



A performance comparison of mobility anchor point selection schemes in Hierarchical Mobile IPv6 networks [☆]

Sangheon Pack ^{*}, Taekyoung Kwon, Yanghee Choi

School of Computer Science and Engineering, 301-518, Seoul National University, Shilim-dong, Kwanak-gu, Seoul 151 742, Republic of Korea

Received 7 October 2005; received in revised form 12 July 2006; accepted 16 September 2006

Responsible Editor: X.S. Shen

Abstract

Hierarchical Mobile IPv6 (HMIPv6) introduces a mobility anchor point (MAP) that localizes the signaling traffic and hence reduces the handoff latency. In addition to processing binding update messages from mobile nodes (MNs) on behalf of MNs' home agents (HAs), the MAP performs data traffic tunneling destined to or originated from MNs, both of which will burden the MAP substantially as the network size grows. To provide scalable and robust mobile Internet services to a large number of visiting MNs, multiple MAPs will be deployed. In such an environment, how to select an appropriate MAP has a vital effect on the overall network performance. In this paper, we choose four MAP selection schemes: the furthest MAP selection scheme, the nearest MAP selection scheme, the mobility-based MAP selection scheme, and the adaptive MAP selection scheme. Then, we compare their performances quantitatively in terms of signaling overhead and load balancing. It can be shown that the dynamic schemes (i.e., the mobility-based and the adaptive MAP selection schemes) are better than the static schemes (i.e., the furthest and the nearest MAP selection schemes), since the dynamic schemes can select the serving MAP depending on the MN's characteristics, e.g., mobility and session activity. In addition, the adaptive MAP selection scheme achieves low implementation overhead and better load balancing compared with the mobility-based MAP selection scheme.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Hierarchical Mobile IPv6; Mobility anchor point (MAP); MAP selection scheme; Analytical modeling; Load balancing

1. Introduction

Mobile IPv6 (MIPv6) [1] is the de facto mobility protocol in IPv6 wireless/mobile networks. Hierarchical Mobile IPv6 (HMIPv6) [2] was proposed by Internet Engineering Task Force (IETF) to mitigate the high signaling overhead that is incurred in Mobile IPv6 networks when mobile nodes (MNs) perform frequent handoffs. In HMIPv6 networks,

[☆] Preliminary version of this paper was presented at *ACM International Workshop on Mobility Management and Wireless Access (MobiWac) 2004*, Philadelphia, USA, September 2004.

^{*} Corresponding author. Tel.: +82 2 880 1832; fax: +82 2 872 2045.

E-mail addresses: shpack@mmlab.snu.ac.kr (S. Pack), tkk-won@snu.ac.kr (T. Kwon), yhchoi@snu.ac.kr (Y. Choi).

the mobility anchor point (MAP) has been introduced in order to handle binding update (BU) procedures due to handoffs within a MAP domain in a localized manner, which reduces the amount of network-wide signaling traffic for mobility.

In HMIPv6 networks, an MN configures two care-of-addresses (CoAs): a regional care-of-address (RCoA) and an on-link care-of-address (LCoA). The RCoA is an address of the MN on the MAP's subnet. The MN configures an RCoA when it receives a Router Advertisement (RA) message with the MAP option. On the other hand, the LCoA is an on-link CoA configured for the MN's interface based on the prefix information advertised by an access router (AR).

Fig. 1 illustrates the basic operations in HMIPv6 networks. An MN entering a MAP domain will receive RA messages containing information on one or more local MAPs. In case of multiple MAPs, the MN selects a MAP by its own criteria, which is the focus of this paper. Then, the MN sends a BU message to the selected MAP. The MN can bind its current location (i.e., LCoA) with an address on the MAP's subnet (i.e., RCoA). The MAP acts as a local home agent (HA) and, as such, it receives all packets on behalf of the MN it is serving and the MAP re-tunnels the received packets to the MN's current address. If the MN changes its current address within a local MAP domain, it only needs to register the new address with the MAP. The RCoA does not change as long as the MN moves within the same MAP domain. This makes the MN's mobility transparent to the correspondent nodes (CNs) and its HA.

In HMIPv6 networks, a MAP can exist at any level in the network hierarchy including at the leaf level of the AR, and several MAPs can be located within the same network hierarchy and function independently of each other. Especially when HMIPv6 is employed in a large-scale wireless/mobile network, multiple MAPs are used to provide scalable and robust mobile Internet services. In such environments, it is important for an MN to select the most suitable MAP among the available MAPs.

An MN needs to consider several factors for selecting an appropriate MAP in a foreign network. In the HMIPv6 specification [2], two MAP selection schemes were recommended. The first of these is a distance-based selection scheme, where an MN may choose the furthest MAP. This scheme is particularly efficient for fast MNs performing frequent handoffs, because by choosing the furthest MAP, the fast MNs can reduce the frequency of changing the serving MAP and informing the HA/CNs of this RCoA change. However, since each MN has different mobility characteristics, the furthest MAP may not constitute an appropriate solution for some MNs (e.g., slow MNs). Furthermore, if all MNs select the furthest MAP as their serving MAPs, this MAP would become a single point of performance bottleneck and it will result in a longer processing latency. The alternative scheme recommended in [2] is to announce the MAP's information (e.g., traffic load on the MAP), so that an MN can choose a MAP by considering MN's mobility characteristics and MAP's current state.

In addition to above schemes, mobility-based MAP selection schemes have been proposed in

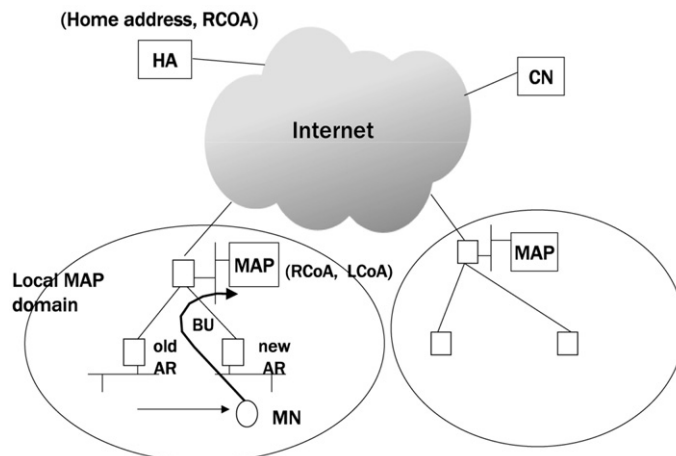


Fig. 1. Overview of HMIPv6 operation.

[5–8]. In these schemes, an MN selects its serving MAP depending on its mobility. For example, the fast MN selects a further MAP while the slow MN chooses a nearer MAP. Another proposal is an adaptive MAP selection scheme, where an MN selects the serving MAP by estimating session-to-mobility ratio (SMR),¹ was proposed in [9]. The SMR is analogous to call-to-mobility ratio (CMR) used in cellular networks. Namely, the SMR is defined by the ratio of the session arrival rate to the mobility rate (i.e., subnet crossing rate). The smaller the SMR of an MN is, the further MAP will be selected by the MN. Note that the SMR of an MN being small means that the MN's mobility rate is relatively higher than the session arrival rate.

