An Adaptive Mobility Anchor Point Selection Scheme in Hierarchical Mobile IPv6 Networks

Sangheon Pack Minji Nam Taekyoung Kwon * Yanghee Choi

301-518, Seoul National University, Shilim-dong, Kwanak-gu, Seoul, Korea

Abstract

In Hierarchical Mobile IPv6 (HMIPv6) networks, the mobility anchor point (MAP) is introduced to localize binding update messages destined to the home agent. In a large-scale wireless/mobile network, multiple MAPs may be deployed in order to provide more scalable and robust mobile services. In this case, it is important for a mobile node (MN) to select the most appropriate MAP among them. In this paper, we propose an adaptive MAP selection scheme for HMIPv6 networks. In the adaptive MAP selection scheme, an MN first estimates its session-to-mobility ratio (SMR). Then, based on its SMR, the MN chooses a MAP that minimizes the total cost, consisting of the binding update cost and packet delivery cost. In addition, the MN calculates two threshold SMR values, which adaptively trigger a new MAP selection procedure. If the estimated SMR is larger (or smaller) than the upper (or lower) threshold SMR value, the MN recalculates the total cost and re-selects a MAP that minimizes the total cost. Simulation results indicate that the adaptive MAP selection scheme achieves a lower total cost and a better load balancing than the previous schemes.

Key words: Hierarchical Mobile IPv6, Mobility anchor point (MAP), MAP selection scheme, Performance analysis.

1 Introduction

In IP-based wireless/mobile networks, there are different types of mobility agent (e.g. the home agent (HA) and foreign agent (FA)) that are used to

* Correspondent author: Taekyoung Kwon, E-mail: tkkwon@snu.ac.kr, Tel: +82-2-880-9105, Fax:+82-2-872-2045

Email addresses: shpack@mmlab.snu.ac.kr (Sangheon Pack), mjnam@kt.co.kr (Minji Nam), tkkwon@snu.ac.kr (Taekyoung Kwon), yhchoi@snu.ac.kr (Yanghee Choi).

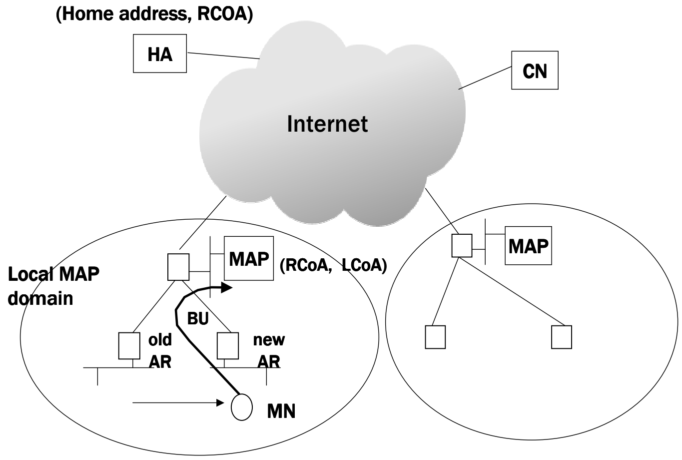


Fig. 1. Overview of HMIPv6 operation.

support host mobility. 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, 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 on the MAP's subnet. An 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 attributed to the MN's interface based on the prefix information advertised by an access router (AR).

Figure 1 illustrates the basic operations that are performed in HMIPv6 networks. An MN entering a MAP domain will receive an RA message containing information on the MAP. Then, the MN sends a BU message to the MAP, which binds its current location (i.e. LCoA) with an address on the MAP's subnet (i.e. RCoA). The MAP acts as a local HA and, as such, it receives all packets on behalf of the MNs it is serving. That is, the MAP decapsulates and forwards the received packets to the MN's current address. If the MN changes its current address within a MAP domain, it only needs to register the new address (i.e. new LCoA) 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 for the correspondent nodes (CNs).

In HMIPv6 networks, a MAP can exist at any level in the hierarchy, including at the level of the AR, and one or more MAPs can be located within the same network hierarchy and operate independently of one another. Especially when HMIPv6 is deployed in large-scale wireless/mobile networks, multiple MAPs are deployed to provide scalable mobile services. In this environment, it is important for an MN to select the most suitable MAP among the available MAPs, in order to reduce the total cost. Here, the total cost, incurred by an MN in an HMIPv6 network, consists of the binding update cost and the packet deliver cost [9–11].

An MN needs to consider several factors when selecting an optimal MAP that minimizes the total cost among the various MAPs available in a foreign network. In the HMIPv6 specification [2], two MAP selection schemes are recommended. The first of these is a distance-based selection scheme. In this scheme, an MN chooses to register with the furthest MAP, in order to avoid frequent re-registrations. This scheme is particularly efficient for fast MNs that are likely to perform frequent handoffs, because by choosing the furthest MAP, they reduce the probability of changing MAPs. Accordingly, the probability of informing the HA and all their CNs of this RCoA change is decreased.

