Performance Analysis of Fast Handover in Mobile IPv6 Networks [*]

Sangheon Pack and Yanghee Choi

School of Computer Science&Engineering
Seoul National University, Seoul, Korea
shpack@mmlab.snu.ac.kr and yhchoi@snu.ac.kr
Tel: +82-2-880-1832, Fax: +82-2-872-2045

Abstract. The Fast Handover protocol [1] provides seamless handover
in wireless IP networks by minimizing handover latency. To reduce han-
dover latency and to provide faster handover, Fast Handover uses an-
ticipation based on layer 2 (L2) trigger information. Therefore, it incurs
higher signaling costs compared with the basic Mobile IP protocol. Fur-
thermore, since the L2 trigger is based on fluctuating wireless channel
states, the handover anticipation using the L2 trigger may sometimes be
incorrect. In the case of incorrect anticipation, unnecessary buffer space
may be used for the purpose of providing a smooth handover. There-
fore, it is essential to analyze these overhead costs, in order to evaluate
and compare the performance of Fast Handover with that of the basic
Mobile IP protocol. In this paper, we analyzed the overhead associated
with Fast Handover including the signaling cost and the packet delivery
cost. We formulated these costs based on a timing diagram and com-
pared Fast Handover with basic Mobile IPv6 in terms of their packet
loss rates and buffer requirements. Also, we studied the impact of the L2
triggering time on the total overhead cost. As a result, we found that the
L2 triggering time is an important factor to consider in the optimization
of handover performance.

Keyword: Fast Handover, Mobile IP, IP-based wireless/mobile networks, L2
trigger, Performance analysis.

1   Introduction

In IP-based mobile/wireless networks, minimizing handover latency is one of the
most important issues. The Mobile IPv6 (MIPv6) specification defines how mo-
bile host (MH) can maintain connectivity to the Internet when its attachment
point is changed from one access router (AR) to another. During the handover
procedure, there is a time period during which the MH is unable to send or receive

any packets, due to both link switching delay and IP protocol operation considerations. This time period is referred to as *handover latency*. In many instances, handover latency resulting from standard MIPv6 handover procedures could be greater than what is acceptable to support real-time applications. Therefore, an enhanced MIPv6 specification, called *Fast Handover* [1], has been proposed by the Internet Engineering Task Force (IETF), in order to reduce handover latency due to IP protocol operations, for it to be as small as possible in comparison with the inevitable link switching latency. Fast Handover provides seamless handover in wireless IP networks using a layer 2 (L2) trigger. To evaluate the performance of Fast Handover in different network environments, several studies have been done [4] [5] [6].

Koodli et al. presented recent works enabling fast handovers and context transfer between two ARs [4]. Also, Koodli et al. proposed a framework for general context transfer solutions. In terms of performance evaluation, handover latency was studied in two different cases, depending on whether the Fast Binding Acknowledgment message is received on an old link or a new link.

Costa et al. compared the handover latency of different IP mobility management schemes currently being discussed within the IETF. They include basic MIPv6, Fast Handover (FMIPv6), and Hierarchical MIPv6 (HMIPv6). FMIPv6 supports a faster handover procedure compared with basic MIPv6, while HMIPv6 provides an approach allowing for different hierarchies of mobility agents [5]. In this work, they studied handover latency for each protocol and concluded that the best option in order to get the better performance is to implement both HMIPv6 and FMIPv6.

Fikouras et al. studied the performance of MIP handover [6]. According to the results of their study, handover latency is largely dependent on the efficiency of the various movement detection methods such as Lazy Cell Switching, Eager Cell Switching, and Prefix Matching. In their report, two generic formulas were derived for determining the average handover latencies of the Lazy Cell Switching and Eager Cell Switching algorithms.

Blondia et al. investigated the performance of two low latency handover protocols called pre- and post-registration [7]. They proposed a simple analytical model to access the packet loss and the delay characteristics of these protocols. Besides, their scalability properties are investigated by means of an OPNET simulation model implementing IEEE 802.11 as a link layer protocol.