In this paper, we conduct a comparative study of the four MAP selection schemes analytically: *the furthest MAP selection scheme, the nearest MAP selection scheme, the mobility-based MAP selection scheme, and the adaptive MAP selection scheme*. Through analytical modeling, we quantify their performances with focus on the binding update traffic and packet tunneling overhead. For the mobility-based MAP selection scheme and the adaptive MAP selection scheme, we explicitly derive the optimal threshold, and investigate the effect of the threshold on the performance. Also, how MNs are distributed among MAPs (i.e., load balancing) depending on schemes is analyzed.

The remainder of this paper is organized as follows. Section 2 details the four MAP selection schemes. We present the system model and the analytical model in Sections 3 and 4, respectively. In Section 5, the numerical results are presented and Section 6 concludes this paper.

2. Overview of MAP selection schemes

For the purpose of mobility support for telephony service in traditional cellular networks, a two-level hierarchy, consisting of the visitor location register (VLR) and home location register (HLR), has been widely used. These mobility agents

are statically deployed in the network and assigned to MNs based on their current locations. Recently, a hierarchical cellular network [3,4] was introduced to improve system capacity where MNs can be serviced by either an upper tier cell (e.g., macro cell) or a lower tier cell (e.g., micro cell). Which tier is selected is dependent on such factors as the MN's mobility rate, the current network load at each tier, etc. However, despite multiple tiers, there is only one available mobility agent (i.e., VLR) for each visiting MN and therefore the problem of mobility agent selection in the traditional cellular networks has not been investigated.

On the contrary, in data-oriented Mobile IP networks, depending on the size of a foreign network, more than one mobility agents (i.e., MAP) is likely to be deployed and dynamically assigned to individual MNs in order to provide more scalable and fault-tolerant mobile Internet services [12]. Consequently, how to select an appropriate mobility agent plays an important role for optimizing the binding update/packet delivery procedures and reducing the signaling overhead in the network. As mentioned earlier, several selection schemes have been reported in an attempt to resolve this mobility agent selection problem.

2.1. Distance-based selection scheme

In the HMIPv6 specification [2], a distance-based scheme was proposed which uses the distance (i.e., the number of hop counts) between the MAP and the MN for the MAP selection. In the distance-based scheme, an MN registers with the furthest MAP, in order to reduce the probability of sending binding update messages frequently. Without loss of generality, we assume that the area of the further MAP is larger than that of the nearer MAP. Accordingly, assuming the same moving speed, an MN at the further MAP domain is likely to request inter-MAP domain handoffs less frequently than an MN at the nearer MAP domain. In this way, the MN can avoid performing frequent binding updates to the HA/CNs. As a result, this selection scheme minimizes the amount of binding update traffic.

However, the furthest MAP selection scheme has some drawbacks. First, the furthest MAP from the MN may be close to (or even collocate with) the gateway of the foreign network. Consequently, if every MN selects the furthest MAP, the MAP may have to deal with most MNs in the foreign network and is likely to become a bottleneck point.

¹ A session is defined as a consecutive packet stream at the IP layer. To identify a session, a timer-based approach is employed [10]. Each MN maintains an active state timer with length T_A . If the time duration between the last received packet and the current received one is greater than T_A , the current packet is considered as the first packet of a new session. Otherwise, the packet is a subsequent packet of an ongoing session. This timer-based technique is similar to the session management scheme in Universal Mobile Telecommunication System (UMTS) [11].

Second, if the MN moves only within a limited area in the foreign network, it is unnecessary to register with the furthest MAP. In this case, using the furthest MAP will increase the handoff delay, because the distance from the MNs to the furthest MAP is relatively longer than that from the MNs to a nearer MAP [7].

2.2. Mobility-based selection scheme

In [5–8], MAP selection schemes based on the MN's mobility were proposed. In [5,6], a suitable MAP is selected based on the estimated velocity of the MN. The schemes in this category normally perform two functions: the determination of the MN's velocity and the selection of which MAP to register with. First, to determine the velocity of the MN, the MN's BU interval is measured. Each MAP records not only the binding information of the MN, but also time at which the BU is requested. When the MN moves into a new MAP domain, it is possible to calculate the MN's residence time in the previous MAP domain by comparing old BU time in the previous MAP domain with new BU time. Then, the velocity is calculated by dividing the distance that the MN has traversed by residence time. However, measuring the distance of the MN exactly is not a simple issue. Instead, a concept of "standard distance" has been introduced where a priori distance value is used for each MAP domain. When the velocity of an MN is estimated, the previously estimated velocity is also taken into consideration, in order to reduce the error in estimating the MN's mobility. After that, the MN selects a MAP by comparing its BU interval with the average BU interval of the MNs serviced by each MAP. The average BU interval is assumed to be contained in the RA message.

In general, it is difficult to estimate the velocity of MNs and the estimation results are often inaccurate. Therefore, MNs may not always register with an appropriate MAP. In [7,8], a new MAP selection scheme that uses the MAP topology instead of the MN's velocity was proposed. In this scheme, an MN first listens to the RA messages from MAPs and figures out the MAP topology. From the topology information, the MN chooses the nearest cross-over MAP. More detailed procedures are as follows. To inform MNs of the MAP topology, the furthest MAP starts with broadcasting a RA message and then the RA message is propagated from the furthest MAP towards MNs. While the RA message

is en route to MNs, intermediate MAPs append their own MAP information. These RA messages can also be cached and broadcasted by ARs. By receiving these RA messages, each MN figures out the topology of available MAPs. If there are common MAPs that appear in every RA message, the MN chooses the nearest common MAP. If there is only one kind of RA message, the MN selects the furthest MAP by default. When the MN moves beyond the coverage of the selected MAP, it will immediately detect this by noticing that the serving MAP is absent in the current RA message.

2.3. Adaptive MAP selection scheme

In general, the total cost in Mobile IP systems can be modeled as the sum of the packet delivery cost and the binding update cost [13,14]. In HMIPv6 networks, since the MAP has to deal with all packet transmissions from and to MNs serviced by the MAP, the session characteristics (e.g., the amount of traffic and session arrival frequency) are also important factors to determine a suitable MAP. However, the mobility-based selection scheme overlooks the effect of session activity, so that it may not be able to select the most appropriate MAP, especially when the session activity is dominant. To address this problem of the mobility-based schemes, an adaptive MAP selection scheme was proposed in [9]. In the adaptive MAP selection scheme, an MN selects the serving MAP depending on its SMR. The SMR is a key factor representing the ratio of the session arrival rate to the mobility rate. The adaptive MAP selection scheme consists of the following four steps.

- **Initialization:** An MN collects all RA messages sent from the available MAPs in a foreign network. From these RA messages, the MN obtains information on each MAP, e.g., the hop distance, network load. Using the MAP information, the MN constructs an available MAP list (AML).
- **SMR estimation:** For each measurement interval, the MN estimates its SMR by measuring the number of subnet crossings and session arrivals. At the same time, the MN updates its SMR and compares the estimated SMR with two SMR threshold values: lower and upper SMR thresholds.
- **Threshold determination:** To select the optimal MAP adaptively, two SMR threshold values are calculated by an iterative method [9].