However, each MN has different mobility characteristics, such as its subnet residence time, subnet crossing rate, etc. Therefore, for some MNs (e.g. slow MNs), the furthest MAP may not constitute the optimal solution. Furthermore, if all MNs select the furthest MAP as their serving MAPs, that MAP would become a single point of performance bottleneck, resulting in a high processing latency. This weakness of the first scheme motivates the preference-based scheme, which uses a `PREFERENCE` field within the MAP option [2]. For example, the `PREFERENCE` field can contain the MAP load information. The MAP option information is appended to and broadcasted with the RA message. However, it is difficult to choose an appropriate preference value for each MAP considering its dynamic traffic load. Furthermore, the MNs' mobility cannot be taken into account by the MAP. Therefore, this preference-based selection scheme may not be a effective solution.

Along with the distance-based selection scheme and the preference-based selection scheme, a few other MAP selection scheme have been proposed in the literature [5–8]. However, these schemes consider only certain specific characteristics and possess both advantages and disadvantages. In this paper, we propose a new MAP selection scheme called the *adaptive MAP selection scheme*. In contrast with the existing schemes, the adaptive MAP selection scheme takes both the MN's mobility and MN's session activity into consideration. Namely, an MN determines its serving MAP based on the estimated session-to-mobility ratio (SMR). By considering the SMR, the MN is able to select a more appropriate MAP with respect to its own mobility and session activity.

The remainder of this paper is organized as follows. In Section 2, we introduce the existing MAP selection schemes. In Section 3, we propose an adaptive MAP selection scheme. In Sections 4 and 5, we present the simulation model and results, respectively. Finally, Section 6 concludes this paper.

2 Previous Work

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 the home location register (HLR), has been widely used [4]. These mobility agents are statically deployed in the network and assigned to MNs based on their current locations. Recently, a two-tier cellular network [3,7] was introduced in order to provide higher system capacity, where MNs can be serviced by an upper tier cell (e.g. macro cell) or a lower tier cell (e.g. micro cell). Which tier to select is dependent on such factors as the MNs' mobility rate, the current network load at each tier, etc. However, in these cellular infrastructures, since there is only one available mobility agent (i.e. VLR) at each tier for visiting MNs, there is no problem of mobility agent selection in the traditional cellular networks.

However, in data-oriented Mobile IP networks, depending on the size of a foreign network, more than one mobility agents (e.g. MAP) may have to be deployed and selected by individual MNs, in order to provide more scalable and fault-tolerant mobile Internet services. Consequently, how an MN selects an appropriate mobility agent plays an important role in optimizing the location update/packet delivery procedures and reducing the signaling overhead in the network as a whole. As mentioned earlier, several selection schemes have been proposed in an attempt to solve this mobility agent selection problem.

2.1 Distance-based Selection Scheme (Furthest MAP Selection Scheme)

In the HMIPv6 specification [2], a distance-based selection scheme was recommended. In this scheme, an MN learns the hop distance between the MAP and the MN from the `DISTANCE` field in the RA message and the MN then registers with the furthest MAP. Without loss of generality, we assume that the area of the further MAP is larger than that of the closer MAP. That is, assuming the same moving speed, an MN at the further MAP's service is likely to request inter-MAP domain handoffs less frequently than an MN at the closer MAP's service. In this way, the MN can avoid performing frequent binding updates to the HA. As a result, this selection scheme minimizes the traffic amount of location registration that occurs at the HA and CNs.

However, the furthest MAP selection scheme has some drawbacks [7]. First, the furthest MAP may be closest to (or even collocated 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. 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 only increase the registration delay, because the hop distance from the MN to the furthest MAP is relatively longer than the one to the closer MAP.

2.2 Velocity-based Selection Scheme

In [5,6], MAP selection schemes based on the MN's velocity, were proposed. In these velocity-based selection schemes, a serving MAP is selected depending on the estimated velocity of the MN. Each of theses schemes can be divided into two functional components: the measurement of the velocity of the MN and the selection of which MAP to register with. First, to measure the velocity of the MN, the MN's binding update interval is used. Each MAP records not only the binding of the MN, but also the binding update time. When the MN moves into a new MAP domain, it is possible for the new MAP to get the MN's residence time in the previous MAP domain, by comparing the old binding update time in the previous MAP domain with the new binding update time. Then, the velocity is calculated by dividing the distance that the MN has traversed in the previous MAP domain by the residence time. However, measuring the actual distance is difficult, so that this scheme makes use of the concept of a *standard distance* that is predefined for each MAP domain, instead of the actual distance. In other words, the estimated velocity is used rather than the actual velocity. When the velocity of an MN is estimated, the previously estimated velocity is also taken into consideration, in order to avoid misinterpreting the fluctuating MN's velocity. Second, the method of selecting a MAP for an MN involves the network's informing the MN of the average binding update interval of all MNs in the MAP domain, so that the MN can select a suitable MAP to register with, by comparing its own binding update interval with the average value in the MAP domain. Here, the binding update interval represents the MN's residence time in a MAP. In addition, to avoid creating any bottleneck points and to achieve load balancing, two techniques are used. The first technique is to limit the number of MNs in a MAP domain. When a MAP receives a BU message, if the number of registered MNs exceeds a threshold value, the MAP forwards the BU message to the next candidate MAP. The other way is to use a new entity called the user agent

