A Study on Optimal Fast Handover Scheme in Fast Handover for Mobile IPv6 (FMIPv6) Networks

Byungjoo Park[1], Youn-Hee Han[2], and Haniph Latchman[1]

[1] Department of Electrical and Computer Engineering,
University of Florida, Gainesville, USA
{pbj0625, latchman}@ufl.edu
[2] School of Internet-Media, Korea University of Technology and Education,
Cheonan-Si, Chungnam, 330-708, Korea
yhhan@kut.ac.kr

Abstract. Mobile IPv6 (MIPv6) is protocol for handling routing of IPv6 packets to mobile nodes that have moved away from their home network. However, in MIPv6, the handover process reveals numerous problems manifested by a time-consuming network layer based movement detection and latency in configuring a new care-of address, with confirmation scheme called duplicate address detection (DAD). To reduce the handover latency in standard MIPv6, the fast handover for Mobile IPv6 (FMIPv6) protocol is proposed. To do this, each time a mobile node moves to a new location, it configures and confirms its temporal IP address during layer 2 handover. In this paper, we study the impact of the address configuration and confirmation procedure on the FMIPv6's IP handover latency. A mathematical analysis comparing the various parameters is provided to show the benefits of our scheme over the current procedure like standard MIPv6 and FMIPv6.

1 Introduction

To accommodate the increasing demand of mobility in the Internet, Mobile IPv6 (MIPv6) has been proposed in the IETF [1]. The protocol provides seamless connectivity to MNs when they move from one wireless point of attachment to another in a different subnet. According to the proposal, a mobile node (MN) should generate a new care-of address (CoA) by using IPv6 stateless address auto-configuration whenever it moves to new link. To verify the uniqueness of this CoA, it should run duplicate address detection (DAD) algorithm [2] before assigning the address to its interface. The algorithm determines if the address chosen by an MN is already in use. MN must perform DAD every time it handovers between IPv6 networks and cannot begin communication until DAD completes. According to the current RFC 2462 DAD algorithm, it takes at least 1000ms to detect that there is no duplicate address in the link. After finishing DAD procedure, the MN has to wait for random delays for router solicitation message (RS) and router advertisement message (RA) [2,3].

F. Stajano et al. (Eds.): ICUCT 2006, LNCS 4412, pp. 120–129, 2007.

Fast handovers for Mobile IPv6 (FMIPv6) [4], has been proposed in IETF to reduce the handover latency in standard MIPv6. This proposal describes a protocol to replace such new care-of address (NCoA) configuration procedure. It enables MN to quickly detect that it is now moving to a new subnet by providing the new access point (NAP) identifier and receiving the associated subnet prefix information. MN formulates a prospective NCoA, if at all possible, when still present on current subnet. Furthermore, in order to make MN allocates NCoA to its interface immediately after attaching to new subnet, FMIPv6 allows the NCoA confirmation procedure to be executed before or while MN switches its subnet. The scenario in which an MN receives the positive result about the confirmation of its prospective NCoA on the current subnet is called predictive mode. The scenario in which an MN checks the uniqueness of NCoA after MN attaches to a new subnet is called reactive mode. Although MN initiates the NCoA confirmation at an early time on the current subnet, FMIPv6 would fall into reactive mode if MN could not receive the confirmation result on the current subnet. In addition, if the proposed NCoA is rejected during the NCoA confirmation procedure, MN may configure NCoA by itself after movement so that handover latency becomes long. In order to achieve more reduction of handover latency, it is required that predictive mode should occur more frequently than reactive mode. So, it is necessary that the NCoA confirmation should be done promptly and its result should be always successful. However, a proper confirmation method has not been provided.

In this paper, we propose new movement detection, address configuration and confirmation scheme (*NAC*) in FMIPv6 networks that remarkably takes off the DAD procedure from the whole layer-3 handover procedure, thereby reducing layer-3 handover latency.

The reminder of this paper is organized as follows. Section 2 introduces our proposed FMIPv6 with *NAC* scheme. The performance evaluations and comparisons in MIPv6, FMIPv6 and proposed FMIPv6 with *NAC* scheme are shown in section 3. Finally we present the conclusion in section 4.