- **MAP selection:** If the updated SMR is smaller than the lower SMR threshold value or larger than the upper SMR threshold value, the MN computes the total costs for the other MAPs available in the current location. Then, the MN selects the MAP with the minimum total cost.

3. System model and assumptions

In this section, we describe the system model and assumptions for the analytical model. As shown in Fig. 2, we use a two-level MAP topology consisting of two types of MAPs: *higher MAP* (HMAP) and *lower MAP* (LMAP) [5,6,9]. In this topology, the LMAP and HMAP domains cover N_L and N_H AR subnets, respectively. Typically, since N_H is larger than N_L , the average residence time in the HMAP domain is longer than that in the LMAP domain. Similar to [8], the AR broadcasts the information on the MAP topology. Hence, the MN learns which MAPs are available at the current location. After that, the MN selects its serving MAP depending on the employed MAP selection scheme. Hereafter, F , N , M , and A represent the furthest, nearest, mobility-based, and adaptive MAP selection schemes, respectively. M is a generalized version of different relevant mobility-based MAP selection proposals. In each scheme, a new MAP selection procedure is triggered when the MN hands off to a new AR subnet. If the newly selected MAP is the same as the previous one, no binding updates to the MAP and HA are performed; otherwise, the MN informs both the MAP and HA of its new

LCoA. The MAP selection procedure of each scheme can be summarized as follows:

- F : The MN always selects the HMAP.
- N : The MN always selects the LMAP.
- M : If the estimated residence time is less than a pre-defined threshold T_{th} , the MN selects the HMAP. Otherwise, the MN selects the LMAP.
- A : If the estimated SMR is less than a pre-defined threshold S_{th} , the MN selects the HMAP. Otherwise, the MN selects the LMAP.

In addition, the followings are assumed without loss of generality.

1. The subnet residence time follows a general distribution with mean $1/\mu_S$; its probability density function (PDF) is $f_S(t)$. $F_S(t)$ and $f_S^*(s)$ denote the cumulative distribution function (CDF) and the Laplace transform of $f_S(t)$, respectively.
2. The HMAP (LMAP) domain residence time follows a general distribution with mean $1/\mu_D^H$ ($1/\mu_D^L$) and its PDF is denoted by $f_{HD}(t)$ ($f_{LD}(t)$). $F_{HD}(t)$ ($F_{LD}(t)$) and $f_{HD}^*(S)$ ($f_{LD}^*(S)$) denote the CDF and Laplace transform of $f_{HD}(t)$ ($f_{LD}(t)$), respectively.
3. The session arrival process follows a Poisson distribution with rate λ_I .
4. Let t_I and t_S be the inter-session arrival time and the subnet residence time, respectively. SMR is the session-to-mobility ratio, which is defined as t_S/t_I , and the CDF of SMR is given by the below equation (see Appendix A). In addition, the PDF of SMR can be obtained from $g(\delta) = \frac{d}{d\delta} G(\delta)$.

$$G(\delta) = \Pr(\text{SMR} < \delta) = \Pr(t_S/t_I < \delta) \\ = f_S^*(S)|_{S=\lambda_I/\delta}.$$

4. Analytical model

For the purpose of performance comparison, we formulate the binding update (BU) cost and the packet deliver (PD) cost in HMIPv6 networks [13–15]. The BU and PD costs are the accumulative traffic loads due to exchanging BU/BACK messages and IP tunneling headers of data packets, respectively [16]. Namely, the BU and PD costs represent the amount of signaling traffic and network overhead, which should be minimized. We model the BU and PD costs during an inter-session arrival time, which is denoted as the time interval between

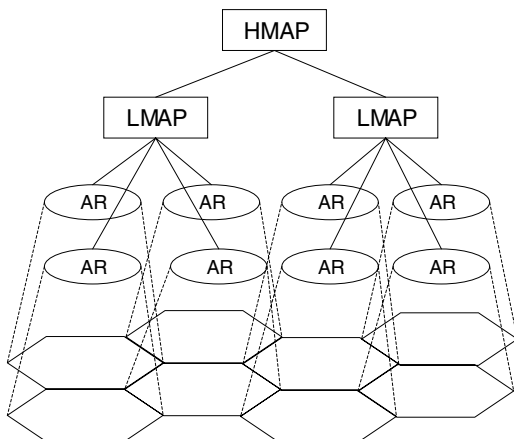


Fig. 2. Two-level MAP topology.

the arrival of first packet of a data session and the arrival of the first packet of the next data session [10].

4.1. Binding update cost

Let N_D and N_S be the numbers of MAP domain crossings and subnet crossings, respectively. Then, the BU cost per session can be expressed as

$$C_{BU} = E(N_D) \cdot B_{HA} + E(N_S) \cdot B_M + B_{HA} + B_M, \quad (1)$$

where B_{HA} and B_M are the unit BU costs to the HA and MAP, respectively. $E(N_D)$ and $E(N_S)$ are the mean numbers of MAP domain crossings and subnet crossings per session, respectively. In Eq. (1), the first and second terms on the right-hand side refer to the BU costs incurred by the MN's movements, whereas the third and fourth terms on the right-hand side are the initial BU costs. B_{HA} and B_M are calculated by

$$B_{HA} = d_{MN-HA} \cdot (BU + BACK) \quad (2)$$

and

$$B_M = d_{MN-MAP} \cdot (BU + BACK), \quad (3)$$

where d_{X-Y} is the hop distance between X and Y . BU and $BACK$ are the sizes of BU and BACK messages, respectively.

As mentioned above, we assume a two-level MAP hierarchy. Let N_D^H and N_D^L be the numbers of HMAP and LMAP domain crossings per session, respectively. Then, the BU costs in F and N are respectively given by

$$C_{BU}^F = E(N_D^H) \cdot B_{HA} + E(N_S) \cdot B_M^H + B_{HA} + B_M^H \quad (4)$$

and

$$C_{BU}^N = E(N_D^L) \cdot B_{HA} + E(N_S) \cdot B_M^L + B_{HA} + B_M^L, \quad (5)$$

where B_M^H and B_M^L represent the unit BU costs to the HMAP and LMAP, respectively, and they are obtained from Eq. (3) by replacing d_{MN-MAP} with $d_{MN-HMAP}$ or $d_{MN-LMAP}$.

Unlike the distance-based selection schemes, M uses a threshold value denoted by T_{th} . If the residence time of the MN is larger than T_{th} , the MN chooses the LMAP as its serving MAP. This is because it is better to reduce the local BU cost (to the MAP) rather than the home BU cost (to the HA) if the MN has low mobility. On the other hand, A uses the SMR not the residence time (or mobility rate). Namely, A chooses the serving MAP depending on the SMR threshold value S_{th} .