(UA), which is created by the network when an MN enters a new MAP domain. In this case, the UA decides the mobility type of the MN (i.e. fast and slow MNs) and selects the serving MAP.

2.3 Topology-based Selection Scheme

In general, it is difficult to measure the velocity of MNs and the measurements/estimates are often inaccurate. Therefore, MNs may not always be registered with an adequate level MAP. In [8], a new MAP selection scheme that uses the MAP topology instead of the MN's velocity measurement was proposed. In this scheme, an MN determines its relative mobility based on the constructed MAP topology. Each MAP advertises its information to all MNs, and the MNs keep track of the MAP topology in the form of a tree. To construct the MAP tree topology, a new IPv6 neighbor discovery option that contains three informational fields was defined: `LIFETIME`, `DISTANCE`, and `IP ADDRESS`. This option travels with the RA message containing the MAP information throughout the foreign network. The gateway MAP (i.e. root of the tree) initiates the tree construction process, by sending out the first RA message with the MAP option. As the RA message propagates away from the gateway MAP, other MAPs append their information to the RA message. Obviously, the length of the RA message gradually increases until the AR is reached. Each AR also maintains a local MAP list, which includes all the MAPs along the shortest tree path from the gateway MAP to the AR itself. The AR periodically sends out the MAP list together with its RA message. Using this list, the MN can find all the MAPs on the tree path from the gateway MAP to the AR and their respective hop distances. When the MN enters a new MAP domain, it selects the gateway MAP (or the furthest MAP) by default. Then, the MN begins to search for a suitable MAP to replace the gateway MAP during a predefined period called the *search interval*. At the end of the search interval, the MN identifies those MAPs that are common in every RA message, and the one closest to the MN is selected to replace the gateway MAP. 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 latest received MAP list. Then, a new MAP will be selected from the latest MAP list. However, the topology-based scheme requires modifications to the RA message and its performance is highly dependent on the length of the search interval.

3 Adaptive MAP Selection Scheme

MAP selection schemes proposed in [5–8] take the mobility rate (i.e. the MN's velocity) into account, but these schemes do not consider the session activity. However, as shown in [9–11], the performance of HMIPv6 is dependent on session activity as well as mobility and hence the cost function of HMIPv6 can be expressed as a sum of the binding update cost and the packet delivery cost. Although the binding update cost can reflect the MN's mobility, it cannot take the MN's session activity into consideration. In short, the previous MAP selection schemes may not select an optimal MAP in terms of the total cost. For example, a fast MN selects a further MAP in some previous schemes. However, if the fast MN's session arrival rate is too high, the packet delivery cost will increase accordingly and the further MAP may then result in a higher total cost than closer MAPs. To overcome this problem, we propose an adaptive MAP selection scheme. In the proposed scheme, an MN selects its serving MAP based on the session-to-mobility ratio (SMR). The SMR is the ratio of the session arrival rate to the mobility rate, which is a key factor to minimize the total cost.

3.1 Overview

The adaptive MAP selection scheme consists of four procedures. Followings are an overview of each procedure. Detailed descriptions for each procedure will be elaborated in the next subsections.

- **Initialization**: An MN collects all RA messages sent from the available MAPs in the foreign network. From these RA messages, the MN can obtain information for each MAP, e.g. the hop distance, network load. Using the MAP information, the MN makes an available MAP list (AML).
- **SMR estimation**: For each measurement interval (MI), the MN estimates its SMR by measuring the number of handoffs and session arrivals. At the same time, the MN updates its SMR and compares the estimated SMR with the two SMR threshold values.
- **Threshold Determination**: To select the optimal MAP adaptively, two SMR threshold values are calculated by an iterative method.
- **MAP Selection**: If the new 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 that minimizes the total cost.