Although the Fast Handover scheme can provide low latency handover, this is only achieved at a certain cost in terms of additional overhead. First, an increased number of signaling messages, between the AR and the MH, are required to provide for the anticipated handover procedure. Also, since Fast Handover is based on the anticipation of future events made using link layer information, there may be some overhead resulting from incorrect predictions. For example, when packet forwarding is supported for smooth handover, the forwarded packets are stored in buffer space. However, if the handover anticipation is wrong, so that the handover doesn't really occur, the allocated buffer space constitutes useless overhead. Therefore, to evaluate the performance of Fast Handover, we need to

take these different sources of overhead into consideration. However, all of the previous studies [4] [5] [6] [7] on this issue focused on handover latency and didn't consider the effects of overhead, when Fast Handover is employed in wireless IP networks.

In this paper, we analyzed the issue of overhead in Fast Handover. Based on our analysis, we concluded that the total amount of overhead is largely dependent on the question of when the L2 trigger is notified to layer 3 (L3). Therefore, we studied the optimal point in time for the L2 trigger to occur, in order to minimize the total overhead. Also, we compared Fast Handover with basic Mobile IP in terms of the packet loss rate and buffer requirements.

The rest of this article is organized as follows. Section II describes the mechanism of the Fast Handover protocol proposed by the IETF. In Section III, we analyze the overhead of associated with Fast Handover. Section III shows the numerical results of our study based on various parameters. Section IV concludes this paper.

2   Fast Handover

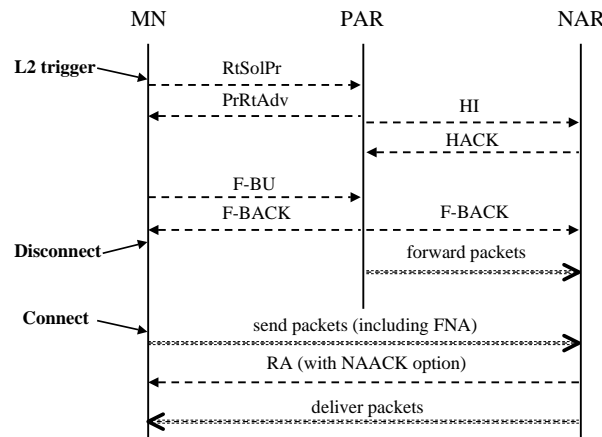Figure 1 shows the sequence of message flows used in the Fast Handover protocol.

Fig. 1. Fast Mobile IPv6 Handover Protocol

While the MH is connected to its previous access router (PAR) and is about to move to a new access router (NAR), Fast Handover in Mobile IPv6 requires:

— the MH to obtain a new care-of address(CoA) at the NAR while still being
  connected to the PAR
— the MH to send a Binding Update message to its PAR to update its binding
  cache with the MH's new CoA
— the PAR to start forwarding packets destined for the MH to the NAR

Either the MH or the PAR may initiate the Fast Handover procedure by us-
ing the L2 trigger. The link-layer information indicates that the MH will soon be
handed over from one wireless access point (AP) to another, with these two APs
being attached to the PAR and the NAR, respectively. If the L2 trigger is re-
ceived at the MH (Mobile-initiated handover), the MH will initiate L3 handover
by sending a Router Solicitation for Proxy (RtSolPr) message to the PAR. On
the other hand, if the L2 trigger is received at the PAR (Network-controlled han-
dover), then the PAR will transmit a Proxy Router Advertisement (PrRtAdv)
message to the appropriate MH, without any solicitation messages.

The MH obtains a new CoA (NCoA), while still being connected to the
PAR, by means of router advertisements from the NAR containing network
information. The PAR will validate the MH's new CoA and initiate the process
of establishing a bidirectional tunnel between the PAR and the NAR, by sending
a Handover Initiate (HI) message to the NAR. Then, the NAR verifies that its
new CoA can be used on the NAR's link. Also, in response to the HI message,
the NAR sets up a host route for the MH's previous CoA (PCoA) and responds
with a Handover Acknowledge (HACK) message.