Let π_H and π_L be the steady state probabilities that the HMAP and LMAP are chosen by the MN, respectively. Then, π_H and π_L in M can be computed as

$$\begin{aligned} \pi_H &= \Pr(t \leq T_{th}) = F_S(T_{th}) \quad \text{and} \\ \pi_L &= \Pr(t > T_{th}) = 1 - F_S(T_{th}). \end{aligned} \quad (6)$$

Then, the BU cost of M is given by

$$\begin{aligned} C_{BU}^M &= \pi_L \cdot (E(N_S) \cdot B_M^L + E(N_D^L) \cdot B_{HA} + B_{HA} + B_M^L) \\ &\quad + \pi_H \cdot (E(N_S) \cdot B_M^H + E(N_D^H) \cdot B_{HA} + B_{HA} + B_M^H). \end{aligned} \quad (7)$$

Similarly, π_H and π_L in A are obtained from

$$\begin{aligned} \pi_H &= \Pr(\text{SMR} \leq S_{th}) = G(S_{th}) \quad \text{and} \\ \pi_L &= \Pr(\text{SMR} > S_{th}) = 1 - G(S_{th}). \end{aligned} \quad (8)$$

Accordingly, the BU cost of A can be calculated from

$$\begin{aligned} C_{BU}^A &= \pi_L \cdot (E(N_S) \cdot B_M^L + E(N_D^L) \cdot B_{HA} + B_{HA} + B_M^L) \\ &\quad + \pi_H \cdot (E(N_S) \cdot B_M^H + E(N_D^H) \cdot B_{HA} + B_{HA} + B_M^H). \end{aligned} \quad (9)$$

4.2. Packet delivery cost

The original packet is not counted as overhead with respect to packet delivery, and therefore the PD cost takes into account the tunneling overhead. Fig. 3 shows two packet deliver paths in HMIPv6 networks. As shown in Fig. 3, a packet is delivered to the destination MN either through the HA or not. According to the HMIPv6 specification [2], packets of a session are first routed to the HA to check the current location of the destination MN (i.e., *indirect path*). Once the MN receives the packets tunneled via the HA, the MN sends a binding update message to the CN, i.e., route optimization. Then, the CN updates its binding cache and sends subsequent packets to the MAP directly (i.e., *direct path*).

By considering these two packet delivery paths, the PD cost per session can be expressed as

$$C_{PD} = \omega \cdot L_S \cdot P_I + (1 - \omega) \cdot L_S \cdot P_D, \quad (10)$$

where ω is the ratio of packets going through the HA before the completion of the binding update procedure. L_S is the average session length in numbers of packets. P_I and P_D are unit costs incurred to deliver a packet in the indirect and direct paths,

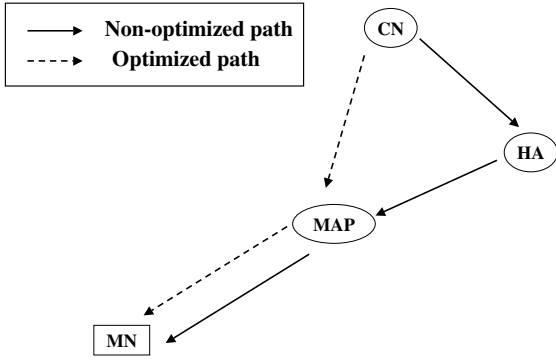


Fig. 3. Packet delivery paths in HMIPv6 network.

respectively. The unit costs for the HMAP (P_I^H and P_D^H) and LMAP (P_I^L and P_D^L) are given by

$$P_I^H = t \cdot d_{HA-HMAP} + 2t \cdot d_{HMAP-MN},$$

$$P_D^H = t \cdot d_{HMAP-MN},$$

$$P_I^L = t \cdot d_{HA-LMAP} + 2t \cdot d_{LMAP-MN},$$

$$P_D^L = t \cdot d_{LMAP-MN},$$

where t is the unit tunneling cost, i.e., the tunneling header size in bytes. Note that two tunneling headers are employed in the path between the MAP and MN. Unlike [14], a weight factor for wireless links is not used because the weight does not affect the performance of MAP selection schemes.

Then, the PD cost in each MAP selection scheme can be summarized as follows:

1. F :

$$C_{PD}^F = \omega \cdot L_S \cdot P_I^H + (1 - \omega) \cdot L_S \cdot P_D^H. \quad (11)$$

2. N :

$$C_{PD}^N = \omega \cdot L_S \cdot P_I^L + (1 - \omega) \cdot L_S \cdot P_D^L. \quad (12)$$

3. M and A :

$$C_{PD}^M \text{ (or } C_{PD}^A) = \pi_H \cdot (\omega \cdot L_S \cdot P_I^H + (1 - \omega) \cdot L_S \cdot P_D^H) + \pi_L \cdot (\omega \cdot L_S \cdot P_I^L + (1 - \omega) \cdot L_S \cdot P_D^L). \quad (13)$$

4.3. Determination of threshold value

In M , the threshold value for the average subnet residence time should be carefully determined. Although each proposal uses different mobility estimation algorithms [5–8], we consider a mobility-

based selection scheme with an optimal threshold. In other words, the optimal threshold for the subnet residence time is derived by the analytical method and it is used for the performance analysis. Similarly, the optimal SMR threshold in A is also obtained by the analytical method and the adaptive selection scheme with the optimal SMR threshold is evaluated.

For tactical analysis, the subnet and domain residence times are assumed to follow exponential distributions.² Then, $E(N_D^H) = \mu_D^H / \lambda_I$, $E(N_D^L) = \mu_D^L / \lambda_I$, and $E(N_S) = \mu_S / \lambda_I$ [21]. In addition, by the fluid flow mobility model [10], μ_D^H and μ_D^L can be approximated to $\mu_S / \sqrt{N_H}$ and $\mu_S / \sqrt{N_L}$, respectively.

The optimal threshold value of M is a point where the BU cost of LMAP is equal to that of HMAP. Therefore, the optimal threshold value can be obtained from Theorem 1.

Theorem 1. *If $\frac{B_{HA} / \sqrt{N_H} + B_M^H}{B_{HA} / \sqrt{N_L} + B_M^L}$ is smaller than 1, there exists an optimal threshold T_{th}^* in $(0, \infty)$ and it is given by³*

$$T_{th}^* = \frac{\left(\frac{1}{\sqrt{N_H}} - \frac{1}{\sqrt{N_L}}\right) \cdot B_{HA} + B_M^H - B_M^L}{\lambda_I (B_M^L - B_M^H)}.$$

Proof. See Appendix B. \square

On the other hand, the optimal SMR threshold value in A is a point where the MAP with a less total cost is changed. Theorem 2 shows the existence of optimal SMR threshold value.

Theorem 2. *If $\frac{B_{HA} / \sqrt{N_H} + B_M^H}{B_{HA} / \sqrt{N_L} + B_M^L}$ is smaller than 1, there exists an optimal threshold (S_{th}^*) in $(0, \infty)$ and it is given by*

$$S_{th}^* = \frac{B_M^H - B_M^L - B_{HA} \cdot \left(\frac{1}{\sqrt{N_L}} - \frac{1}{\sqrt{N_H}}\right)}{C_{PD}^L - C_{PD}^H + B_M^L - B_M^H}.$$

Proof. See Appendix C. \square

² Although the residence time is typically non-exponential in wireless/mobile networks, the analysis based on the simplified exponential assumption has been widely used [18–20] and provides useful mean value information.