Figure 2 gives an overview of the adaptive MAP selection scheme. First of all, the MN learns the information of MAPs by the initialization procedure. After

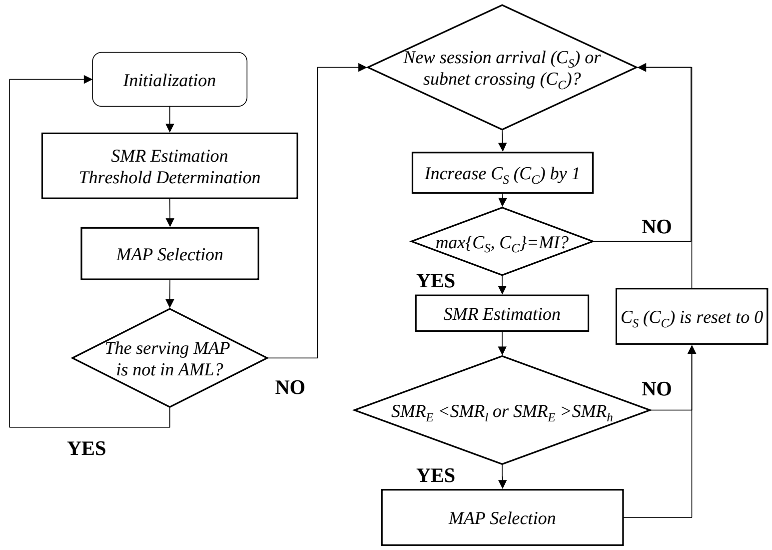


Fig. 2. Flow diagram for adaptive MAP selection: C_S and C_C denotes the counters for session arrivals and subnet crossings, respectively.

then, the MN estimates its SMR and determines two SMR thresholds. Based on the estimated SMR, the MN can choose the serving MAP. In addition, the MN monitors the number of session arrivals and intra-MAP handoff (i.e. subnet crossing). When the monitored count of either handoffs or session arrivals is equal to the measurement interval, the MN re-estimates the SMR. At the same time, the estimated SMR is compared with high and low SMR thresholds. When the estimated new SMR is larger than the upper SMR threshold value or smaller than the lower SMR threshold value, the new MAP selection procedure is invoked. On the other hand, if the estimated new SMR is between the lower SMR threshold value and and the upper SMR threshold value, the serving MAP is not changed. If the current serving MAP does not exist in the AML, the MN performs the above procedures again.

3.2 Initialization

In the multiple MAP environments, an MN receives multiple RA messages. Then, the MN configures an AML. The AML consists of the MAP id list and hop distance vector ($\mathbf{D}$), which is a data structure used to store the hop distance to each MAP. The hop distance vector is defined as follows:

$$\mathbf{D} = (D[1], D[2], ..., D[N])$$

where N is the number of available MAPs at the current location and $D[i]$ is the hop distance from MAP i to the MN. Since the latency from each MAP to the MN is not the same, RA messages arrive at the MN at different times. Therefore, the MN collects RA messages during a predefined time interval (T), and the configuration of the hop distance vector is completed after this time interval.

3.3 SMR Estimation

In our scheme, each MN measures the number of session arrivals and subnet handoffs, in order to calculate its SMR. The SMR of an MN is defined as follows:

$$SMR_M = \frac{\text{Session arrival rate}}{\text{Mobility rate}} = \frac{N_S}{N_C}$$

where N_S and N_C are the number of sessions that arrive and the number of subnets visited by the MN during a specific time duration (i.e. the measurement interval), respectively.

For stable estimation of the SMR, we use an *exponentially weighted moving average (EWMA)* technique, in order to avoid the effect of sudden changes of session activity or mobility rate. Eq. (1) shows the SMR estimation using EWMA used in the adaptive selection scheme.

$$SMR_E(i+1) = \omega \cdot SMR_E(i) + (1-\omega) \cdot SMR_M \tag{1}$$

where $SMR_E(i)$ and SMR_M are the estimated SMR at the time i and the currently measured SMR, respectively. ω is a weighting parameter, where $0 < \omega < 1$.

In the adaptive MAP selection scheme, the frequency of invoking the new MAP selection procedure should be carefully decided. If this frequency is too high, the processing overhead will be too high. On the other hand, if this frequency is too low, the interval between MAP selection times is too long, which will make the selected MAP suboptimal as time goes by. In addition, in the case of IP networks, the session size could be very small and the subnet residence time of an MN with high mobility could be very short. Therefore, estimating the SMR during every session or every handoff may result in a significant processing overhead for an MN with a limited capability. To solve this problem, the measurement interval (MI) is introduced to avoid excessive computation overhead arising from the SMR estimation procedure. In other

