315.01, by DWTC Belgium under project IAP P5/11 MOTION (Mobile Multimedia Communication Systems and Networks) and by IWT under project 020152 – End-to-end QoS in IP based Mobile Networks.

The IETF has proposed a number of protocols that handle local movements locally. These so-called micro-mobility protocols incorporate a number of important design features related to location management, routing and handoff schemes. They should fulfill requirements such as simplicity to implement, scalability with respect to the induced signaling, efficiency and performance with respect to packet loss and introduced delay.

Two approaches to hierarchical mobility management have been proposed [8]:

(i) Host-based routing schemes like Cellular IP (CIP) [6] and HAWAII [20] use mobile-specific routing without tunneling to route packets inside the access network. We have evaluated these mechanisms in [1], [2] and [22].

(ii) Hierarchical tunneling techniques like Hierarchical Mobile IP (HMIP) [12] rely on a tree-like structure of FAs. Traffic destined to the MN is encapsulated at the HA and then delivered to the root FA. Each FA in the tree decapsulates and then re-encapsulates the packets as they are forwarded down the tree of FAs 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.

These protocols are designed without any assumption regarding underlying layers. This allows the widest possible applicability and a clean separation between Layer 2 and Layer 3 of the protocol stack. However, this layer separation results in lower performance. Indeed, the MN may only communicate with a directly connected FA and therefore it can only start the registration process after completion of the L2 handoff. Moreover, the MN is unreachable during the registration process, a property that may contribute to a non-negligible handoff latency and packet loss.

In this paper we analyze two mobility protocols [10] that aim at low latency Layer 3 handoff based on Layer 2 information received by means of L2 triggers. In Pre-Registration, the MN communicates with the new FA through the old FA for registration purposes. In the second type, Post-Registration, the data is delivered to the MN by the new FA, even before the actual registration process has been completed. These methods can support both the normal Mobile IP model [17] in which the MN is receiving packets from a HA and the Hierarchical Mobile IP model [11] in which the MN receives packets from a Gateway Foreign Agent (GFA).

This paper focuses on the performance evaluation of these schemes and their comparison. We describe possible implementations of the low latency handoff mechanisms over an IEEE 802.11 network and analyze various implementation issues.

We present a simple analytical model that allows a detailed analysis of the delay characteristics and the buffer requirements for a single MN involved in a handoff. In particular we investigate the influence of the timing of the L2 triggers on the delay and the packet loss of UDP streams towards an MN that is involved in a handoff. In order to investigate the scalability of the protocols, we have developed an OPNET simulation model that allows considering more realistic systems with respect to the number of MNs and their mobility pattern.

2. Low Latency Handoff Mechanisms: Pre- and Post-Registration

Mobile IP [17] defines how a MN performs a Layer 3 handoff between subnets served by different FAs. In order to keep a clear separation between L2 and L3 functionality and to allow for the widest possible applicability, Mobile IP has been designed without any assumption concerning the underlying link layer. This separation between layers results in the following sources of handoff delays. First, an MN involved in a handoff may only begin the registration process after the L2 handoff to the new Foreign Agent has been completed. Secondly, 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. This may lead to a handoff latency that is unacceptable for the support of real-time services. For these reasons, the Mobile IP workgroup of the IETF has proposed so-called low latency handoff schemes [10] based on information of the L2 handoff process received using L2 triggers [23]. Two types of schemes may be distinguished, depending on the above mentioned problem they address. A first type 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 [10]. First we introduce the reference architecture that is used throughout the paper.

2.1 The Reference Architecture

Consider the network configuration depicted in Figure 1. The access network is connected to the Internet via a gateway router. The two access points (AP) belong to different subnets with access routers referred to as respectively the old access router and the new access router. The APs may or may not be co-located with the access routers. The access routers and the

gateway router are provided with mobility agents, respectively the old Foreign Agent (oFA), the new Foreign Agent (nFA) and the Gateway Foreign Agent (GFA).

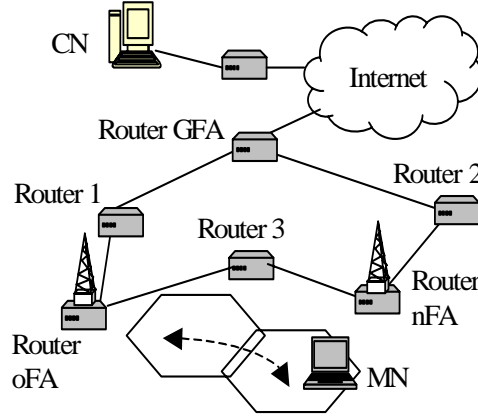


Figure 1: Reference Network

Remark that this reference network has no tree-like structure as encountered in most networks. Router 3 has been added to allow independence between the distance (oFA,nFA) and the distance (nFA,GFA). A more detailed discussion on this issue, together with the impact of the network topology on the performance of the low latency schemes, will be given in Section 5.4.

We assume that a Corresponding Node (CN) sends packets to the Mobile Node. We consider both CBR UDP packet streams and TCP flows.

2.2 L2 Triggers

The handoff schemes investigated in this paper make use of L2 triggers. Such a trigger is a signal related to the L2 handoff process. A first trigger that is used is an early notice of an upcoming change in the L2 point of attachment of the MN, referred to as *anticipation trigger*. A second trigger, the *Line Down trigger* (L2-LD), indicates that the L2 link between the MN and the old AP is lost. The *Line Up trigger* (L2-LU) occurs when the L2 link between the MN and the new AP is established. A trigger initiated at the old FA is referred as a *source trigger* and a trigger initiated at the new FA is referred as a *target trigger*.

For a detailed discussion on these L2 triggers, we refer to [10].

2.3 The Pre-Registration Handoff Scheme

Pre-Registration realizes an anticipated L3 handoff. 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 the L2 anticipation trigger is received at the MN informing it that it will shortly move to the nFA. The L2 trigger contains information such as the nFA's IP address identifier.

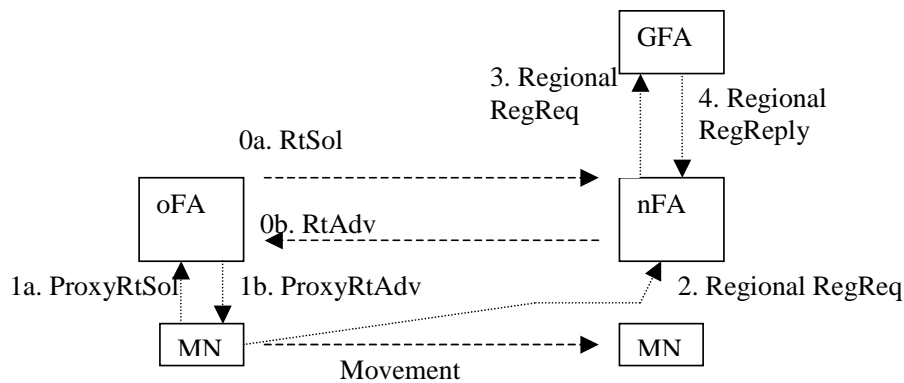


Figure 2: Mobile-initiated Pre-Registration Handoff protocol

The following messages are involved (see Figure 2).