³ This condition indicates that an optimal T_{th} may not exist in $(0, \infty)$ in some environments. However, it does not mean that M does not work in such environments. In this cases, M performs the same MAP selection procedures as F or N by choosing a sufficiently large or small threshold value. This fact is also applied to A (see Theorem 2).

5. Numerical results

Table 1 shows the default parameter values used in the numerical analysis. The sizes of tunneling header and BU/BACK messages follow the specification of MIPv6 [1] and HMIPv6 [2]. From [22,23], the average session length is assumed to be 10 Kbytes and the packet size is fixed at 1 Kbytes. Hence, L_S is set to 10. It is assumed that route optimization is performed when the first packet of a session is received, and thus ω is given by $1/10 = 0.1$. Default N_H and N_L are set to 64 and 9, respectively; however, the effect of difference MAP domain sizes is investigated in Section 5.1. In terms of hop distances among the MN, HMAP, LMAP, and HA, suitable values satisfying the conditions in Theorems 1 and 2 are chosen. Note that similar results can be observed even for different distance values.

5.1. Optimal threshold

Fig. 4 shows the variation of the optimal threshold as the session arrival rate λ_1 is varied. As shown in Fig. 4, T_{th}^* of M is dependent on λ_1 , while S_{th}^* is constant regardless of λ_1 . This result demonstrates that A is a more feasible solution. In other words,

when the session arrival rate is changed, M should determine a new threshold to achieve the optimal performance. It indicates that M results in higher MAP selection overhead and instability. On the contrary, A can use a fixed threshold as long as the network topology parameters (i.e., N_L and N_H) are fixed.

The effect of the HMAP domain size is also shown in Fig. 4. In this result, N_L is fixed to 9, whereas N_H is varied over 25, 64, and 81. For M , as N_H increases, the optimal threshold decreases. In terms of reducing the BU cost, the HMAP can reduce the home BU cost to the HA, whereas the LMAP is efficient to reduce the local BU cost to the MAP. Consequently, the MN in M selects the HMAP when reducing the home BU cost is more effective than reducing the local BU cost. On the other hand, the LMAP is chosen when the local BU cost is more dominant. Apparently, the frequency of home BUs is reduced as N_H increase. Therefore, as N_H becomes large, it is more effective to reduce the cost due to local BUs, which are more frequently performed than home BUs. To this end, T_{th}^* should become smaller as N_H increases. This is because the LMAP is selected with a higher probability when T_{th} is small.

Table 1
Default parameter values used in numerical analysis

d_{MN-HA}	$d_{MN-HMAP}$	$d_{MN-LMAP}$	L_S	ω	N_H	N_L	t	BU	BACK
8	4	3	10	0.1	64	8	40 bytes	72 bytes	52 bytes

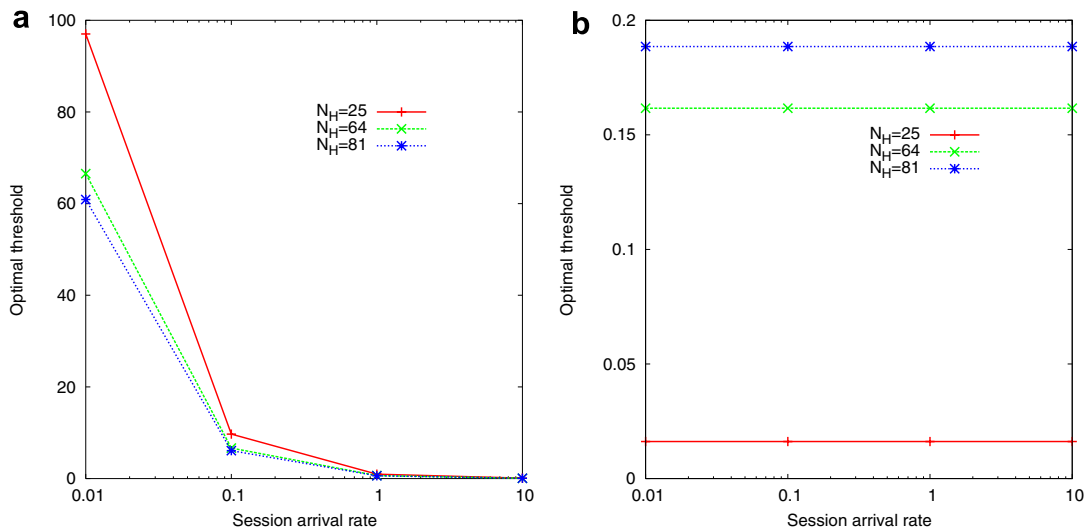


Fig. 4. Optimal threshold vs. session arrival rate. (a) Optimal T_{th} and (b) optimal S_{th} .

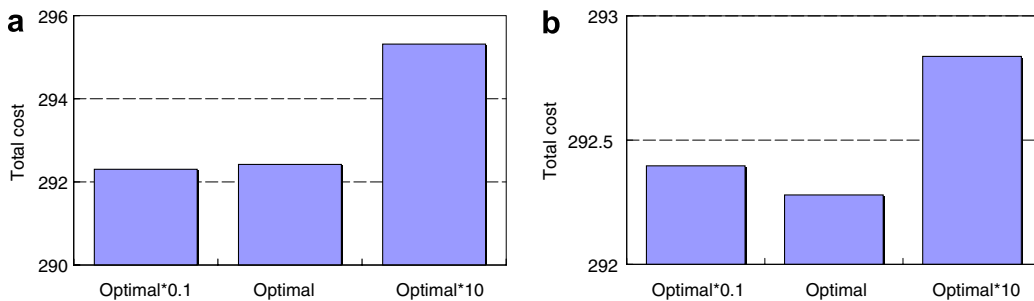


Fig. 5. Effect of different threshold values. (a) Effect of T_{th} and (b) effect of S_{th} .

On the other hand, S_{th}^* increases with the increase of N_H . In A , the MN selects the serving MAP depending on which is the most dominant cost between the BU and PD costs. In other words, if the BU cost is more important than the PD cost in reducing the total cost, the HMAP is chosen; otherwise, the LMAP is selected as the serving MAP. As mentioned before, a larger N_H represents that the HMAP covers more subnets, and hence the BU cost can be reduced more significantly than the PD cost by choosing the HMAP. Therefore, the probability that the HMAP is chosen should be increased as N_H increases. Accordingly, S_{th}^* increases proportionally to N_H .

Fig. 5 indicates the sensitivity of M and A when different threshold values are used. To this end, three different values are used: $T_{th}^* \times 10$ ($S_{th}^* \times 10$), T_{th}^* (S_{th}^*), and $T_{th}^* \times 10$ ($S_{th}^* \times 10$). In the case of M , a low total cost can be observed when a small threshold (i.e., $T_{th}^* \times 10$) is employed. This is because T_{th}^* is an optimal threshold to minimize the BU cost not the total cost. Consequently, the MAP selection in M may not be optimal when we consider both the PD cost and the BU cost. On the other hand, A shows the lowest total cost when S_{th}^* is used. Also, even if different threshold values are used, the variation in the total cost is low for A . This implies that A is less sensitive to threshold values and hence it can provide higher robustness in HMIPv6 networks.