When the MH receives a PrRtAdv message, it should send a Fast Binding
Update (F-BU) message, preferably prior to disconnecting its link. When the
PAR receives an FBU message, it must verify that the requested handover is
accepted by the NAR as indicated in the HACK message status code. Then,
it begins forwarding packets intended for PCoA to the NAR and sends a Fast
Binding Acknowledgement (F-BACK) message to the MH.

In basic Mobile IP, there exists a service disruption period during the han-
dover procedure. The period of service disruption usually occurs due to the time
required for an MH to inform the Home Agent (HA) of its current location, after
it moves into the area of the NAR. During this period, the MH cannot resume
or continue communication. On the other hand, Fast Handover [1] involves the
use of an L2 trigger which allows L3 handover to be anticipated rather than
being performed after the completion of the L2 handover procedure. Thus, Fast
Handover ensures that the L3 handover delay is minimized, and also that the
period of service disruption, which normally occurs when an MH moves between
two ARs, is eliminated. Figures 2 and 3 are cited from [4]. They show timing
diagrams corresponding to basic Mobile IPv6 handover and Fast Handover. In
basic Mobile IPv6 handover, $t_L + t_I + t_P$ is the service disruption period. On the
other hand, in Fast Handover, packet reception latency is equal to IP connec-
tivity latency ($t_P = t_I$). Namely, $t_L + t_I$ is the service disruption time in Fast
Handover. However, since packet forwarding from PAR to NAR is supported in
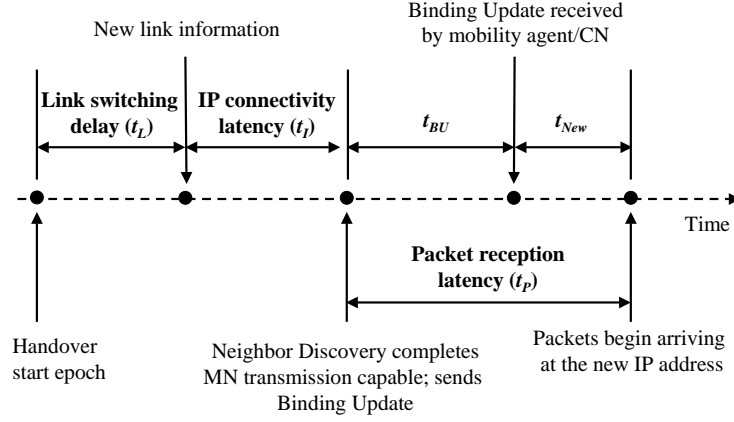Fast Handover, no packets arriving during this period ($t_L + t_I$) are lost.

Fig. 2. Timing Diagram in Basic Mobile IPv6

3   Performance Analysis

In this section, we present analytic cost functions to evaluate Fast Handover and basic MIPv6. We classify the total cost into two types: signaling cost (C_{signal}) and packet delivery cost (C_{packet}). The signaling cost is incurred by the signaling messages used for a handover procedure. On the other hand, the packet delivery cost is related to the packet loss rate associated with a handover procedure. Then, the total cost (C_{total}) can be expressed as follows.

$$C_{total} = C_{signal} + C_{packet} \tag{1}$$

3.1   Signaling Cost

Unlike localized mobility management schemes [2] [3], Fast Handover does not concern itself with the minimization of signaling costs. Therefore, we analyze the additional signaling costs incurred in Fast Handover. The message flows associated with Fast Handover are described in Figure 1. Since Fast Handover is based on anticipation using the L2 trigger, signaling costs due to additional messages may vary according to the probability that the handover anticipation is correct. Let P_s and P_f be the probabilities that L3 handover does or does not occur after the L2 trigger, respectively. If there is no real handover after the L2 trigger, the RtSolPr/PrRtAdv and HI/HACK messages may be useless. M_f denotes the signaling cost for this case. On the other hand, M_s denotes the
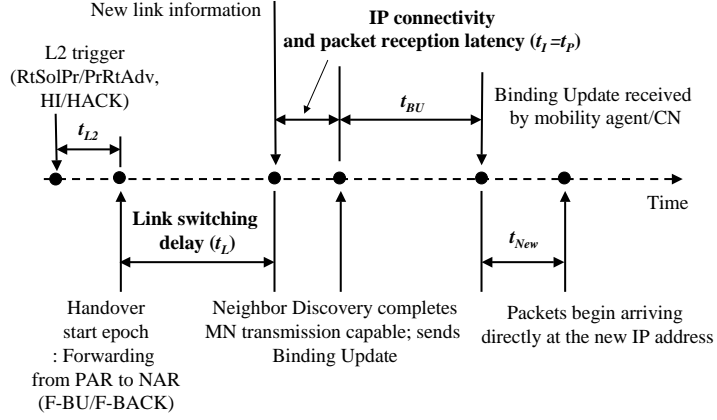
Fig. 3. Timing Diagram in Fast Handover