- *Messages 0a and 0b* contain a solicitation for a Router Advertisement from the oFA to the nFA and a reply Router Advertisement from nFA to the oFA. The oFA should solicit and cache advertisements from the nFA in advance of the Pre-Registration handoff in order not to delay the handoff. These messages will not be taken into account in the performance analysis.
- *Message 1a*, a Proxy Router Solicitation, is issued by the MN as a consequence of the L2 anticipation trigger.
- *Message 1b*, a Proxy Router Advertisement, is sent by the oFA as a result of the MN solicitation message.
- *Message 2*, the Registration Request to the nFA should be sent via the oFA if the L2 handoff is not completed (i.e. before L2-LD) or directly to the nFA if the L2 handoff is finished (i.e. after L2-LU).

- *Message 3*, a Regional Registration Request with the GFA, issued by the nFA upon receiving GFA information within message 2 from the MN.
- *Message 4*, the Regional Registration Reply, sent by the GFA to the nFA .

Until the MN actually completes the L2 handoff to the new FA and fully establishes the new L2 link, the nFA may receive packets for the MN 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.

Between the time that the Registration Reply is sent from the GFA to the nFA till the MN's connection on L2 at the nFA is fully established, the GFA can also bicast traffic for the MN to both the oFA and 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).

A *source-initiated handoff* is initiated at the oFA by a received L2 trigger that informs the oFA of a MN's upcoming movement from oFA to nFA. The difference with the mobile-initiated case previously described is that *Message 1a* of Figure 2 is not used.

A *target-initiated handoff* is initiated at the nFA by a received L2 trigger that informs the nFA of a MN's upcoming movement from oFA to nFA. The difference with the mobile-initiated case is that *Message 1a* of Figure 2 is not used and that *Message 1b* is sent by the nFA to the MN via the oFA.

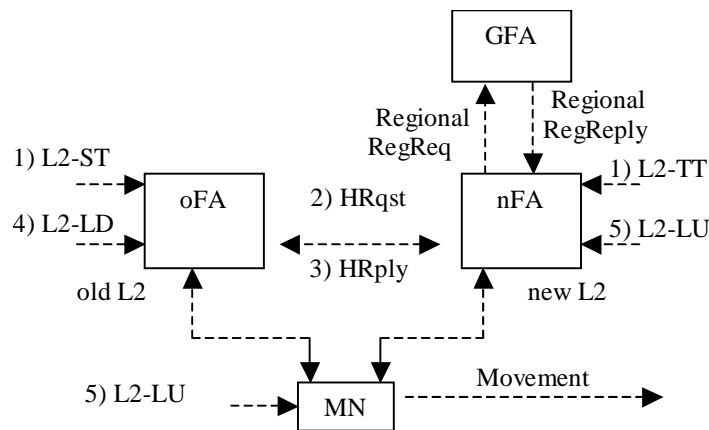


Figure 3. Post-Registration handoff

2.4 The Post-Registration Handoff Scheme

The Post-Registration handoff method is based on a network-initiated model of a handoff which 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 a bi-directional edge tunnel (BET) to perform a 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. The MN delays its registration with the nFA, while maintaining connectivity using the BET between the oFA and nFA. In [10], two different Post Registration handoff schemes are defined: Source and Target Trigger Post Registration. The sequence of messages for both schemes is depicted in Figure 3.

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 assumes the role of oFA.

3. Implementation of Low Latency Handoff Mechanisms Over IEEE 802.11

We describe a possible implementation of the L3 handoff mechanisms described previously in a wireless network having an 802.11 link layer [26]. Our goal is to illustrate by means of a realistic example the abstraction made in previous sections about the L2 triggers and the interaction with the L3 layer protocol. Other authors have also discussed the integration of L3 and 802.11 handoffs, e.g. [9] and [11].

We first introduce the 802.11 signals issued during handoff and we describe how to interpret them as L2 triggers used to implement various L3 handoff mechanisms. Subsequently, we discuss some implementation issues.

3.1 IEEE 802.11 Handoffs

The IEEE 802.11 standard defines two operational modes commonly referred to as *ad hoc* and *infrastructure*. In *ad hoc* mode MNs communicate on a peer-to-peer basis. This mode is usually conceived for a scenario where a group of MNs dynamically set up a network to communicate among them. In *infrastructure* mode all MN transmissions go through an *Access Point* (AP). Usually, APs are not mobile and form part of a wired network.

We assume that the wireless network operates in *infrastructure mode*. In this mode, before an MN can send data packets, it has to be authenticated and associated with an AP. The simplest authentication mechanism is the so-called *Open-system authentication* (the only method required by 802.11). This method is in fact a null authentication mechanism with a two-way message exchange between the MN and the AP. After being authenticated, the MN associates with the AP by sending an Association Request frame (or Re-association if the MN is already associated), which in turn is answered with an Association (or Re-association) Reply. In case of re-association, the Re-Association Request frame conveys the MAC address of the old AP. Another management frame usually involved in 802.11 handoffs is the so-called *beacon*, sent periodically by the APs (typically every 100 ms).

Handoffs are always initiated by the MNs and the standard defines two procedures for the MN to decide which AP to handoff: *active scanning* and *passive scanning*. Using *active scanning*, the MN sends Probe Requests frames on every available channel. The APs answer these frames with Probe Responses that are collected by the MN. Based on the probe responses, the MN selects the AP that would make the best connection. Using *passive scanning* the MN listens for the beacons sent by the APs in order to decide which one to handoff. The event triggering the L2 handoff is implementation specific (and may be initiated by the card firmware using proprietary algorithms).

3.2 Pre- and Post-Registration over IEEE 802.11

The flexibility of the handoff mechanism defined by the 802.11 standard allows a large number of possible scenarios. For sake of simplicity, we shall assume the following: (i) The AP is an embedded entity in the FA such that the L2 triggers can easily be implemented by means of some internal interface. (ii) The MN uses the same channel with all APs. This allows for the following simple way to trigger the L2-handoffs: (iii) The movement at L2 layer is detected upon receiving the first beacon from the new AP. Note that a timeout is needed to avoid ping-ponging between both APs. (iv) We do not take into account the

authentication phase. These assumptions are discussed in section 3.3. The traces shown in this section have been obtained with the ns simulator [24].

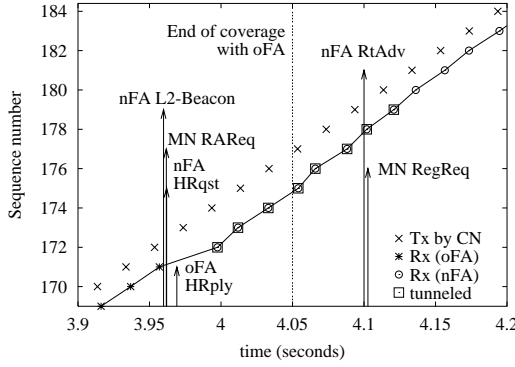


Figure 4: Target Trigger-Post Registration handoff.