5.2. Effect of mobility rate

Tables 2–4 demonstrate the effect of the mobility rate. It is assumed that the session arrival rate λ_I is normalized to 1.0, and therefore the unit of the mobility rate μ_S is λ_I . As shown in Table 2, F shows the lowest BU cost when the mobility rate is high. On the other hand, N exhibits the lowest BU cost when

Table 2

Binding update cost: effect of μ_S

μ_S (Unit: λ_I)	M	A	F	N
0.01	1.383	1.382	1.506	1.382
0.1	1.453	1.448	1.563	1.446
1	2.104	2.089	2.125	2.083
10	7.751	8.021	7.750	8.458
100	64.000	64.478	64.000	72.208

Table 3

Packet delivery cost: effect of μ_S

μ_S (Unit: λ_I)	M	A	F	N
0.01	1.487	1.485	1.875	1.484
0.1	1.510	1.491	1.875	1.484
1	1.674	1.539	1.875	1.484
10	1.874	1.726	1.875	1.484
100	1.875	1.852	1.875	1.484

Table 4

Total cost: effect of μ_S

μ_S (Unit: λ_I)	M	A	F	N
0.01	2.870	2.867	3.381	2.866
0.1	2.963	2.938	3.438	2.930
1	3.778	3.628	4.000	3.568
10	9.625	9.746	9.625	9.943
100	65.875	6.331	65.875	73.693

the mobility rate is low, which indicates that it is better to select the LMAP with a lower PD cost if the MN's mobility is low. Regarding M and A , they exhibit a slightly higher BU cost than F when μ_S is higher than 1.0. On the other hand, when μ_S is lower than 1.0, they show comparable total costs to N . In short, M and A choose a MAP just like either F or N dynamically depending on the MN's characteristics (e.g., mobility and session activity) and thus they achieve more adaptive performances in diverse mobile environments.

Unlike the BU cost, since the PD costs of F and N are not dependent on the mobility, F and N have constant PD costs. At the same time, since the distance between the MN and the HMAP is longer than the one between the MN and the LMAP, the HMAP incurs a higher tunneling cost and hence a higher PD cost is observed. Consequently, the PD cost of F is larger than that of N . As shown in Table 3, A shows a less PD cost than M . This is because M does not take the PD cost into account during the MAP selection procedure. However, both A and M show more adaptive PD costs to the mobility rate than F and N .

Table 4 summarizes the total cost as a function of μ_S . Similar to the BU and PD costs, when μ_S is low, N is the best scheme in reducing the total cost, while F is preferable if μ_S is high. On the other hand, the total costs of M and A are close to that of N for low mobility, whereas their total costs approach that of F for high mobility. This observation indicates that M and A are adaptive to varying mobility. Comparing A and M , A is better with the low μ_S , but M shows a less total cost if μ_S is high. In other words, M is a more appropriate solution in a highly mobile environment.

5.3. Load balancing

When we design a MAP selection scheme, it is important to distribute total MAP loads as evenly as possible, as well as to select a MAP with the lowest total cost. To investigate load distribution, we first classify MNs into two classes: *fast MN* and *slow MN*. The mobility rates of fast and slow MNs are 10 and 0.1, respectively. Then, the average HMAP load L can be defined as

$$L = \alpha \times T \times \pi_H^{\text{fast}} + (1 - \alpha) \times T \times \pi_H^{\text{slow}}, \quad (14)$$

where α is the fraction of fast MNs to the total MNs and T is the total number of MNs. π_H^{fast} and π_H^{slow} are the probabilities that the fast and slow MNs select the HMAP as their serving MAPs, respectively. Intuitively, $T - L$ refers to the LMAP load.

Fig. 6 plots the HMAP load as a function of α . As α approaches 1.0, there exist more fast MNs and hence the HMAP load increases. In terms of load balancing, when L is equal to $T/2$, the optimal load balancing performance can be achieved. As shown in Fig. 6, A provides a more balanced HMAP load if α is larger than 0.61; otherwise, M shows a better load balancing performance. Specifically, when α is 1.0, L in M and A are 99.87 and

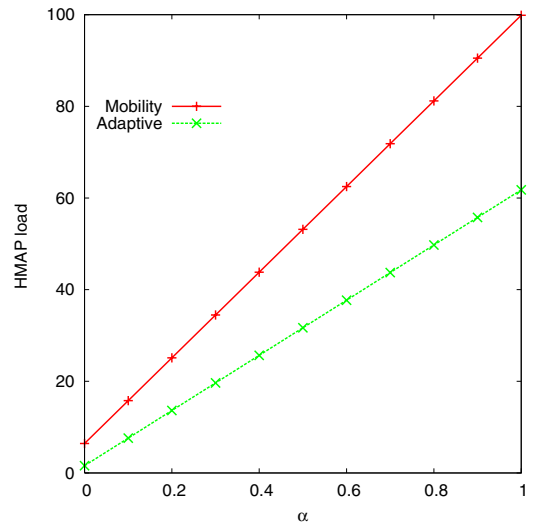


Fig. 6. HMAP load: $T = 100$.

61.77, respectively. In other words, M shows a highly skewed HMAP load in a mobile network where most MNs are fast. This is because M considers only mobility for the MAP selection and most fast MNs select the HMAP regardless of their session activities. On the contrary, since A uses a combined factor (i.e., SMR), some fast MNs choose the LMAP instead of the HMAP and therefore a better load balancing can be achieved.

6. Conclusion

In this paper, we have conducted a comparative study for four MAP selection schemes: the furthest MAP selection scheme, the nearest MAP selection scheme, the mobility-based MAP selection scheme, and the adaptive MAP selection scheme. Based on the analytical models, we compared the signaling overhead incurred by binding update procedures and IP tunneling. Also, how the MAP load is balanced among MAPs was investigated. Overall, the dynamic schemes (i.e., the mobility-based and the adaptive MAP selection schemes) exhibit more desirable performances than the static schemes (i.e., the furthest and the nearest MAP selection schemes), since they dynamically select the serving MAP depending on the MN's characteristics (e.g., mobility rate and session activity). Especially, the adaptive MAP selection scheme is not sensitive to the change of threshold values and hence does not require overhead for the threshold determination. Therefore, the adaptive MAP selection scheme

achieves low implementation overhead compared with the mobility-based MAP selection scheme.

Acknowledgements

This work was supported in part by the Brain Korea 21 project of the Ministry of Education and in part by the National Research Laboratory project of the Ministry of Science and Technology, 2005, Korea.