signaling cost for a successfully anticipated handover. The two signaling costs can be calculated based on the message flows depicted in Figure 1, by means of the equations listed below. TC_{MP} is the transmission cost incurred in the wireless link between MH and PAR. TC_{PN} is the transmission cost incurred in the wired link between PAR and NAR. PC_N and PC_P are processing costs in NAR and PAR, respectively. In general, since the transmission cost is proportional to the number of hops in the link, TC_{MP} and TC_{PN} can be expressed as $\rho\theta_U$ and $l_{PN}\theta_U$, respectively. (l_{PN} is the average hop counts between PAR and NAR. θ_U and ρ are the unit update cost and the weighing factor for the wireless link, respectively.)

$$M_f = 2TC_{MP} + 2TC_{PN} + PC_N + 2PC_P$$
$$M_s = 4TC_{MP} + 3TC_{PN} + 2PC_N + 3PC_P$$

M_s and M_f only take into account the additional costs incurred when Fast Handover is used. Along with these two costs, the normal signaling cost associated with basic MIPv6 handover also needs to be considered. We denote the signaling cost of the basic Mobile IPv6 scheme as M_n. In other words, C_{signal} in the basic MIPv6 scheme is equal to M_n. Since this signaling cost is common to both Fast Handover and basic Mobile IPv6, we assume the normal case signaling cost (M_n) as being a specific constant in our numerical analysis for the simplicity of analysis. In Eq. 2, α, β, and γ are weighting factors.

$$C_{signal} = \alpha M_s P_s + \beta M_f P_f + \gamma M_n \tag{2}$$

Definition of Success Probability (P_s) To calculate the signaling cost referred to in the above section, we need to define the success probability (P_s). According to Figure 3, t_{L2} is the time taken from the occurrence of the L2 trigger event to the start of the real L2 switching process. Therefore, a small t_{L2} value means that the L2 trigger occurred at a time which was close to the start of the real link switching process. In this case, the value of P_s, which indicates the probability that L3 handover really occurs after the L2 trigger, will be larger. Therefore, we assume the success probability to be dependent on the timing of the L2 trigger. Taking this relationship between t_{L2} and P_s into consideration, we used the following equation for the determination of the success probability. In Eq. 3, τ is a decreasing factor, which is introduced to consider a variety of decreasing patterns.

$$P_s = \frac{1}{e^{\tau t_{L2}}} \tag{3}$$

According to Eq. 3, the success probability is inversely proportional to $e^{\tau t_{L2}}$. In the ideal case, where t_{L2} is 0, the success probability will be 1. On the other hand, if the value of t_{L2} approaches ∞, the success probability will be 0. This means that handover anticipation is done too early without any exact information as to the link condition. Figure 4 shows the relationship between P_s and L_2 as τ is changed.
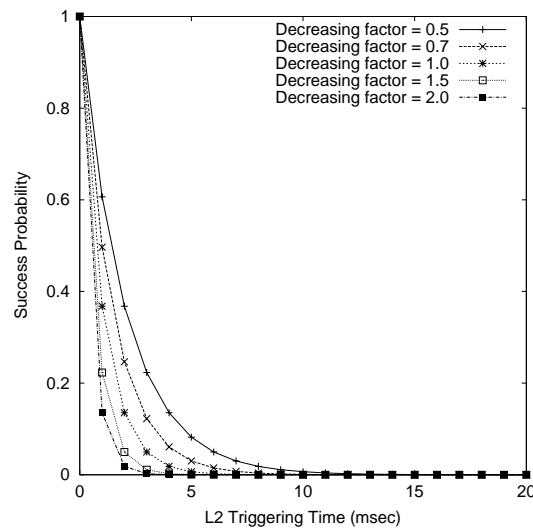
Fig. 4. The determination of P_s as t_{L2}