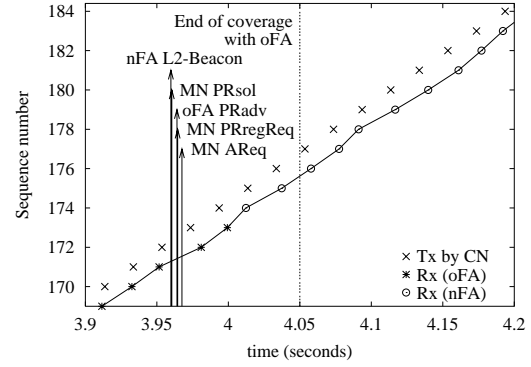


Figure 5: Target Trigger-Pre Registration handoff.

3.2.1 Post-Registration Handoff

Figure 4 shows a trace with a possible handoff using the Post-Registration scheme. In this trace the CN sends a CBR stream to the MN. The figure shows: (i) the instants when the packets are sent (indicated as “Tx by CN”); (ii) the instants when the MN receives the packets (these are indicated as “Rx (oFA)” and “Rx (nFA)” according to which FA they are sent by); (iii) the signaling packets sent during the handoff

When the MN approaches the nFA, the L2 beacons sent by the nAP trigger the L2 handoff at the MN, which sends a *Re-association Request* (RAReq in the figure) to the nAP. Upon receiving this frame, there is a *target trigger* at the nFA (LU trigger), which sends the Handoff Request (HRqst in the figure) to the oFA. Upon receiving the HRqst, the oFA sends the Handoff Reply (HRply in the figure) and establishes a tunnel with the nFA. In this way, the packets can reach the MN via the nFA after the coverage with the oFA has been lost. The oFA starts sending the packets addressed to the MN through the tunnel immediately after it is established. This instant corresponds with the occurrence of the LD trigger introduced in section 2.3. Finally, when the nFA sends the Router Advertisement, the MN makes a Registration with the nFA.

3.2.2 Pre-Registration Handoff

Figure 5 depicts a trace analogous to Figure 4 for the Pre-Registration mechanism. The signaling packets sent during the handoff are the following. The new AP issues L2 beacons

which are interpreted as L2 triggers at the MN indicating the imminent movement to this AP. Then the MN sends the Proxy Router Solicitation (indicated as “MN PRsol”) to oFA. Upon reception of this packet, the oFA sends a Proxy Router Advertisement (indicated as “oFA Pradv”) to the MN. Upon reception of this packet, the MN sends a Registration Request to the oFA with destination address the nFA (indicated as “MN PregReq”). Then, the MN sends the Re-association to the nAP.

A more detailed description of possible implementations of Pre- and Post-Registration handoff mechanisms using IEEE 802.11 can be found in [4] and [5].

3.3 Implementation Issues

The assumptions made in section 3.2 need to be discussed. First, we assumed that the AP is an embedded entity in the FA. In some networks however, the FA and the AP may be different entities. For example, a number of APs distributed over a geographical area may be connected to the same FA. In this case, a change from AP does not imply an L3 handoff. This physical separation between the FA and the AP would prevent the Re-association Request sent by the MN to the new AP to be the target trigger as described in the Post-Registration handoff. One possible solution could be the following. After re-association an L2-source trigger takes place at the MN, which in turn triggers the standard MIP registration process, i.e. the MN broadcasts an Agent Solicitation and sends a Registration Request to nFA upon receiving the Agent Advertisements from the candidate FAs. The MN uses the PFANE (Previous Foreign Agent Notification Extension, see [18]) in the Registration Request. Knowing the Previous Foreign Agent IP address, the nFA sends the Handoff Request to the oFA. This mechanism is equivalent to the *Smooth Handoff* described in [19].

The second assumption (MN uses the same channel with all APs) may also not be satisfied in many situations. This is because it is convenient to allocate adjacent APs in non interfering channels, in order to increase the maximum network load. In fact, a recent study [15] analyzed a real 802.11 network where adjacent APs used different channels. MNs use active scanning to find out new channels and available APs when the 802.11 handoff is triggered. This study shows that the scanning phase is the dominant component in the 802.11 handoff latency and may range from tens to several hundred ms depending on the cards. Note that the MN chooses the AP after the scanning, furthermore, during the scanning the MN cannot send nor receive data packets. Therefore, the Pre-Registration may not be feasible since the connection with the old AP may have been lost due to the long latency required for the 802.11 handoff. In this situation, the Post-Registration could be more effective. Additionally, the oFA

would allow the Smooth Handoff mechanism described in [19]. This mechanism consists of the oFA buffering a certain amount of packets sent to the MN. After the BET has been established between the nFA and oFA in the Post-Registration, packets arriving to the oFA during the 802.11 handoff destined to the MN could have the chance to reach the MN along the nFA.

4. Performance Evaluation of Low Latency Handoff Mechanisms

In this part we analyze and compare the performance of the two mechanisms. An analytical model is introduced and used to compare the delay characteristics and the buffer requirements for an MN that switches from one access point to another. A similar model has been used to investigate other protocols (Cellular IP [1], HAWAII [2], Optimized Smooth Handoff [3] and Post-Registration [5]).

4.1 An Analytical Model for Pre- and Post-Registration Handoff

In this section we propose a simple analytical model that enables us to compute and compare performance characteristics of the Pre- and Post-Registration handoff procedure. In particular, we are interested in the delay distribution of packets involved in a handoff, and in the buffer requirements of the relevant FAs.

Consider the network architecture as depicted in Figure 1. We model all routers in the network as ordinary M/M/1 queues. Hence, each packet passing through some router has an exponentially distributed random service time, which is assumed to both include the processing time in the router and the transmission time. Moreover the response time of a packet is also exponentially distributed.

Now consider an MN moving from the oFA to the nFA. We assume that a handoff is initiated when the anticipation trigger occurs and we denote this instant by t_0 . In this section we furthermore assume that we have a source-triggered handoff, but the model can also be applied to target-triggered handoff. The timing of the relevant L2 triggers (L2-LD and L2-LU) is considered to be constant. We define D_{LD} and D_{LU} as the time between t_0 and the respective triggers and we have that $0 < D_{LD} < D_{LU}$.

The performance of the handoff procedures strongly depends on the timing of these triggers as will be shown in the next section. Additionally, there are some important random time instants for both methods.

- For Pre-Registration, notably important is the moment the regional registration request arrives at the GFA. From that moment on packets will be directed to the nFA instead

of the oFA. In our M/M/1 queuing model this moment ($t_{1,pre}$) is random and is distributed as the sum of several exponential variables (the response times of the routers on the path from oFA via nFA to GFA), and several constant values (the propagation delays on the links between these routers).