Appendix A. SMR distribution

The SMR is defined as t_S/t_I . It is assumed that t_I follows an exponential distribution with rate λ_I whereas t_S follows a general distribution. Then, the SMR distribution function can be derived from

$$\begin{aligned} \Pr(\text{SMR} < \delta) &= \Pr(t_S/t_I < \delta) = \Pr(t_S < \delta \cdot t_I) \\ &= \int_0^\infty \Pr(t_S < \delta\tau) \cdot \lambda_I e^{-\lambda_I\tau} \cdot d\tau. \end{aligned} \quad (\text{A.1})$$

If we set t as $\delta\tau$, Eq. (A.1) becomes

$$\frac{\lambda_I}{\delta} \cdot \int_0^\delta \Pr(t_S < t) \cdot e^{-\frac{\lambda_I t}{\delta}} \cdot dt. \quad (\text{A.2})$$

By the property of the Laplace transform [17], $\int_0^\infty \Pr(t_S < t) \cdot e^{-st} \cdot dt = \frac{f_S^*(s)}{s}$, where $f_S^*(s)$ is the Laplace transform of $f_S(t)$, and Eq. (A.2) becomes

$$\frac{\lambda_I}{\delta} \cdot \frac{f_S^*(S)}{S} \Big|_{S=\lambda_I/\delta} = f_S^*(S) \Big|_{S=\lambda_I/\delta}. \quad (\text{A.3})$$

Appendix B. Proof of Theorem 1

First, the BU costs of F and N (shown in Eqs. (4) and (5), respectively) can be expressed as functions of mobility rate, μ_S , as

$$C_{\text{BU}}^F(\mu_S) = \frac{\mu_S}{\lambda_I} \left(\frac{1}{\sqrt{N_H}} \cdot B_{\text{HA}} + B_M^H \right) + B_{\text{HA}} + B_M^H$$

and

$$C_{\text{BU}}^N(\mu_S) = \frac{\mu_S}{\lambda_I} \left(\frac{1}{\sqrt{N_L}} \cdot B_{\text{HA}} + B_M^H \right) + B_{\text{HA}} + B_M^L.$$

Since B_M^H is larger than B_M^L , the following condition is satisfied when μ_S approaches 0:

$$\lim_{\mu_S \rightarrow 0} \frac{C_{\text{BU}}^F(\mu_S)}{C_{\text{BU}}^N(\mu_S)} = \frac{B_{\text{HA}} + B_M^H}{B_{\text{HA}} + B_M^L} > 1. \quad (\text{B.1})$$

On the other hand, if μ_S becomes infinity, Eq. (B.2) is met by the assumed condition, $\frac{B_{\text{HA}}/\sqrt{N_H} + B_M^H}{B_{\text{HA}}/\sqrt{N_L} + B_M^L} < 1$

$$\lim_{\mu_S \rightarrow \infty} \frac{C_{\text{BU}}^F(\mu_S)}{C_{\text{BU}}^N(\mu_S)} = \frac{B_{\text{HA}}/\sqrt{N_H} + B_M^H}{B_{\text{HA}}/\sqrt{N_L} + B_M^L} < 1. \quad (\text{B.2})$$

By Eqs. (B.1) and (B.2), there exists an intersection point where C_{BU}^F and C_{BU}^N have the same value in $(0, \infty)$ and the point represents the optimal threshold. The optimal threshold can be simply derived from a difference function that is denoted as $\Delta C_{\text{BU}}(\mu_S) = C_{\text{BU}}^F(\mu_S) - C_{\text{BU}}^N(\mu_S)$. That is, the optimal threshold value in M satisfies the condition,

$$\Delta C_{\text{BU}}(\mu_S^*) = 0.$$

This condition can be reduced as

$$\begin{aligned} \Delta C_{\text{BU}}(\mu_S) &= \frac{\mu_S}{\lambda_I} \left(\frac{1}{\sqrt{N_H}} \cdot B_{\text{HA}} + B_M^H \right) + B_{\text{HA}} \\ &\quad + B_M^H - \left(\frac{\mu_S}{\lambda_I} \left(\frac{1}{\sqrt{N_L}} \cdot B_{\text{HA}} + B_M^H \right) \right. \\ &\quad \left. + B_{\text{HA}} + B_M^L \right) = 0. \end{aligned}$$

Note that the average residence time is the inverse of the average mobility rate. Consequently, the optimal threshold is given

$$\frac{1}{\mu_S^*} = \frac{\left(\frac{1}{\sqrt{N_H}} - \frac{1}{\sqrt{N_L}} \right) \cdot B_{\text{HA}} + B_M^H - B_M^L}{\lambda_I (B_M^L - B_M^H)} = T_{\text{th}}^*. \quad \square \quad (\text{B.3})$$

Appendix C. Proof of Theorem 2

The total costs of F and N can be expressed as functions of SMR (i.e., $\delta = \lambda_I/\mu_S$) as

$$C_T^F(\delta) = \frac{1}{\delta} \left(\frac{1}{\sqrt{N_H}} \cdot B_{\text{HA}} + B_M^H \right) + B_{\text{HA}} + B_M^H + C_{\text{PD}}^F$$

and

$$C_T^N(\delta) = \frac{1}{\delta} \left(\frac{1}{\sqrt{N_L}} \cdot B_{\text{HA}} + B_M^H \right) + B_{\text{HA}} + B_M^L + C_{\text{PD}}^N.$$

If δ becomes infinity, Eq. (C.1) can be satisfied because B_M^H and C_{PD}^F are larger than B_M^L and C_{PD}^N , respectively

$$\lim_{\delta \rightarrow \infty} \frac{C_T^F(\mu_S)}{C_T^N(\mu_S)} = \frac{B_{HA} + B_M^H + C_{PD}^F}{B_{HA} + B_M^L + C_{PD}^N} > 1. \quad (C.1)$$

On the other hand, if δ approaches 0, Eq. (C.2) is met by the assumption

$$\lim_{\delta \rightarrow 0} \frac{C_T^F(\mu_S)}{C_T^N(\mu_S)} = \frac{B_{HA}/\sqrt{N_H} + B_M^H}{B_{HA}/\sqrt{N_L} + B_M^L} < 1. \quad (C.2)$$

Therefore, by Eqs. (C.1) and (C.2), there exists an intersection point where C_T^E and C_T^N have the same value in $(0, \infty)$ and the point represents the optimal threshold. The optimal threshold can be derived from a difference function denned as $\Delta C_T(\delta) = C_T^F(\delta) - C_T^N(\delta)$. That is, the optimal threshold value S_{th}^* satisfies the following condition:

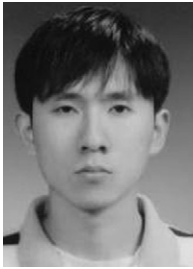
$$\begin{aligned} \Delta C_T(\delta) &= \frac{1}{\delta} \left(\frac{1}{\sqrt{N_H}} \cdot B_{HA} + B_M^H \right) + B_{HA} \\ &\quad + B_M^H + C_{PD}^F - \left(\frac{1}{\delta} \left(\frac{1}{\sqrt{N_L}} \cdot B_{HA} + B_M^H \right) \right. \\ &\quad \left. + B_{HA} + B_M^L + C_{PD}^N \right) = 0. \end{aligned}$$