2 New Address Configuration and Confirmation Algorithm in Fast Handover for Mobile IPv6 (*NAC*)

In this section, we describe our new optimized address configuration and confirmation scheme called *"NAC"* to reduce total handover latency. We can define the handover procedure like movement detection, NCoA configuration and confirmation (DAD) procedure.

2.1 New Fast Movement Detection in FMIPv6

In FMIPv6, the movement detection is based on an indication from a wireless Layer 2 (L2) trigger which informs that MN will soon be handover. First, we assume that the L2 trigger signaling message from NAP includes stored router advertisement (SRA) message based on EAP [5,6]. To begin a fast handover, an

MN sends the router solicitation for proxy (RtSolPr) message to the previous access router (PAR).

The RtSolPr contains the L2 identifier of a target AP which the MN will move to. At this time, PAR starts to map the L2 identifier into proper target new access router (NAR). In response, The MN will receive the proxy router advertisement (PrRtAdv) message from PAR. Based on SRA and PrRtAdv messages, the MN compares the prefix of the SRA message with existing prefixes in the cache. If the prefix is different, the MN starts to configure a prospective NCoA using the IPv6 stateless (or stateful) address auto-configuration method. And then, the MN immediately sends the fast binding update (FBU) message with the prospective NCoA. When PAR receives FBU message, it sends the modified new handover initiation (*NHI*) carrying a newly define 1-bit D-flag, named "*NCoA DAD Request bit (D bit)*" to NAR, which validate the MN's new CoA and initiates the process of establishing a bidirectional tunnel between the PAR and the MN at its NCoA. This *NHI* message contains the "*previous MN's CoA*" and "*previous AR's global address*" to support interoperability with normal nodes by using a bit in the reserved field.

2.2 New NCoA Configuration and DAD Scheme in FMIPv6

To reduce DAD processing delay, we propose new NCoA configuration and DAD scheme using modified neighbor cache in NAR. This modified neighbor cache supports new enhanced lookup algorithm which can reduce DAD processing delay from 1000 ms to 5.28 μ sec. That is, the DAD using lookup algorithm consumes an extremely short amount of time, typically a few micro second units, such as Longest Prefix Matching speeds in routing table.

In the current FMIPv6, there is no specific address confirmation scheme. So, in our paper, we assume that RFC 2462 DAD is also used for the confirmation scheme. If the period of address confirmation procedure is long, then the delivery of handover acknowledgement (HAck) message and fast binding acknowledgement (FBAck) message would be delayed. That is, the MN can not receive FBAck before it disconnects with PAR. This means FMIPv6 could fall into the reactive mode and the MN has to resend FBU message as soon as it attaches to NAR. As a result, at this case, it requires to deliver an additional FBU message, which will be encapsulated in fast neighbor advertisement (FNA), with the consumption of wireless bandwidth. On the other hand, if NAR receives FNA and an encapsulated FBU, and detects that NCoA is duplicated, it must discard the inner FBU and notify this fact of MN. It will cause handover latency to be extended. This kind of case can occur even when the period of confirmation procedure is very short. If the result of confirmation shows that the prospective NCoA is invalid, the MN should itself configure its NCoA and run RFC 2462 DAD after moving to NAR.

However, if the NAR adopts proposed *NAC* scheme, these problems can be obviously removed. After receiving *NHI* message, NAR starts new DAD procedure using a lookup algorithm in modified neighbor cache in NAR. As soon as DAD procedure finishes, the NAR can unicast the HAck message to the PAR.

This HAck could also be modified like the NHI message by adding a 2-bit F-flag to the reserved flag and containing the "*New MAC address*", "*New link-local address*" and "*NCoA DAD Reply option*" in the option field in case of address duplication. We name this new HAck message as '*NHA*'. Table 1 defines the F-flag in proposed FMIPv6.