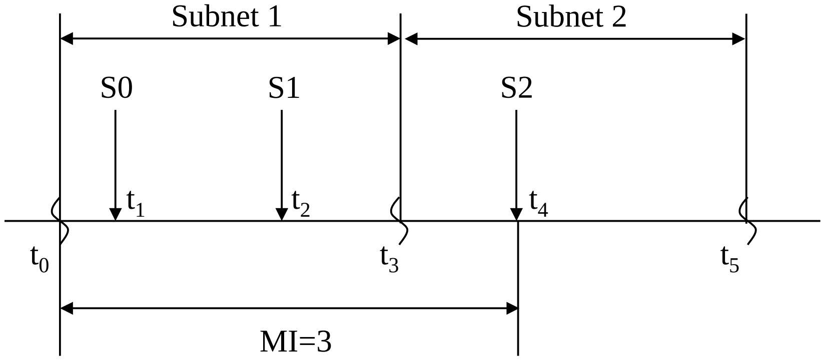


Fig. 3. Timing diagram for the SMR estimation.

words, an MN estimates its SMR only when either the number of handoffs or the number of session arrivals is equal to MI.

As an illustrative example, Figure 3 shows the timing diagram when MI is set to 3. Let t_0, t_3, and t_5 be the entering time to subnet 1, the entering time to subnet 2, and the departure time from subnet 2, respectively. In addition, let t_1, t_2, and t_4 be the arrival times of sessions 0, 1, and 2, respectively. Then, in this example, N_S is 3, which is equal to MI whereas N_C is 2. Therefore, SMR_M is 3/2=1.5.

Algorithm 1 SMR estimation procedure

while do
  $N_S \leftarrow 0$;
  $N_C \leftarrow 1$;
  if the MN crosses the subnet **then**
    $N_C \leftarrow N_C + 1$;
  end if
  if a new session arrives at the MN **then**
    $N_S \leftarrow N_S + 1$;
  end if
  if $\max\{N_C, N_S\}$ is equal to MI **then**
    $SMR_M \leftarrow N_S/N_C$;
    $SMR_E \leftarrow \omega \cdot SMR_E + (1-\omega) \cdot SMR_M$;
  end if
end while
return SMR_E;

Algorithm 1 shows the operation performed by the MN in order to calculate the estimated SMR that is used in the following sections.

3.4 MAP Selection

In the proposed scheme, an MN performs the MAP selection procedure based on the analytical cost function, which will be elaborated in Appendix A. In this analytical function, the hop distances between the MN and the MAPs are variables whereas the other hop distances (e.g. CN-HA, CN-MAP) are assumed to be constant despite different MAP locations. The other parameter values (e.g. unit transmission/processing cost) are the same for all available MAPs. Then, we formulate the total cost of MAP i, which is the sum of the binding update cost and packet delivery cost, as a function of the SMR and hop distance vector (refer to Appendix A).

$$C_T(SMR, D[i]) = C_{BU}(SMR, D[i]) + C_{PD}(SMR, D[i]) = f(SMR, D[i])$$

In the MAP selection procedure, the MN calculates the total cost function for each MAP i using the hop distance vector and the estimated SMR, and selects the MAP with the minimum total cost. Algorithm 2 shows how the optimal MAP is selected. In this procedure, id represents the identifier of the selected MAP.

Algorithm 2 MAP Selection

estimate the current SMR value (SMR_E);
$id \leftarrow 1$;
$i \leftarrow 1$;
while $i \leq N$ **do**
    if $C_T(SMR_E, D[i]) < C_T(SMR_E, D[id])$ **then**
        $id \leftarrow i$;
    end if
    $i \leftarrow i + 1$;
end while
return id;

3.5 Threshold Determination

Even though the optimal MAP at a given moment can be selected by the above Algorithm 2, the MN's SMR may change as time goes by and the selected MAP becomes suboptimal. To adaptively tackle this issue, the proposed MAP selection scheme defines two threshold SMR values: *lower threshold SMR* and *upper threshold SMR*. These threshold SMR values represent the two points at which there are likely to be other MAPs with a lower total cost compared to the current serving MAP. In other words, the threshold SMR values are the upper and lower bounds of the area between which the current serving MAP has the optimal or near-optimal total cost. The SMR threshold values are determined by Algorithm 3. In this procedure, we increase/decrease the SMR value by a unit of δ and check whether the current optimal MAP still holds for that increased/decrease value. This comparison is repeated up to the value of SMR_{min} (for the lower SMR threshold value) and SMR_{max} (for the upper SMR threshold value). $OptimalMAPSelection(N, D, SMR_E)$ refers to the procedure presented in Algorithm 2, where SMR_E is the estimated SMR.