After some manipulations, we get

$$S_{th}^* = \frac{B_M^H - B_M^L - B_{HA} \cdot \left(\frac{1}{\sqrt{N_L}} - \frac{1}{\sqrt{N_H}} \right)}{C_{PD}^L - C_{PD}^H + B_M^L - B_M^H}. \quad \square \quad (C.3)$$

References

- [1] D. Johnson, C. Perkins, J. Arkko, Mobility Support in IPv6, IETF RFC 3775, June 2003.
- [2] H. Soliman, C. Castelluccia, K.E. Malki, L. Bellier, Hierarchical Mobile IPv6 mobility management (HMIPv6), IETF RFC 4140, August 2005.
- [3] Y. Lin, I. Chlamtac, Heterogeneous personal communications services: integration of PCS systems, IEEE Communications Magazine 34 (9) (1996) 138–145.
- [4] X. Xu, B. Mukherjee, B. Bhargava, Low-cost, low-delay location update/paging/scheme in hierarchical cellular networks, in: Proceedings of the ACM MobiDE 2003, September 2003.
- [5] K. Kawano, K. Kinoshita, K. Murakami, Multilevel hierarchical distributed IP mobility management scheme for wide area networks, in: Proceedings of the IEEE ICCCN 2002, October 2002.
- [6] K. Kawano, K. Kinoshita, K. Murakami, A mobility-based terminal management in IPv6 networks, IEICE Transactions on Communications E85-B (10) (2002) 2099.
- [7] Y. Xu, H.C.J. Lee, V.L.L. Thing, A local mobility agent selection algorithm for mobile networks, in: Proceedings of the IEEE ICC 2003, May 2003.
- [8] V. Thing, H. Lee, Y. Xu, Designs and analysis of local mobility agents discovery, selection and failure detection for Mobile IPv6, in: Proceedings of the IEEE MWCN 2002, September 2002.
- [9] S. Pack, M. Nam, T. Kwon, Y. Choi, An adaptive mobility anchor point selection scheme in hierarchical Mobile IPv6 networks, Computer Communications 29 (16) (2006) 3066–3078.
- [10] X. Zhang, J. Castellanos, A. Campbell, P-MIP: paging extensions for mobile IP, ACM/Kluwer Mobile Networks and Applications (MONET) 7 (2) (2002) 127–141.
- [11] Y. Lin, A. Pang, H. Rao, Impact of mobility on mobile telecommunications networks, Wiley Wireless Communications and Mobile Computing 5 (8) (2005) 713–732.
- [12] S. Pack, T. You, Y. Choi, Performance analysis of robust Hierarchical Mobile IPv6 for fault-tolerant mobile services, IEICE Transactions on Communications E87-B (5) (2004) 1158–1165.
- [13] S. Pack, Y. Choi, A study on performance of Hierarchical Mobile IPv6 in IP-based cellular networks, IEICE Transactions on Communications E87-B (3) (2004) 462–469.
- [14] J. Xie, I. Akyildiz, A distributed dynamic regional location management scheme for mobile IP, IEEE Transactions on Mobile Computing 1 (3) (2002) 163–175.
- [15] Y. Fang, Movement-based location management and trade-off analysis for wireless mobile networks, IEEE Transactions on Computers 52 (6) (2003) 791–803.
- [16] S. Lo, G. Lee, W. Chen, J. Liu, Architecture for mobility and QoS support in all-IP wireless networks, IEEE Journal on Selected Areas on Communications 22 (4) (2004) 691–705.
- [17] L. Kleinrock, Queuing Systems, vol. 1: Theory, John Wiley & Sons, 1975.
- [18] L. Ortigoza-Guerrero, A.H. Aghvami, A prioritized handoff dynamic channel allocation strategy for PCS, IEEE Transactions on Vehicular Technology 48 (4) (1999) 1203–1215.
- [19] W. Zhuang, B. Bensaou, K.C. Chua, Adaptive quality of service handoff priority scheme for mobile multimedia networks, IEEE Transactions on Vehicular Technology 49 (2) (2000) 494–505.
- [20] Y. Haung, Determining the optimal buffer size for short message transfer in a heterogeneous GPRS/UMTS network, IEEE Transactions on Vehicular Technology 52 (1) (2002) 216–225.
- [21] Y. Xiao, Y. Pan, J. Li, Design and analysis of location management for 3G cellular networks, IEEE Transactions on Parallel and Distributed Systems 15 (4) (2004) 339–349.
- [22] D. Staehle, K. Leibnitz, K. Tsipotis, QoS of internet access with GPRS, ACM Wireless Networks 9 (3) (2003) 213–222.
- [23] A. Downey, The structural cause of file size distributions, in: Proceedings of the IEEE MASCOT 2001, August 2001.



Sangheon Pack received his B.S. (2000, magna cum laude) and Ph.D. (2005) degrees from Seoul National University, both in computer engineering. He is a post-doctor fellow in Seoul National University. From 2005–2006, He was a post-doctor fellow in the Broadband Communications Research (BBCR) Group at University of Waterloo, Canada. Since 2002, he has been a recipient of the Korea Foundation for Advanced

Studies (KFAS) Computer Science and Information Technology Scholarship. He has been also a member of Samsung Frontier Membership (SFM) from 1999. He received a student travel grant award for the IFIP Personal Wireless Conference (PWC) 2003. He was a visiting researcher to Fraunhofer FOKUS, German in 2003. His research interests include mobility management, multimedia transmission, and QoS provision issues in the next-generation wireless/mobile networks. He is a member of the IEEE and ACM.



Taekyoung Kwon is an assistant professor in Multimedia & Mobile Communications Lab., School of Computer Science and Engineering, Seoul National University. He received his Ph.D., M.S., and B.S. degrees in computer engineering from Seoul National University in 2000, 1995, and 1993, respectively. He was a visiting student at IBM T.J. Watson Research Center in 1998 and a visiting scholar at the University of North

Texas in 1999. His recent research areas include radio resource

management, wireless technology convergence, mobility management, and sensor network.



Yanghee Choi received his B.S. in electronics engineering from Seoul National University, M.S. in electrical engineering from Korea Advanced Institute of Science, and Doctor of Engineering in Computer Science from Ecole Nationale Supérieure des Telecommunications (ENST) in Paris, in 1975, 1977 and 1984 respectively. Before joining the School of Computer Engineering, Seoul National University in 1991, he has been with

Electronics and Telecommunications Research Institute (ETRI) during 1977–1991, where he served as the director of Data Communication Section, and Protocol Engineering Center. He was a research student at Centre National d'Etude des Telecommunications (CNET), Issy-les-Moulineaux, during 1981–1984. He was also a Visiting Scientist to IBM T.J. Watson Research Center for the year 1988–1989. He is now leading the Multimedia Communications Laboratory in Seoul National University. He is also the director of Computer Network Research Center in Institute of Computer Technology (ICT). He was the editor-in-chief of Korea Information Science Society journals. He was the chairman of the Special Interest Group on Information Networking. He has been the associate dean of research affairs at Seoul National University. He was the president of Open Systems and Internet Association of Korea. His research interest lies in the field of multimedia systems and high-speed networking.