- For Post-Registration, the first essential random value is the moment the BET is established. This instant ($t_{1,post}$) is also distributed as the sum of exponential variables (the response times of the routers on the path from oFA to nFA and back to oFA) and constants (the propagation delays). A second essential value is the moment the regional registration request arrives at the GFA ($t_{2,post}$). Once more this instant is randomly distributed following the sum of exponential variables (the response times of the routers on the path from nFA to GFA) and constants (D_{LU} and the propagation delays).

We now look at a CBR UDP stream originating from a CN with destination the MN. Suppose that every T ms a packet arrives at the GFA (we ignore the jitter introduced by the backbone network). We can then observe the stream starting from a packet arriving some time before t_0 , and compute the distribution of the end-to-end delay (GFA to MN) of this packet and each subsequent packet involved in the handoff.

The delay of each packet is again a random variable made up from exponential variables and constants, and its specific form depends on the path the packet follows. The latter is in turn dependent on stochastic events involving the random time instants defined above. Denote by t_{GFA}^1 the predetermined arrival instant of the first observed packet at the GFA. The arrival of the k -th packet then equals $t_{GFA}^k = t_{GFA}^1 + (k-1)T$. Denote by ${}_{xFA}\Delta_{yFA}$ the random time needed for a packet to travel from xFA to yFA through the connecting routers. In the case of Pre-Registration the following stochastic events determine the path of the k -th packet and thus its delay distribution:

1. $t_{GFA}^k < t_{1,pre} \rightarrow k$ -th packet routed via the oFA
 - a. $t_{GFA}^k + {}_{GFA}\Delta_{oFA} < t_0 + D_{LD} \rightarrow$ packet forwarded from oFA to MN
 - b. $t_{GFA}^k + {}_{GFA}\Delta_{oFA} > t_0 + D_{LD} \rightarrow$ packet lost
2. $t_{GFA}^k > t_{1,pre} \rightarrow k$ -th packet routed via the nFA
 - a. $t_{GFA}^k + {}_{GFA}\Delta_{nFA} < t_0 + D_{LU} \rightarrow$ packet needs to be buffered at nFA, in which case it will be forwarded to MN at $t_0 + D_{LU}$.

- b. $t_{GFA}^k +_{GFA} \Delta_{nFA} > t_0 + D_{LU} \rightarrow$ packet forwarded from nFA to MN

And similarly, for Post-Registration the following scheme holds:

1. $t_{GFA}^k < t_{2,post} \rightarrow k$ -th packet routed via the oFA
 - a. $t_{GFA}^k +_{GFA} \Delta_{oFA} < t_0 + D_{LD} \rightarrow$ packet forwarded from oFA to MN
 - b. $t_0 + D_{LD} < t_{GFA}^k +_{GFA} \Delta_{oFA} < t_{1,post} \rightarrow$ packet needs to be buffered at oFA, in which case it will be forwarded through the BET at $t_{1,post}$
 - c. $\max(t_0 + D_{LD}, t_{1,post}) < t_{GFA}^k +_{GFA} \Delta_{oFA} \rightarrow$ packet forwarded through the BET
 - i. $t_{GFA}^k +_{GFA} \Delta_{oFA} +_{oFA} \Delta_{nFA} < t_0 + D_{LU} \rightarrow$ packet needs to be buffered at nFA, in which case it will be forwarded to MN at $t_0 + D_{LU}$
 - ii. $t_{GFA}^k +_{GFA} \Delta_{oFA} +_{oFA} \Delta_{nFA} > t_0 + D_{LU} \rightarrow$ packet forwarded from nFA to MN
2. $t_{GFA}^k > t_{2,post} \rightarrow k$ -th packet routed via the nFA, forwarded from nFA to MN

Taking into account these schemes allows us to compute the overall delay distribution of the k -th packet. In particular, denoting by $_{GFA} E_{MN}^k$ the end-to-end delay from GFA to MN for packet k , we have that

$$P[_{GFA} E_{MN}^k > t] = \sum_{j \in J} P[(_{GFA} E_{MN}^k > t) \& (event \ j)]$$

where $J = \{1a, 1b, 2a, 2b\}$ for Pre-Registration and $J = \{1a, 1b, 1ci, 1cii, 2\}$ for Post-Registration.

The end-to-end delay takes on different forms according to the event it is associated with. As an example, for Post-Registration in case of events (1c i) and (1c ii) we have that

- i. $_{GFA} E_{MN}^k = t_0 + D_{LU} + F_{MN} - t_{GFA}^k$
- ii. $_{GFA} E_{MN}^k =_{GFA} \Delta_{oFA} +_{oFA} \Delta_{nFA} + F_{MN}$

where F_{MN} denotes the forwarding of the packet from the nFA to the MN, which is modeled as one more response time for the router nFA. The propagation delay between an FA and the MN is assumed to be negligible. Based on the M/M/1 assumptions, the delay distribution can now be computed in a rather tedious but fairly straightforward way.

Remark that we can compute the delay distribution either on the assumption of available buffer size at the FAs or on the assumption of absence of buffers. In the latter case we can e.g. compute the expected number of lost packets due to the handoff (see the next section).

To determine the size of the buffer that should be installed in the FAs in order to avoid or minimize packet loss, we can proceed as follows. We focus on buffering at the nFA, although for Post-Registration, a buffer at the oFA could also be considered. At the nFA the buffer is needed for packets that arrive before the LU trigger occurs. (Note that this is assumed to be a separate buffer meant to store the waiting packets destined to the MN and this is not modeled as an M/M/1 queue.) The required buffer size can be determined by computing the distribution of N_B , the number of packets that would be lost in the absence of a buffer. If we denote by $p_{loss}(M)$ the probability that at least one packet is lost during a handoff at the nFA with buffer capacity M , then we have

$$p_{loss}(M) = P[N_B > M].$$

Specific random time intervals I_B can be determined for which it holds that a packet is lost (in the absence of a waiting buffer) if and only if it arrives in that interval at the nFA. These are given by

$$Pre: I_B = [t_{1,pre} +_{GFA} \Delta_{nFA} ; D_{LU}]$$

$$Post: I_B = [\max(t_{1,post}, t_0 + D_{LD}) +_{oFA} \Delta_{nFA} ; D_{LU}]$$

Regarding Post-Registration we have assumed here that no waiting buffer is present at the oFA. Hence the packets that would get lost at the nFA are those arriving in the oFA after the link is down and after the BET is established, and arriving via the BET at the nFA before the link is up.

The respective lengths L_B of these intervals have a distribution composed of exponential variables and constants, as again follows from our M/M/1 model. Conditioned on L_B it is straightforward to approximate the distribution of N_B , so we can consider

$$p_{loss}(M) = \int P[N_B > M \mid L_B = z] f_{L_B}(z) dz$$

where f_{L_B} is the density function of L_B .