Table 1. The F-Flag of NHA message

F-Flag	Mean
00	Must change MAC address. (Can not apply in IEEE-802 case)
01	Can allocate a link-local address and a new CoA
10	Must change the link-local address allocated into the Alternative Address
10	Can not use

After the PAR receives *NHA* message from NAR, it sends FBAck message carrying the *NHA* message's "*NCoA DAD Reply option*" to MN. As soon as the MN receives the FBAck, it configures the address specified in the NCoA DAD Reply option into its interface. In the proposal, it takes a very short time to configure and confirm NCoA such as 5.28 μ sec in worst case. Also, in the proposed scheme, NCoA does not become invalid, since the unique NCoA is provided by NAR.

2.3 Lookup Procedure for Fast DAD Procedure in FMIPv6

We denote t_{LU} as the address lookup delay, which is the time required to check an MN's MAC address for movement detection and DAD in the Patricia Trie search. Accordingly the address lookup delay (t_{LU}) is given as:

$$t_{LU} = t_{DAC} \cdot N \tag{1}$$

Where t_{DAC} is the delay for access and comparison operations in RAM and N is the number of lookups in Patricia Trie. This Patricia Trie has the worst performance in line per minute (LPM). We use this algorithm in order to show the lookup time of the worst performance. Under the present circumstance, since a memory access requires from 60 to 100 nsec [7] and a comparison requires 10nsec in DRAM [8], we can use the value of t_{DAC} as 70 and 110nsec. In the Patricia Trie case, since lookups require accessing memory 48 times in the worst case, the N value is 48. Hence, t_{LU} is 3.36 μ sec and 5.28 μ sec and the calculated lookup delay is very small.

We describe the analysis method by using the queuing system. We assume that arrival packets are stored in the buffer and processed by the FIFO policy. We also suppose that the packet interarrival times can be modeled by a poisson process. Then, we use an M/G/1 queuing model to calculate the average performance of the MAC address in lookup algorithm. We denote λ_p as the *NHI* packet arrival rate at the AR. An average of lookup processing time ($E[t_{LU}]$) is determined according to the corresponding neighbor cache lookup delays and the probability density of addresses determined by the memory access times. We define the traffic intensity φ :

$$\varphi = \lambda_p \cdot E[t_{LU}] \tag{2}$$

The traffic intensity φ is the quantity that governs the stability of the system. Let us introduce *LU* as the lookup delay, which is defined as the time duration from when an *NHI* packet arrives at the AR to when an *NHA* message is forwarded to the output link. By applying the M/G/1 queuing model, the mean lookup processing delay is derived by

$$E[LU] = E[W] \cdot E[t_{LU}] \tag{3}$$

Where $E[W]$ is the expected mean waiting time of a packet in queue. Using the Pollaczek-Khinchin (P-K) formula, the mean waiting time is derived by

$$E[W] = \left(\frac{\lambda_p \cdot E[t_{LU}]^2}{2(1-\varphi)} \right) = E[t_{LU}] \cdot \left(\frac{\varphi}{1-\varphi} \right) \tag{4}$$

Where C_B^2 denotes the squared coefficient of variation of the processing time. An important observation is that, clearly, the mean waiting time only depends upon the first two moments of the lookup processing time.

3 Performance Analysis

In this section, we will calculate the handover latency per movement for each protocol. Handover latency is defined for a receiving MN as the time that elapses between the disconnection with the previous attachment of point and the arrival of the first packet after the MN moves to NAR. We use a simple model for the data packet traffic, although the self similar nature of it has been noticed. Our packet traffic model has two layers namely session and packets. During a session, several packets are generated by a CN at an arbitrary rate and they reach an MN at the same rate. We assume that the session duration time has the exponential distribution with mean $E[t_o] = 1/\lambda_o$.

3.1 Network System Model and Mobility Model

We assume that a homogeneous network of which all wireless AP areas in a subnet domain have the same shape and size. First, we can define some parameters used for performance analysis. Let t_s and t_p be i.i.d. random variables representing the subnet domain residence time and the AP area residence time, respectively. Let $f_s(t)$ and $f_p(t)$ be the density function of t_s and t_p, respectively. In our paper, we suppose that an MN visit k AP areas in a subnet domain for a period t_s^k. During t_s^k, the MN resides at AP area i for a period t_i.