3.2   Packet Delivery Cost

In terms of the packet delivery cost, we consider the costs associated with both
the forwarded packets (*forwarding cost*) and the lost packets (*loss cost*). Eq. 4
shows the packet delivery cost consisted of the forwarding cost and the loss cost,
where δ and ϵ are weighting factors.

$$C_{packet} = \delta C_{forwarding} + \epsilon C_{loss} \tag{4}$$

In Fast Handover, packet forwarding from PAR to NAR is supported to avoid
packet losses and to support smooth handover. Using the tunnel established after
RtSolPr/PrRtAdv and HI/HACK message exchanges, PAR starts the packet
forwarding to NAR, after receiving the F-BU message sent by the MH just
prior to its disconnecting its link. As shown in Figure 3, there is a delay before
the MH begins receiving packets directly at the NCoA in the NAR ($t_L + t_I +
t_{BU} + t_{New}$). We assume that if an MH sends an F-BU message to the PAR,
then it will definitely start L3 handover to the NAR with no exceptions. In
this paper, we consider the forwarding cost as the additional buffer space used
by the forwarded packets during this period. Therefore, the forwarding cost
($C_{forwarding}$) is proportional to the packet arrival rate (λ_p) and the forwarding
time ($t_L + t_I + t_{BU} + t_{New}$).

In general, since Fast Handover buffers all incoming packets after the L2
trigger, no packets are lost during the handover procedure. However, if an MH
moves to another AP's area before establishing a forwarding tunnel, some pack-
ets may be lost. Specifically, in Fast Handover, a tunnel is established after the
RtSolPr/PrRtAdv and HI/HACK message exchanges. t_{L2} denotes the time pe-
riod from the L2 trigger to the starting point of link switching and t_R denotes
the time required to establish the tunnel. We can obtain the value of t_R using an
analytic method similar to that described in [5]. In most cases, since t_{L2} is larger
than t_R, packets received during the handover procedure are forwarded to NAR
using the already established tunnel. However, in the case of fast moving MHs,
t_{L2} may be so small. If t_{L2} is less than t_R, packets arriving at PAR during the
($t_R - t_{L2}$) period may be lost, because the tunnel is not yet established. There-
fore, the cost of packet losses (C_{loss}) can be expressed as $\lambda_p max\{(t_R - t_{L2}), 0\}$.

In short, the packet delivery cost (C_{packet}) in Fast Handover can be expressed
as Eq. 5.

$$C_{packet} = \delta\lambda_p(t_L + t_I + t_{BU} + t_{New}) + \epsilon\lambda_p max\{(t_R - t_{L2}), 0\} \tag{5}$$

On the other hand, packet delivery cost in the basic Mobile IP model can
be obtained as Eq. 6. In basic Mobile IP, we assume that packet forwarding
begins after the Neighbor Discovery procedure. Therefore, the forwarding time
is $t_{BU} + t_{New}$. Also, since no handover anticipation is supported in the basic
Mobile IP model, all packets arriving during $t_L + t_I$ will be lost.

$$C_{packet} = \delta\lambda_p(t_{BU} + t_{New}) + \epsilon\lambda_p(t_L + t_I) \tag{6}$$

4   Numerical Results

As mentioned above, Fast Handover is initiated by the L2 trigger. The L2 trigger signals an layer 2 event to layer 3. In the handover procedure, an L2 event can occur due to a change in the signal strength of one or more APs. Therefore, the time when layer 3 is notified of the L2 event is determined by the threshold value of the signal strength used in the protocol implementation. In this paper, we focus on the optimal L2 trigger timing, required to minimize total overhead cost. This value is important in the implementation of Fast Handover, because the total cost is largely dependent on this value.