The integral can be computed numerically. The latter equation can then be used as an alternative method to calculate the expected number of packets that need buffering and furthermore we can determine the required buffer size at the nFA by applying

$$\min M : p_{loss}(M) < 10^{-\alpha}.$$

Here we can choose e.g. $\alpha=5$. In the next section we present some numerical results concerning M , as well as results for the delay distribution of the stream of packets.

4.2 Numerical Results

We assume a network topology as depicted in Figure 1. The results that will be shown are all obtained with the following network characteristics. The propagation delays on the links connecting the GFA and the oFA, as well as the links connecting the GFA and the nFA are all set to $\tau_1 = 5\text{ms}$, while the links connecting the oFA and the nFA (via router 3) have a propagation delay of $\tau_2 = 3\text{ms}$. The service rate μ in each router is set to 1 packet per ms and all routers have a load of $\rho = 0.8$. This leads to an exponentially distributed response time for each router with rate equal to $\mu(1-\rho) = 0.2/\text{ms}$. Furthermore we assume that the CN transmits a packet every $T = 10\text{ms}$ (UDP stream) destined for the MN. These characteristics are arbitrarily chosen and are not essential since the model is used mainly for comparison between Pre- and Post-Registration.

First we present some results for the delay distribution of a stream of packets involved in a handoff. The start of the handoff is set to $t_0 = 0$, and we observe a stream of 30 packets, the first of which arrives at the GFA at $t_{GFA}^1 = -80\text{ms}$. We define the playout time as the maximum allowed end-to-end delay from GFA to MN: if a packet's end-to-end delay exceeds the playout time it will be dropped.

In Figure 6 we assume that there is no buffer capacity available, which implies that packets can get lost due to the handoff. The curves depict the expected number of packets from the stream that are dropped due to the expiration of the play-out time or the absence of a buffer. 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 to 60ms since t_0 . 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 results in 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.

In Figure 7 we show the impact of providing infinite buffers at the FAs. For Post-Registration the expected number of dropped or lost packets in case of infinite buffers is compared with the case of the absence of buffers. Note that the packets that would be lost when there is no buffer available will obviously not be lost when buffers are installed, but they will experience a larger average end-to-end delay.

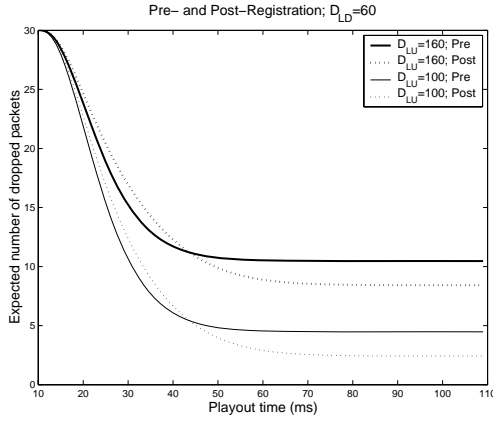


Figure 6: Delay – Post- and Pre-Registration

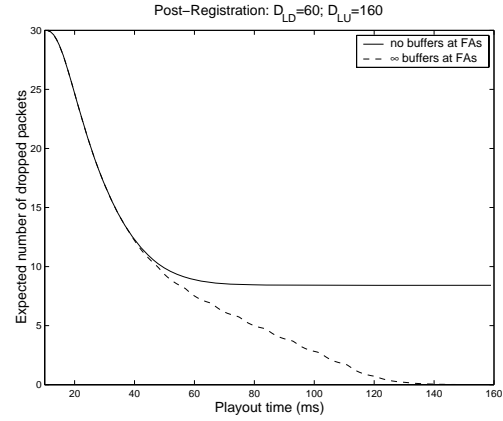


Figure 7: Delay – Post-Registration

Figure 8 and Figure 9 depict the delay distribution of each individual packet of the observed stream. In particular the curves present the probability that the GFA-MN end-to-end delay of the k -th packet exceeds t . The timing of the LD and LU trigger is fixed at 80ms and 120ms respectively and we assume infinite buffer sizes. The curves for the first few packets and the last few packets are essentially the same, since these packets are not influenced by the handoff. It can then be seen that for Post-Registration the influence of the handoff on the stream's delay starts later as well as lasts longer. This corresponds to the fact that Post-Registration (optimally) keeps using the link between the oFA and the MN, while Pre-Registration possibly already redirects packets before this link is down. Such packets (e.g. $k=12$ and $k=13$) then have a longer delay since they have to wait for the LU trigger. The disadvantage of Post-Registration is reflected by the fact that it lasts until packet 25 for the influence of the handoff procedure to have disappeared completely. Note that packets 19 to 22 have the same delay distribution. These packets, according to our model, all travel with a high probability through the BET, straight to the MN without having to be buffered. In summary, these figures show that Post-Registration is better adjusted to the timing of the triggers, but more packets have a slightly larger delay. A Pre-Registration handoff affects fewer packets, but some of these can have unnecessary high delays.

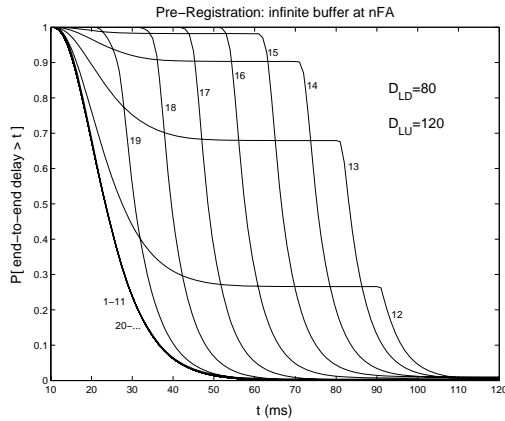


Figure 8: Individual delays – Pre

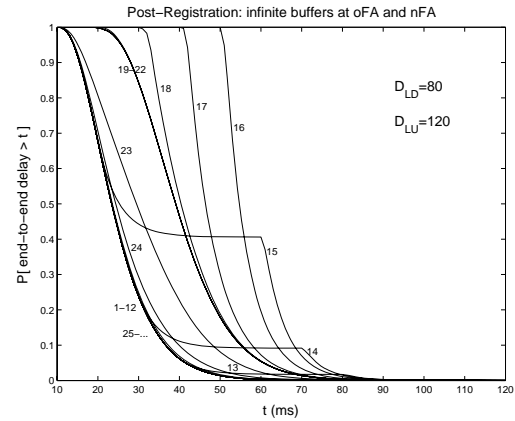


Figure 9: Individual delays – Post

Finally we show some results on the required buffer capacity as a function of the timing of the L2 triggers. Figure 10 and Figure 11 each depict both the expected number of packets that would be lost if there would be no buffer available at the nFA (thick line), as well as the minimum buffer size M to ensure that $P_{loss}(M) < 10^{-5}$ (thin line). Each figure compares Pre-Registration (solid line) to Post-Registration (dotted line). In Figure 10 the timing of the LU trigger is fixed at $D_{LU} = 200\text{ms}$, and D_{LD} is varied. In case of Pre-Registration, only the timing of the LU trigger affects the situation at the nFA, and not the timing of the LD trigger, hence the horizontal curves. As the time between the LD and LU trigger decreases, the number of packets that need to be buffered at the nFA in case of Post-Registration decreases as well. As can be seen from the figure, Post-Registration offers considerable gain over Pre-Registration when the time between the triggers is limited.