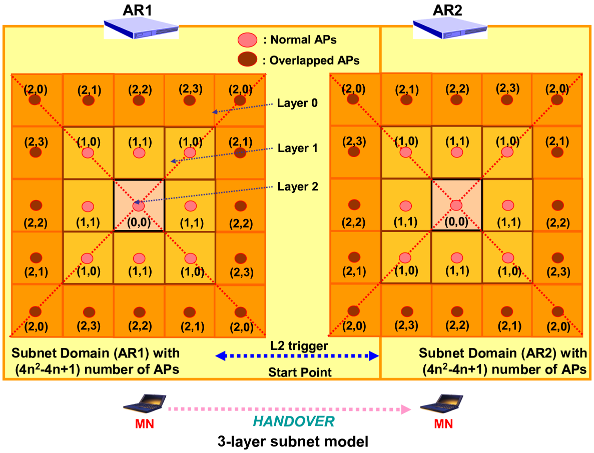


Fig. 1. 3-Layer Subnet Area Structure

Then, $t_s^k = t_1 + t_2 + t_3 + \cdots + t_{k-1} + t_k$ has the following density function

$$f_s^{(k)}(t) = \int_{t_1=0}^{t}\int_{t_2=0}^{t-t_1}\cdot\cdot\int_{t_{k-1}=0}^{t-t_1-\cdots-t_{k-2}} f_p(t_1)f_p(t_2)f_p(t_3)$$

$$\cdots f_p(t_{k-1})f_p(t-t_1-\cdots-t_{k-1})dt_{k-1}\cdots dt_2dt_1. \tag{5}$$

Using the Laplace transform convolution, we can determine the Lasplace transform for $f_s^{(k)}(t)$ as follows:

$$f_s^{(k)*}(s) = [f_p^*(s)]^k. \tag{6}$$

where $f_p^*(s)$ is the Laplace transform of $f_p(t)$.

We describe a two-dimensional random walk model for mesh planes in order to compute the subnet domain residence time density function. Our model is similar to reference [9] and considers a regular AP area/subnet domain overlay structure. We assume that an MN resides in an AP area for a period and moves to one of its four neighbors with the same probability, i.e. with probability 1/4. A subnet is referred to as a *n-layer* domain if it overlays with $N = 4n^2 - 4n + 1$ AP areas.

Fig. 1 shows the *3-layer* subnet domain architecture in which each of the 25 small squares and the entire square represents each of the AP areas and one subnet domain area, respectively. The AP area at the center of the subnet is called the *layer 0 AP area*. The AP areas that surround *layer x-1* AP areas are called *layer x* AP areas.

There are $8x$ AP areas in *layer x* except exactly one AP area which is in *layer 0*. An *n-layer* subnet overlays AP areas from *layer 0* to *layer n-1*. Particularly the AP areas that surround the *layer n-1* AP areas are referred to as boundary neighbors, which are outside of the subnet. According to the equal moving probability assumption, we classify the AP areas in a subnet domain into several AP area types. An AP area type is of the form $< x, y >$, where x indicates that the AP area is in *layer* x and y represents the $y+1$st type in *layer* x. AP areas of the same type have the same traffic flow pattern because they are at the symmetrical positions on the mesh domain. For example, in Fig. 1, the AP type $< 1, 1 >$, $< 2, 1 >$ represent that this AP is in ring 1 and ring 2 and it is the AP of 2nd type in ring 1 and ring 2, respectively.

In the random walk model, a state (x, y) represents that the MN is in one of the AP areas of type $< x, y >$. The absorbing state (n, j) represents that an MN moves out of the subnet from state $(n-1, j)$, where $0 \leq j \leq 2n-3$. We assume that the AP area residence time of an MN has a Gamma distribution with mean $1/\lambda_p$ ($=E[t_p]$) and variance ν. The Gamma distribution is selected for its flexibility and generality. The Laplace transform of a Gamma distribution is

$$f_p^*(s) = \left(\frac{\gamma\lambda_p}{s+\gamma\lambda_p}\right)^{\gamma}, \qquad where\ \gamma = \frac{1}{\nu\lambda_p^2}. \tag{7}$$