4.1   Impact of L2 Triggering Timing

In the first analysis, we consider only the influences of t_{L2} and assume that other latency values are determined in advance by the corresponding message procedures. Table 1 shows the parameter values used in the numerical analysis. In this analysis, t_R is set to 10ms. $t_L + t_I$ and $t_{BU} + t_{New}$ are set to 165ms and 320ms using reference values defined in [4].

Table 1. System Parameters for Numerical Analysis

α	β	γ	δ	ϵ	θ_U	ρ	l_{PN}	PC_P	PC_N	λ_p	M_n
0.2	0.8	1.0	0.2	0.8	1	10	2	5	5	1	100

(a) Signaling Cost   (b) Packet Deliver Cost
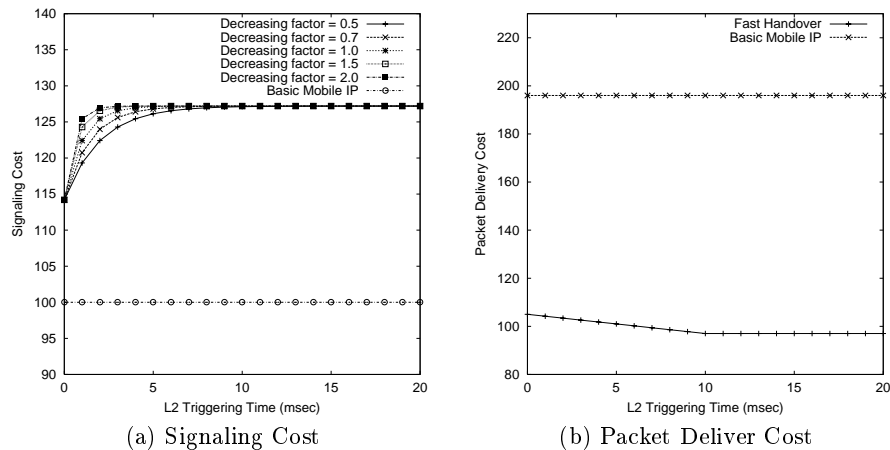
Fig. 5. Cost Comparison: Fast Handover vs. Basic Mobile IP

Figure 5 shows the signaling cost and packet delivery cost as t_{L2} is changed. In terms of the signaling cost, the basic Mobile IP exhibits a constant signaling cost as a function of the L2 triggering time. On the other hand, the signaling cost in Fast Handover increases as the L2 triggering time increases. The rate of increase of the signaling cost is determined by the value of the decreasing factor which is used. However, the signaling cost no loger increases once the L2 triggering time attains a certain critical value. These values are different from the values of decreasing factors which are used.

In contrast to the signaling cost, the packet delivery cost is not dependent on the value of the success probability (P_s). In terms of the packet delivery cost, the basic Mobile IP imposes a higher packet delivery cost than Fast Handover. This is because the handover latency in the basic Mobile IP is larger than that of Fast Handover. Although additional packet loss can occur due to the hasty handover anticipation which sometimes takes place in Fast Handover, the amount of packet loss is negligible compared to that caused by handover latency in basic Mobile IP.

Figure 6 shows the total costs incurred in Fast Handover and basic Mobile IP. In this figure, when the value of t_{L2} is about 10ms, the total cost is minimized. Namely, the total cost is minimized when t_{L2} is equal to t_R. This fact can be proved theoretically using the defined cost functions (refer to Appendix). Therefore, it is necessary to adjust t_{L2} to be close to the value of t_R for exact handover anticipation and optimized handover performance.
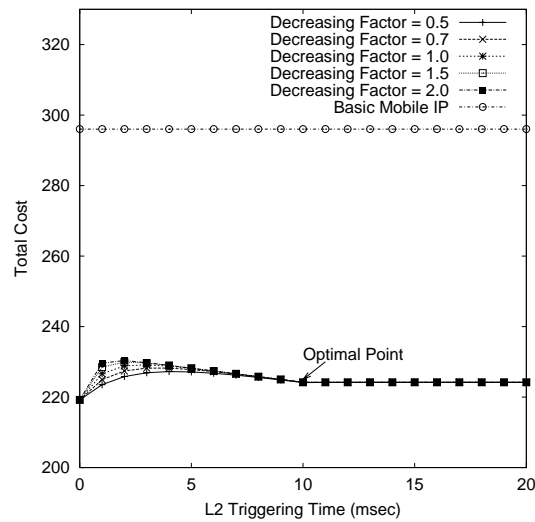
Fig. 6. Total Cost in Fast Handover and Basic Mobile IP