In Figure 11 both LD and LU vary, but the time between LD and LU is fixed. The required buffer capacity for Pre-Registration is obviously linearly related to the timing of the LU trigger, while for Post-Registration it is essentially the difference (which is fixed here) between the two triggers that determines the buffer requirements. Concerning Post-Registration, the fact that for smaller values of D_{LU} fewer packets need to be buffered, is explained by noting that in these cases fewer packets are transmitted through the BET because the BET might not yet be established at the time of D_{LD} and we did not consider a buffer at the oFA. These last two figures indicate that Pre-Registration in general, but not necessarily, requires more buffer capacity than Post-Registration.

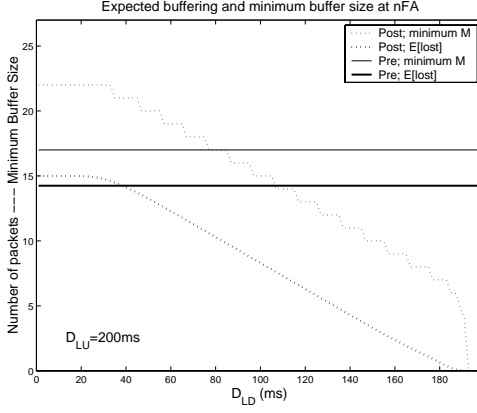


Figure 10: Buffering in nFA

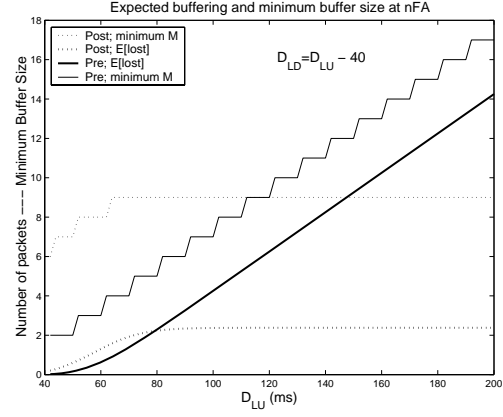


Figure 11: Buffering in nFA

5. Scalability of the Handoff Protocols

5.1 OPNET Model Description

The OPNET simulation models used in this paper (see also [21]) are built into OPNET's standard models [16], extending their functionality and allowing extensive use of the stock router, link and host models. The implemented protocols are Mobile IP [17] extended with Route Optimization [18], Hierarchical MIP [12], Optimized Smooth Handoff [19] and Low Latency Handoffs [10].

Mobility agents are modeled as regular hosts, not as one co-located router and access point node. These entities are derived from standard workstation models, but extended with several new processes: mobility registration service, (de-)tunneling service, agent advertisement mechanism and proxy ARP (used at the HA). These are located on the same IP (sub)network as a Layer 2 bridge, which in this case is an 802.11b WiFi <-> Ethernet bridge. Mobile nodes are very similar to the agents detailed above as they are descendant from the same base model and process additions, but they are equipped with an arbitrary number of Layer 2 interfaces, one for each network the node can roam unto. While they do not need to be of the same type, all interfaces are 802.11b for the course of this paper.

The handover protocols discussed here rely on a level of integration between IP and the underlying access technology in a way that the IP layer needs to receive certain input triggers from the underlying one. For the sake of evaluating these protocols per IETF standards, all triggers are delivered in a “deus ex machina” way. Each MN in the scenario executes a pre-defined script in which all handoff events are given (either constant time or according to some

distribution). Execution blocks in this script can be repeated any number of times to allow for very controlled movement scenarios.

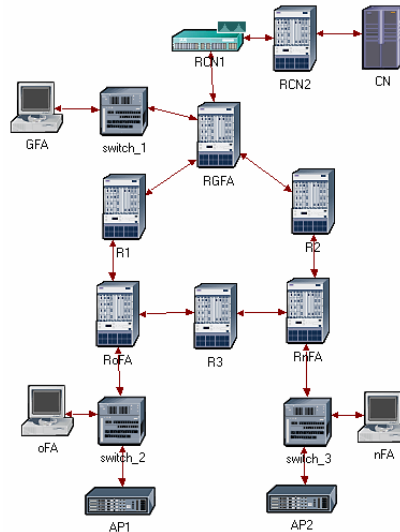


Figure 12: OPNET Network Model

Figure 12 shows a typical OPNET setup. Only the hierarchical branch and the correspondent node are shown here, not the home network or any mobile nodes. The foreign agent entities are connected to the access points through Ethernet switches. The GFA is also a simple host on an Ethernet domain, which is 10Mbps throughout. Special attention is needed when configuring this network for a large number of MNs, e.g. ARP cache sizes in the foreign agents must be large enough to support the number of nodes.

5.2 Results for CBR Traffic

The CBR traffic consists of constant-sized (500 bytes) UDP packets sent to all the mobile nodes every 50ms. The mobility script for each MN is as follows: it starts in one of the two domains and draws a uniformly distributed random value between 0 and 5s to initiate its first move; the MNs are divided equally among the APs at the start of the simulation. Every MN then stays a uniformly distributed dwell time (between 4 and 6 seconds) in one domain before switching to the other.

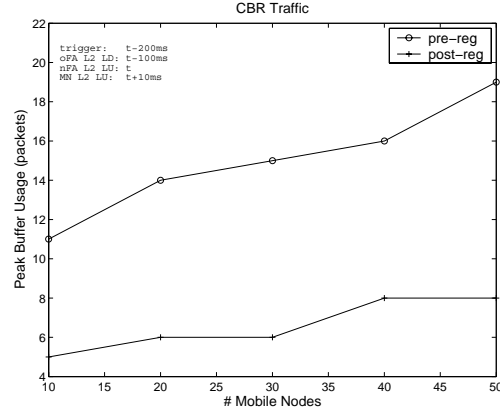


Figure 13: Peak buffer usage

In Figure 13 the peak buffer usage is measured in both foreign agents when each MN conducts 100 handovers. Post-Registration needs significantly less buffers as the time needed to buffer packets depends on the spacing of oFA L2-LD and nFA L2-LU since the bi-directional edge tunnel (BET) starts forwarding traffic the moment the link goes down. In Pre-Registration however, the anticipation trigger which occurs before L2-LD causes the MN to register with its nFA through the oFA. This registration goes up to the GFA who will change the routing path for the MN's traffic.