Algorithm 3 Determination of SMR threshold values

```
set δ to a positive value;
SMR_min ← 0;
SMR_max ← INFINITE;
SMR_l ← SMR_E − δ;
SMR_h ← SMR_E + δ;
id ← OptimalMAPSelection(N, D, SMR_E);
while SMR_l > SMR_min do
   if id ≠ OptimalMAPSelection(N, D, SMR_l)) then
      SMR_l ← SMR_l + δ;
      break;
   else
      SMR_l ← SMR_l − δ;
   end if
end while
while SMR_h < SMR_max do
   if id ≠ OptimalMAPSelection(N, D, SMR_h)) then
      SMR_h ← SMR_h − δ;
      break;
   else
      SMR_h ← SMR_h + δ;
   end if
end while
return SMR_l and SMR_h;
```

4 Simulation Model

To show the effectiveness of the adaptive MAP selection scheme, we have conducted the comprehensive simulations. In this simulation, we compare four MAP selection schemes, i.e. the distance-based scheme (or the furthest MAP selection scheme), the velocity-based scheme, the topology-based scheme, and the adaptive scheme. For simplicity of description, we represent the furthest, velocity-based, topology-based, and adaptive MAP selection schemes as F, V, T, and A, respectively. In terms of simulation topology, a wrap-around model consisting of 64 subnets is used to avoid boundary effects in mobility [13]. These subnets are serviced by a two-level MAP hierarchy consisting of the higher level MAP (HMAP) and the lower level MAP (LMAP). A HMAP

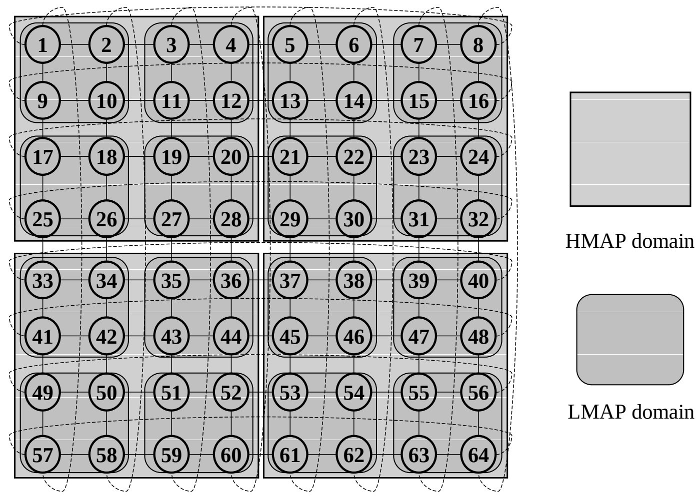


Fig. 4. Simulation topology.

domain covers 16 subnets whereas a LMAP domain covers 4 subnets. Figure 4 shows the two-level simulation topology.

In the simulation, the followings are assumed.

(1) Random walk mobility model [14] is used for user mobility. Thus, an MN moves to one of neighbor subnets with the same probability (i.e. 1/4) over the simulation topology shown in Figure 4.
(2) The residence time in a subnet follows a Gamma distribution with mean of $1/\lambda_m$ [15,16] as shown in Eq. (2).

$$f_R(t) = \frac{b^k t^{k-1}}{\Gamma(k)} e^{-bt} \tag{2}$$

where k is the shape parameter and $b = k\lambda_m$ is the scale parameter. $\Gamma(k)$ is the Gamma function, which is defined as $\int_0^\infty t^{k-1}\mathrm{e}^{-\mathrm{t}}dt$. The mean and variance of the Gamma distribution are $1/\lambda_m$ and $1/k\lambda_m^2$, respectively.
(3) Initial location of MNs are uniformly distributed among 64 subnets.
(4) There are two types of MNs: fast and slow MNs. The mean subnet residence times of fast and slow MNs are $60sec$ and $600sec$, respectively. Their variances are $600sec^2$ and $6000sec^2$, respectively.
(5) The session arrival process follows a Poisson distribution with rate of λ_S. Therefore, inter-session arrival time follows an exponential distribution in Eq. (3).

$$f_S(t) = \lambda_S e^{-\lambda_S} \tag{3}$$

(6) The session duration process follows a Pareto distribution with parameter a and k [17]. In our simulation, a is set to 0.78 and k is set to 180.

$$f_D(t) = \frac{a}{k}\left(\frac{k}{t}\right)^{a+1} \tag{4}$$

(7) Total simulation time is 10000 sec.

5 Simulation Results

In our simulations, we consider the following four scenarios depending on the ratio of fast MNs. Let ρ denote the ratio of fast MNs to total MNs. Thus, $1-\rho$ is the ratio of slow MNs to total MNs.

- Scenario 1: ρ is equal to 0.8 (high mobile environment).
- Scenario 2: ρ is equal to 0.6.
- Scenario 3: ρ is equal to 0.4.
- Scenario 4: ρ is equal to 0.2 (low mobile environment).