4.2   Buffer Space Analysis

Fast Handover anticipates whether or not an MH is likely to move to another cell area using L2 information. To avoid packet losses during the anticipation period, Fast Handover should support packet forwarding during an extended time period compared with basic Mobile IP. In this section, we analyze the buffer requirement of Fast Handover and basic Mobile IP. In this analysis, we assume the average packet size is 200 bytes and consider the buffer space required for one mobile host's handover procedures. Figure 7 shows the required buffer space in Fast Handover and basic Mobile IP. As mentioned above, since Fast Handover begins packet forwarding at an earlier time than basic Mobile IP, it requires more buffer spaces. The buffer space required increases in proportion to the packet arrival rate. Also, the buffer space required increases according to the number of mobile hosts performing handover procedures. In addition, the buffering time may affect the on-time delivery of data packets, especially in real-time applications. For example, if some packets are stored in buffer during a longer time period than the admissible end-to-end delay, they may become useless. Therefore, it is essential to manage the forwarding buffer efficiently to support Fast Handover with minimum overhead and provide more scalable services.
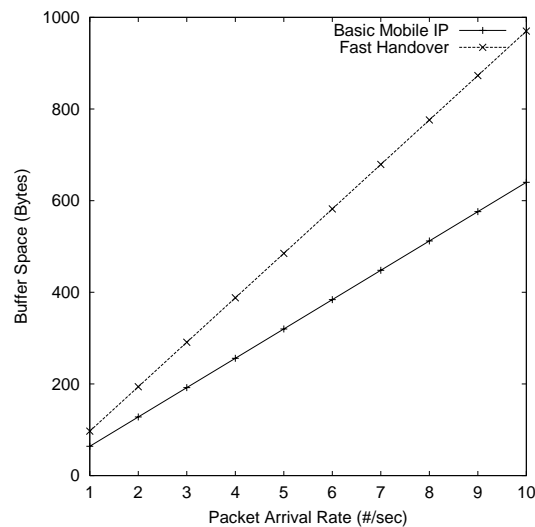

Fig. 7. Buffer Requirement: Basic Mobile IP vs. Fast Handover

5   Conclusion

In IP-based wireless/mobile networks, minimizing handover latency is one of the most important issues. Compared with the existing protocols, Fast Handover has several advantages in terms of its ability to reduce handover latency and support smooth handover. However, this is only achieved at a certain cost in terms of additional overhead.

In contrast to the results published in previous reports, in this paper, we presented several analysis results of the overhead costs incurred in Fast Handover. First, we classified overhead costs into signaling cost and packet delivery cost. Then, we compared and evaluated these two costs both in Fast Handover and in basic Mobile IPv6. In the first analysis, we studied the change in the overhead costs as a function of the L2 triggering time. Based on the results of this study, we found that the cost is minimized when t_{L2} is equal to t_R. For this condition to be met, a more exact notification of the L2 trigger is required. In other words, the support of the lower layers is essential for optimized handover performance. In terms of the buffer space required for packet forwarding, efficient buffer management and optimized forwarding timing are needed to minimize buffer management overhead and to support real-time applications.

The analysis proposed in this paper can be utilized in the determination of the signal threshold value for the L2 trigger, and as a reference for the design of buffer management schemes, which can be used in access routers designed for smooth handover.