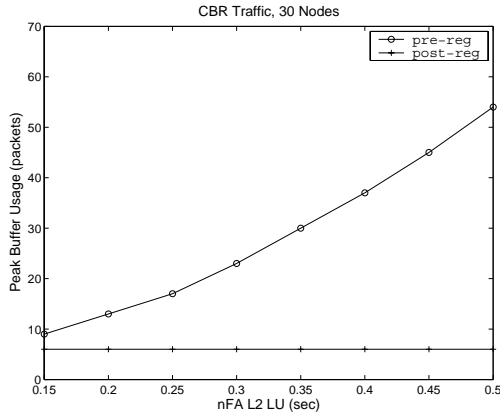


Figure 14: Peak buffer usage varying L2-LU

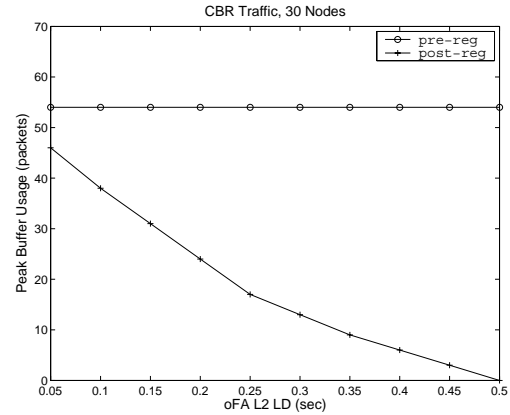


Figure 15: Peak buffer usage varying L2-LD

Figure 14 and Figure 15 show peak buffer usage measured in the foreign agents with a constant number of MNs but when varying both the nFA L2-LU and the oFA L2-LD trigger. In Figure 14 the anticipation trigger arrives at 0 secs, and oFA L2-LD precedes the LU event by 100ms. In Figure 15, oFA L2-LD varies between the anticipation trigger and nFA L2-LU at 500ms. They clearly illustrate the dependence of buffer requirements on one of the triggers:

Post-Registration is not influenced by the timing of the anticipation trigger, as long as there is time enough for BET setup (this is dictated by the protocol requirements). Similarly, Pre-Registration's buffer requirements remain unaffected when varying the oFA L2-LD trigger. Remark that these results are completely in correspondence with the conclusions drawn from the analytical results (see Figure 10 and Figure 11).

A certain amount of traffic delay is of course associated with these handovers. Apart from the delay experienced due to buffering, which will usually be higher in Pre-Registration, the Post-Registration will have an additional delay due to use of the BET. This tunnel is used until the MN node chooses to register. Neither Pre- nor Post-Registration induce an extra delay when the number of nodes is increased: the buffers in place are logical per-node buffers.

5.3 Results for TCP Traffic

TCP traffic consists of FTP file transfers (10MB) to each of the mobile nodes. FTP response time is then measured at each of the mobile nodes (time between start and end of the file request) and averaged over the total of nodes. TCP Reno is used in all the runs, but the MSS has been reduced to 1440 bytes to avoid fragmentation induced by tunneling (an extra IP header is added to tunneled packets). Ethernet links are loaded with background traffic to avoid overloading the wireless channel.

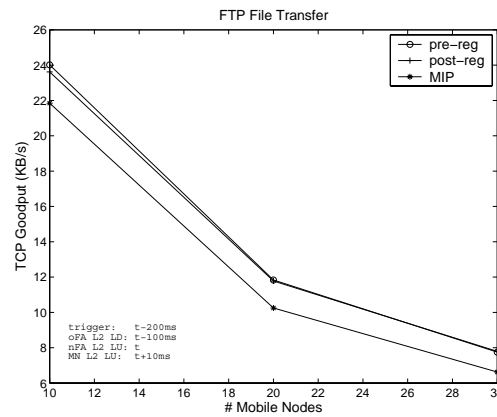


Figure 16: TCP Goodput

Figure 16 shows the advantage of using a low latency handover method as opposed to plain MIP. TCP flows clearly benefit from smoothness of the handoffs, but there is no significant difference between the two methods discussed here. Post-Registration might behave slightly worse because there is a possibility of packet re-sequencing: packets

following the old route through the tunnel might arrive later than packets sent directly to the nFA if the MN has just performed normal registration.

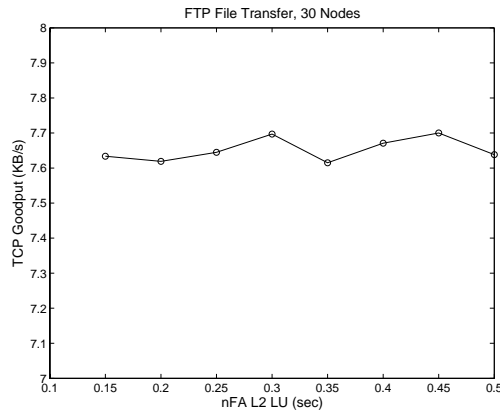


Figure 17: TCP Goodput, Pre-Reg

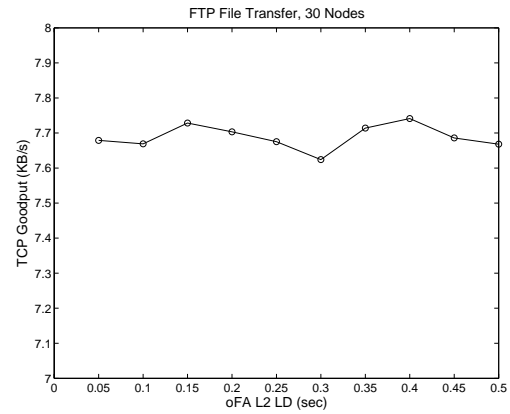


Figure 18: TCP Goodput, Post-Reg

The influence of trigger timing on TCP's performance is negligible: the time that packets spend in the FA's buffer is not large enough to cause performance degradation as the timeout values are larger: TCP commonly has an absolute bottom value of 500ms for timeout timers. This is illustrated in Figure 17 and Figure 18.

5.4 Impact of Network Topology

Let us now consider at a tree-structured network depicted in Figure 19 instead of the reference network used throughout this paper. This shows a more tree-like physical setup and is identical to the architecture of Figure 12, except for R3 which has been moved up one level to become the crossover router CR.

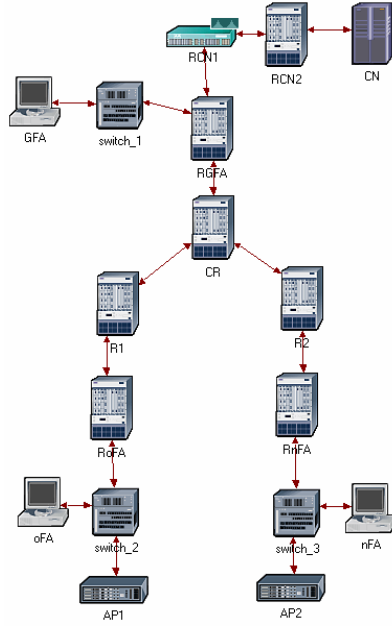


Figure 19: Alternate network topology

Using the mobility scenario described above, we can examine protocol performance differences between these two networks for e.g. Pre-Registration. To emphasize the impact of the router R3, the 1-hop distance (delay) between oFA and nFA in the reference network has been made very small. Figure 20 shows that the original network has higher buffer requirements: when the routing distance between oFA and nFA is very small, the registration request to the nFA routed through the oFA will reach the GFA much sooner, resulting in more packets being forwarded for the same trigger values. However, both curves behave very similarly.