5.1 Binding Update Cost

First, we measured the number of BU messages to the MAP and to the HA in each scheme. The MAP BU message and the HA BU message have different effects on network performance in terms of signaling traffic. Therefore, we define a *weighted binding update cost* as follows:

$$BU = \alpha \cdot N_{HA} + \beta \cdot N_{HMAP} + \gamma \cdot N_{LMAP}$$

where α, β, and γ are weight values for the HA binding update, the HMAP binding update, the LMAP binding update, respectively. These weight values are set in proportion to the hop distances between the MN and the corresponding entities. α, β, and γ are set to 10, 2, and 1, respectively, unless stated otherwise.

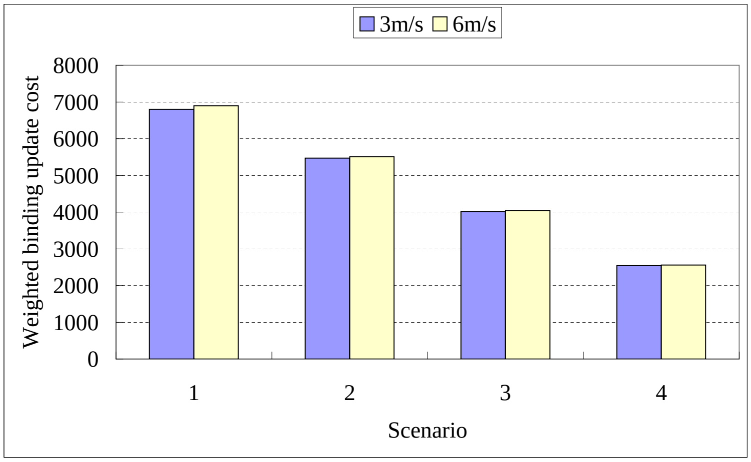


Fig. 5. Effect of threshold value: Velocity-based scheme (velocity thresholds: 3m/s and 6m/s).

5.1.1 Effect of Threshold Value

As mentioned before, V, T, and A use threshold values for MAP selection. [1] Therefore, the threshold value affects the performance of each scheme. In this section, we investigate the sensitivity of V, T, and A by changing threshold values.

Figures 5, 6, and 7 show the weighted binding update cost when two different threshold values are used in each scheme. Intuitively, the binding update cost is the highest in the case of scenario 1, while scenario 4 shows the lowest binding update cost. As shown in Figures 5, 6, and 7, the effect of threshold value is not significant. This is because only two types of MNs are used in simulations and their variance of subnet residence time is not high (i.e. the variance is 10 times of the average). However, the difference is expected to be notable in real environments where there are various types of MNs.

Specifically, the difference of the weight binding update costs in V is about 0.7%-1.4% whereas the difference in T is about 0.9%-1.5%. In the case of V, the MN's velocity is normalized by the EWMA scheme. Therefore, the effect of threshold on the binding update cost in V is minimal. On the other hand, A shows a litter higher difference of 1.8%-3.2%. In A, the SMR threshold should be dynamically changed depending on the mobility patterns in order to obtain a better performance. However, this simulation uses a static SMR threshold value under different mobility scenarios, in order to exhibit effects of threshold value.

[1] In the case of T, the search interval can be considered a kind of threshold value.

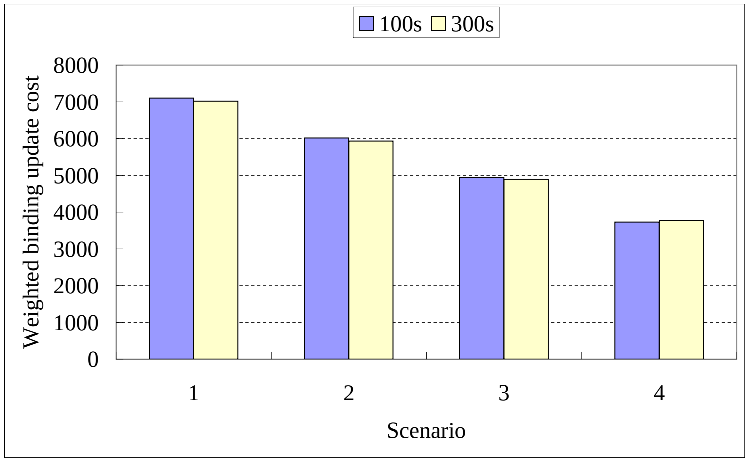


Fig. 6. Effect of threshold value: Topology-based scheme (search intervals: 100s and 300s).