Also, we can get the Laplace transform $f_s^*(s)$ of $f_s(t)$ and its expected subnet domain residence time $E[t_s]$ from [9]. For an MN, in the end, the probabilities $\Pi_p(i)$ and $\Pi_s(j)$ that the MN moves across i AP areas and j subnets during a session duration, can be derived as follows [10]:

$$\Pi_p(i) = \begin{cases} 1 - \frac{E[t_o]}{E[t_p]}(1 - f_p^*(\frac{1}{E[t_o]})) & , i = 0 \\ \frac{E[t_o]}{E[t_p]}(1 - f_p^*(\frac{1}{E[t_o]}))^2 (f_p^*(\frac{1}{E[t_o]})^{i-1} & , i > 0 \end{cases} \tag{8}$$

$$\Pi_s(j) = \begin{cases} 1 - \frac{E[t_o]}{E[t_s]}(1 - f_s^*(\frac{1}{E[t_o]})) & , j = 0 \\ \frac{E[t_o]}{E[t_s]}(1 - f_s^*(\frac{1}{E[t_o]}))^2 (f_s^*(\frac{1}{E[t_o]})^{j-1} & , j > 0 \end{cases} \tag{9}$$

3.2 Handover Latency Comparisons

At first, we introduce distance parameters used for handover latency functions. t_{WD} is the wireless component of the delay for a new AP re-association and authentication latency (MN's switching delay between APs). t_{RS} and t_{RA} are the transmission delays for the RS/RA messages in standard MIPv6, ($t_{RS} + t_{RA} = 2t_R$). $^*t_{RD}$ is the random delay for RS, RA defined as the RFC 3775 ($^*t_{RD} = t_{RD\_RS} + t_{RD\_RA}$).

t_{BU} and t_{BAck} are the transmission delays for *BU*/*BAck* messages respectively ($t_{BU} + t_{BAck} = 2t_B$). t_{Packet} is the packet transmission delay from CN to MN. t_{DAD} is the DAD processing delay defined as the RFC 2462. t_{LU} is the lookup delay for DAD. ζ is the weighting factor of packet tunneling. ψ is the total delay between the time to exchange FBU/FBAck and the time of disconnection (Link-Down) with the current AP. t_{RS_FNA} and t_{RA_NAAck} are the transmission delays for RS with Fast Neighbor Advertisement and RA with Neighbor Advertisement Acknowledgment ($t_{RS\_FNA} + t_{RA\_NAAck} = 2t_{FR}$). $t_{NHI/HI}$ and

$t_{NHA/HAck}$ are the transmission delays for new NHI/NHA messages in proposed FMIPv6 with *NAC* and regular HI/HAck messages in standard FMIPv6 for address confirmation and to setup the tunnel between PAR and NAR. t_{Packet_MN} is the packet forwarding delay between MN and NAR. t_{Packet_PN} is the buffered packets forwarding delay from PAR to NAR. Using such parameters, for the standard MIPv6, FMIPv6 and proposed MIPv6 with *NAC*, the total handover latency per session duration is defined as follows:

$$T_{MIPv6} = \sum_{i=0}^{\infty}\{i\Pi_p(i) \cdot t_{WD}\} + \sum_{j=0}^{\infty}\{j\Pi_s(j) \cdot (2t_R +^* t_{RD} + t_{DAD} + 2t_B + t_{Packet})\} \tag{10}$$

In FMIPv6, MN sends FBU to PAR prior to disconnection with PAR. At this time, the handover procedure of FMIPv6 is divided into two independent procedures; H_I, the procedure to be executed by MN itself with PAR and NAR, and H_{II}, the procedure to be executed by both PAR and NAR to establish the bidirectional tunnel. The two separated procedures will combine into one when NAR receives FNA from MN after MN's subnet movement. We first assume that NAR has already received at least HI from PAR, when it receives FNA from MN. Before the two procedures H_I and H_{II} combine into one, the completion times of each procedure are defined as follows:

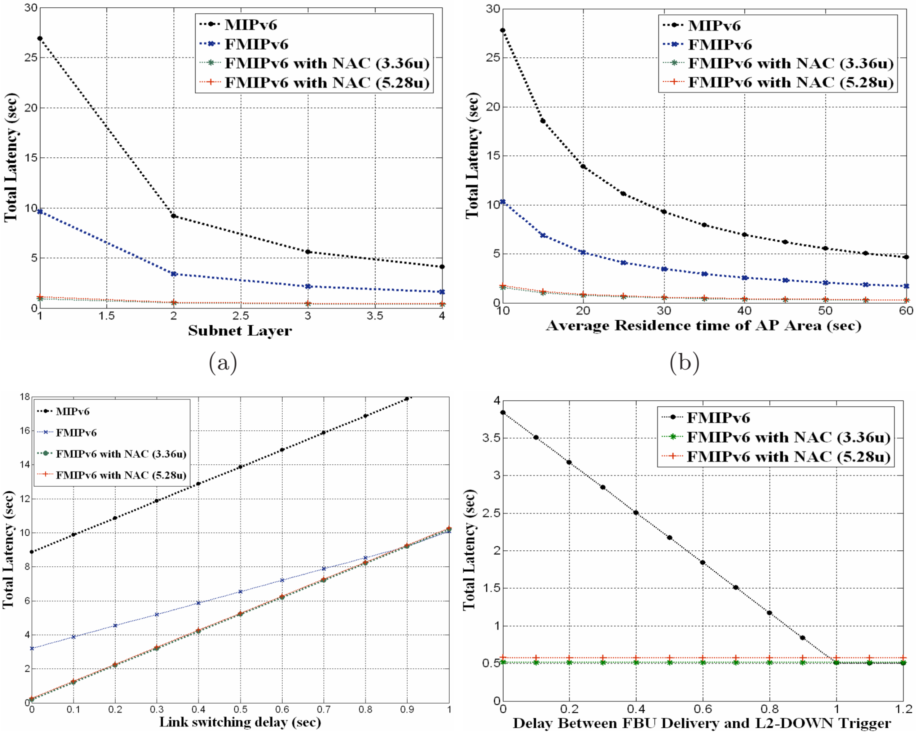
$$T_{H_I} = \psi + t_{WD} + t_{RS\_FNA} + t_{RA\_NAAck} \tag{11}$$

$$T_{H_{II}} = (t_{NHI/HI} + t_{NHA/HAck} + \zeta \cdot t_{Packet\_PN}) + t_{LU/DAD} \tag{12}$$

If H_{II} finishes before the completion of H_I(that is, $T_{H_I} > T_{H_{II}}$), NAR has buffered the packets tunneled from PAR and forwards them to MN when it receives FNA. if not, NAR waits the packets which will be tunneled from PAR when it receives FNA. At the latter case, NAR have to wait the completion of the address confirmation procedure. After announcing its attachment to NAR and receiving the tunneled packets, MN sends binding update messages with its new CoA to HA, and to CNs consecutively. In FMIPv6, the total handover latency per a session time are defined in Eq. 13. In our paper, the SRA message and L2 information are triggered together with an association response message. We assume that $t_{NHI/HI}$, $t_{NHA/HAck}$, t_{RS_FNA} and t_{RA_NAAck} have the same value in transmission time. Also, t_{BU} and t_{BAck}have the same value in transmission time. In our proposed scheme, t_{LU} is the most important factor determining the performance.

$$\begin{aligned} T_{FMIPv6} = & \sum_{i=0}^{\infty}\{i\Pi_p(i) \cdot t_{WD}\} \\ & + \sum_{j=0}^{\infty}\{j\Pi_s(j) \cdot (MAX\{T_{H_I}, T_{H_{II}}\} + \zeta \cdot t_{Packet\_MN} - t_{WD} - \psi)\} \end{aligned} \tag{13}$$

And, from the above function with $t_{LU} = 3.36\ \mu$ sec and $5.28\ \mu$ sec in Eq.12, we can get the handover latency for the enhanced FMIPv6 equipped with *NAC* scheme.