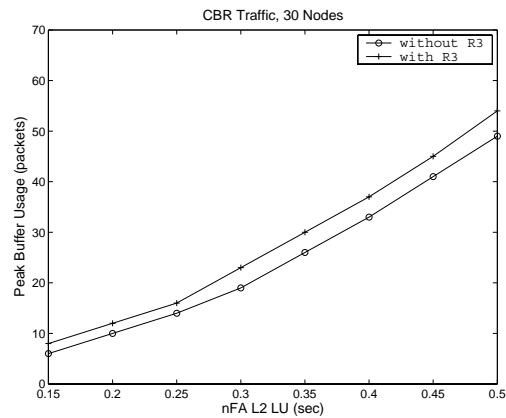


Figure 20: Peak buffer usage for both architectures

6. Conclusion

In this paper, we have analyzed two Layer 3 low latency mobility protocols, Pre- and Post-Registration. Their operation relies on the availability of L2 triggers. Possible implementations when using IEEE 802.11 as link layer are discussed and illustrated using an ns simulation. A simple analytical model is presented that allows a detailed investigation of the delay characteristics and the buffer requirements for a single mobile node involved in a handoff. It is shown that Pre-Registration needs more buffers to obtain zero losses. With respect to delay, Post-Registration is better adjusted to the timing of the triggers, but more packets have a slightly larger delay. A Pre-Registration handoff affects fewer packets, but some of these can have unnecessary high delays. OPNET simulations confirm these conclusions in more realistic systems with respect to the number of mobile nodes and their mobility pattern. The difference in TCP throughput for both schemes is negligible.

References

- [1] C. Blondia, O. Casals, P. De Cleyn and G.Willems, "Performance analysis of IP Micro-Mobility Handoff Protocols", Proceedings of Protocols for High Speed Networks 2002 (PfHSN 2002), Berlin 2002, pp 211-226.
- [2] C. Blondia, O. Casals, Ll. Cerdà and G.Willems, "Performance analysis of a forwarding scheme for handoff in HAWAII", Proceedings of Networking 2002, Pisa 2002, pp 504-514.
- [3] C. Blondia, O. Casals, N. Van den Wijngaert, G. Willems, "Performance analysis of smooth handoff in Mobile IP", Proceedings of The Fifth ACM International Workshop on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM2002), Atlanta, USA, September 2002
- [4] C. Blondia, O. Casals, Ll. Cerdà, N. Van den Wijngaert, G. Willems, P. De Cleyn, "Low Latency Handoff Mechanisms and their implementation in an IEEE 802.11 Network". To appear in the Proceedings of ITC 2003.
- [5] O. Casals, Ll. Cerdà G. Willems, C. Blondia, N. Van den Wijngaert, "Performance evaluation of the Post-Registration method, a low latency handoff in MIPv4", to appear in Proceedings of ICC 2003.
- [6] A.T. Campbell, J. Gomez, C-Y. Wan, S. Kim, Z. Turanyi, and A. Valko, "Cellular IP", draft-ietf-mobileip-cellular-00.txt, December 1999.
- [7] A.T. Campbell, J. Gomez, C-Y. Wan, S. Kim, Z. Turanyi, and A. Valko, "Cellular IP Performance", draft-gomez-cellularip-perf-00.txt, October 1999.

- [8] A.T. Campbell, J. Gomez-Castellanos, "IP Micro Mobility Protocols", ACM SIGMOBILE Mobile Computing and Communications Review, October 2000, Vol. 4, No. 4, pp 45-54.
- [9] P. McCann, "Mobile IPv6 Fast Handovers for 802.11 Networks", IETF draft-mccann-mobileip-80211fh-00.txt, 2002.
- [10] K. El Malki and others, "Low Latency Handoffs in Mobile IPv4", IETF draft-ietf-monileip-lowlatency-handoffs-v4-04.txt, 2002.
- [11] S. Goswami, "Simultaneous Handoff of Mobile-IPv4 and 802.11", IETF draft-goswami-mobileip-simultaneous-handoff-v4-01.txt, 2002.
- [12] E. Gustafsson, A. Jonsson, C. Perkins. "Mobile IP Regional Registration", draft-ietf-mobileip-reg-tunnel-02.txt, March 2000.
- [13] H. Haverinen, J. Malinen, "Mobile IP Regional Paging", draft-haverinen-mobileip-reg-paging-00.txt, June 2000.
- [14] J. Kempf, P. R. Calhoun, C. Pairla, "Foreign Agent Assisted Handoff", draft-calhoun-mobileip-proactive-fa-01.txt, June 2000.
- [15] A. Mishra, M. Shin and W. Arbaugh. "An Empirical Analysis of the IEEE 802.11 MAC Layer Handoff Process". Technical report number CS-TR-4395, University of Maryland, Department of Computer Science, September 2002.
- [16] OPNET Modeler, www.opnet.com.
- [17] C.E. Perkins. "IP Mobility Support", rfc 2002, October 1996.
- [18] C. Perkins and D. Johnson, "Route Optimization in Mobile IP", IETF Internet Draft, draft-ietf-mobileip-optim-11.txt, September 2001.
- [19] C. Perkins and K-Y. Wang. "Optimized smooth handoffs in Mobile IP", Proceedings of IEEE Symposium on Computers and Communications, Egypt, July '99.
- [20] R. Ramjee, T. La Porta, S. Thuel, K. Varadhan, and L. Salgarelli, "IP micro-mobility support using HAWAII", draft-ietf-mobileip-hawaii-00.txt, June 1999.
- [21] N. Van den Wijngaert and C. Blondia, "Integration of IP mobility in OPNET: modeling and simulation", proceedings of OPNETWORK 2002, Washington D.C., August 2002.
- [22] F. Vena, Ll. Cerdà O. Casals. "Study of the TCP Dynamics over Wireless Networks with Micromobility Support Using the ns Simulator", European Wireless 2002, Florence, February 2002, pp 114-120.

- [23] A.E. Yegin and others, “Supporting Optimized Handover for IP Mobility – Requirements for Underlying Systems”, IETF draft-manyfolks-12-mobilereq-02.txt, 2002.
- [24] ns with micromobility support, <http://comet.ctr.columbia.edu/micromobility>
- [25] The Network Simulator - ns-2, <http://www.isi.edu/nsnam/ns>
- [26] ANSI/IEEE Std 802.11, 1999 Edition.