References

1. Rajeev Koodli: Fast Handovers for Mobile IPv6. Internet Draft, draft-ietf-mobileip-fast-mipv6-06.txt, Work in Progress, March 2003.
2. E. Gustafsson, A. Jonsson, and C. Perkins: Mobile IP Regional Registration. Internet Draft, draft-ietf-mobileip-reg-tunnel-02, Work in Progress, March 2000.
3. H. Soliman, C. Castelluccia, K. E. Malki, and L. Bellier: Hierarchical Mobile IPv6 mobility management (HMIPv6). Internet Draft, draft-ietf-mobileip-hmipv6-07.txt, Work in Progress, July 2000.
4. Rajeev Koodli and Chales E. Perkins: Fast Handovers and Context Transfers in Mobile Networks. ACM Computer Communication Review, Vol. 31, No. 5, October 2001.
5. X. P. Costa, R. Schmitz, H. Hartenstein, and M. Liebsch: A MIPv6, FMIPv6 and HMIPv6 Handover Latency Study: Ananlytic Approach. Proc. of IST Mobile & Wireless Telecommunications Submit, June 2002.
6. N. A. Fikouras, K. El Malki, S. R. Cvetkovic: Performance Analysis of Mobile IP Handoffs. Proc. of Asia Pacific Microwave Conference (APMC), December 1999.
7. C. Blondia, O. Casals, Ll. Cerda, N. Wijngaert, G. Willems, P. Cleyn: Performance Comparison of Low Latency Mobile IP Schemes. Proc. of Modeling and Optimization in Mobile Ad Hoc and Wireless Networks (WiOpt 2003), March 2003.

Appendix: Proof of the Optimal L2 Triggering Time

To obtain the optimal L2 triggering time minimizing the total cost, we formulated the total cost shown in Eq. 1 as the function of t_{L2} as follows:

$$C_{total}(t_{L2}) = \alpha M_s P_s + \beta M_f P_f + \gamma M_n + \delta \lambda_p (t_L + t_I + t_{BU} + t_{New}) + \varepsilon \lambda_p max\{(t_R - t_{L2}), 0\} \tag{7}$$

A. $0 < t_{L2} < t_R$

Differentiating $C_{total}(t_{L2})$ with respect to t_{L2}, and set it to zero, we obtain the stationary equation

$$C'_{total}(t_{L2}) = -\tau \alpha M_s e^{-\tau t_{L2}} + \tau \beta M_f e^{-\tau t_{L2}} - \varepsilon \lambda_p = 0 \tag{8}$$

Then, the root (t_{L2}^*) of the equation shown in Eq. 8 is

$$t_{L2}^* = \frac{\log(\tau \beta M_f - \tau \alpha M_s) - \log(\varepsilon \lambda_p)}{\tau} \tag{9}$$

If t_{L2} is smaller than t_{L2}^*, $C'_{total}(t_{L2})$ is larger than zero. On the other hand, if t_{L2} is larger than t_{L2}^*, $C'_{total}(t_{L2})$ is smaller than zero. In other words, the total cost is maximized when t_{L2} is equal to t_{L2}^* in the period of $[0, t_R)$. In addition, the total cost is minimized when t_{L2} approaches to t_R, because $C_{total}(0)$ is larger than $\lim_{t_{L2} \to t_R} C_{total}(t_{L2})$.

B. $t_{L2} > t_R$

Differentiating $C_{total}(t_{L2})$ with respect to t_{L2}, we obtain the following equation.

$$C'_{total}(t_{L2}) = -\tau \alpha M_s e^{-\tau t_{L2}} + \tau \beta M_f e^{-\tau t_{L2}} \tag{10}$$

In the period of $[t_R, \infty)$, $C'_{total}(t_{L2})$ is larger than zero. Namely, $C_{total}(t_{L2})$ is a monotone increasing function in that period. Therefore, the total cost is minimized when when t_{L2} is equal to t_R.

Conclusively, the total cost is minimized when t_{L2} is equal to t_R (note that $C_{total}(t_{L2})$ is a continuous function), which is the required time for establishing a bidirectional tunnel between NAR and PAR.