(a) (b)

(c) (d)

Fig. 2. Total Handover Latency Comparison

3.3 Numerical Results

For examinations, the following fixed parameters are used: $t_{RS/RA} = 0.015$, $^{*}t_{RD} = 1.5$, $t_{DAD} = 1$, $t_{BU/BAck} = 0.065$, $t_{NHI/HIandNHA/HAck} = 0.01$, ν=1.0, $\zeta = 1.2$, λ_0=0.0033 (session duration is 300sec) and $t_{Packet/Packet\_MN/Packet\_PN}$ = 0.065/0.015/0.01. As the target of investigation, we select the following changeable parameters and their default values: n=*2* (subnet layer is 2), λ_p(mean of AP area residence time is 30 sec.), t_{WD} = 0.03sec., and ψ = 0.1. While we select one parameter and change its value, the remaining parameters values are set to their default values during the following investigation. Fig. 2 explains the total handover latency per session duration with respect to each changeable parameter. From the figures, we can know that proposed FMIPv6 with *NAC* handover latency are considerably reduced the address configuration and confirmation process.

Fig. 2 (a) shows the total handover latency of each protocol with respect to the subnet layer. It shows that the reduction of latency becomes high when a subnet contains many AP areas. The figure show that proposed FMIPv6 with *NAC* is under little influence of such system deployment. Fig. 2 (b) shows that the handover process occupies much time within the whole session duration when

MN moves across AP areas and subnets more frequently. Fig. 2 (c) shows the relationship between the handover latency and the delay of link switching in a session duration. When the switching delay (t_{WD}) becomes high, all protocols' handover latency become high, too. When the link switching delay is 1, the procedure H_I becomes the dominant factor of handover latency. Fig. 2 (d) shows that FMIPv6's handover latency becomes low if MN sends FBU to PAR more early before it disconnects with PAR. If FBU can be delivered to PAR as soon as possible, NAR receives HI early in FMIPv6 handover process.

4 Conclusion

In this paper, we have introduced the proposed FMIPv6 with NAC. The use of a modified neighbor cache with look up algorithm has merits, such as a faster DAD checking speed, which solves the short-comings of normal DAD when a router has more than two links. We also can obtain alternative addresses by managing addresses in the network. In the numerical analysis, we developed packet traffic, system and mobility models. Based on the numerical results, we can see that the major benefits of our scheme are to remarkably reduce CoA configuration and confirmation latency concerned in any seamless handover schemes, and preventing address collision from occurring provided there is no packet loss.

References

1. D.Johnson, C. Perkins, J. Arkko, "Mobility Support in IPv6", RFC 3775, June 2004.
2. S. Thomson, T. Narten, "IPv6 Stateless Address Auto-configuration", RFC 2462, Dec. 1998.
3. Narten, T., Nordmark, E. and W. Simpson, "Neighbor Discovery for IP version 6 (IPv6)", RFC 2461, December 1998.
4. Koodli, R., "Fast Handovers for Mobile IPv6", RFC 4068, July 2005.
5. Byungjoo. Park, Y-H. Han, H. A. Latchmann, "EAP: New Fast Handover Scheme based on Enhanced Access Point in Mobile IP Networks", International Journal of Computer Science and Network Security, Vol. 6, No.9, pp. 69-75, Sep. 2006.
6. JinHyoeck Choi, DongYun Shin, "Router Advertisement Cashing in Access Point (AP) for Fast Router Discovery", draft-jinchoi-mobileip-frd-00.txt, June 2002.
7. V. Srinivasan, G. Varghese, "Fast Address Lookups Using Controlled Prefix Expansion", ACM Transactions on Computer System, Vol.17, Feb.1999.
8. R. Kawabe, S. Ata, M. Murata, "On Performance Prediction of Address Lookup Algorithms of IP Routers through Simulation and Analysis Techniques", IEEE International Conference on Communications 2002 (ICC 2002).
9. I. F. Akyildiz, Y.B. Lin, W. R. Lai, and R. J. Chen, "A new Random Walk Model for PCS Networks," IEEE JSAC, Vol.18, No.7, pp.1254-1260, July 2000. 10.
10. Y. H. Han, "Hierarchical Location Chacing Scheme for mobility Managament", Dept. of Computer Science and Engineering, Korea University, Dec. 2001.