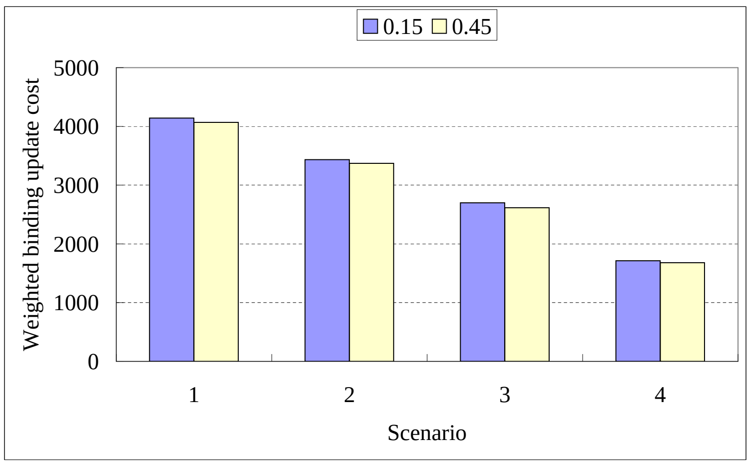


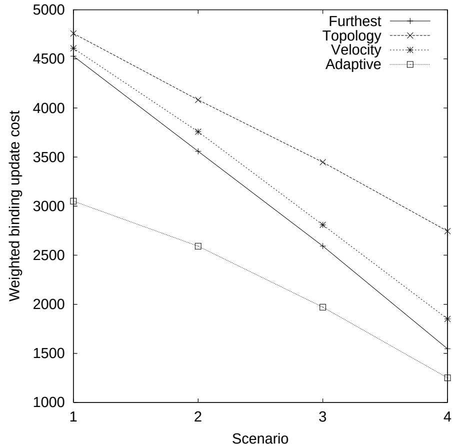
Fig. 7. Effect of threshold value: Adaptive scheme (SMR thresholds: 0.15 and 0.45).

Table 1
Different weight sets

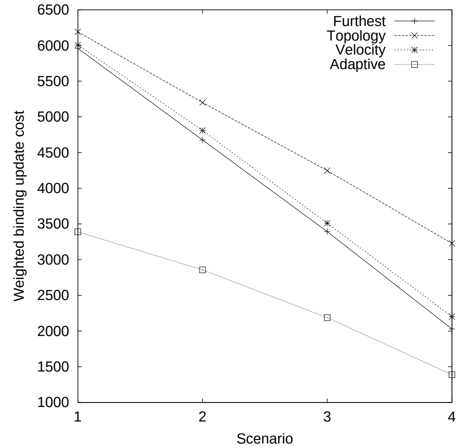
Set	α	β	γ
1	8	1	1
2	8	2	1
3	8	3	1
4	20	2	1

5.1.2 Effect of Different Weight Values

The weighted binding update cost is dependent on α, β, and γ values. Therefore, we evaluate the weight binding update costs in different weight sets shown in Table 1.



(a) Weight set 1

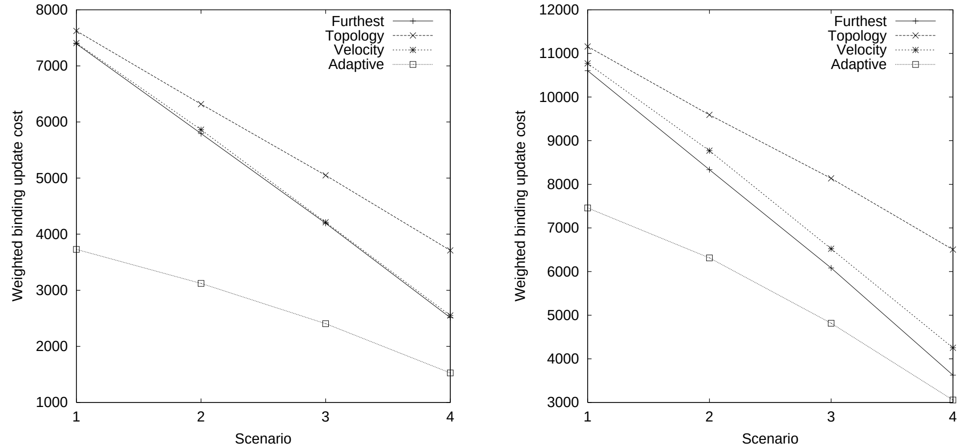


(b) Weight set 2

Fig. 8. Weighted binding update cost: Weight sets 1 and 2.

Overall, A shows the best performance in terms of binding update cost whereas the binding update cost of T is highest among four schemes. This is because T always selects the HMAP as a default MAP and searches a more suitable MAP during the search interval. In simulations, the search interval is set to $300sec$, so that the fast MN, which average residence time is $60sec$, uses the HMAP in most cases. Consequently, T results in a higher local binding update cost, because the HMAP binding update cost is larger than LMAP binding update cost. On the other hand, since V selects its serving MAP based on the estimation velocity, V enables for an MN to select a more suitable MAP depending on its mobility. Although V shows a less binding update cost than T, it requires a consistent monitoring of the MN's velocity. Therefore, V results in an implementation overhead.

In the case of F, the HMAP is always selected as the serving MAP. Although F can reduce the number of the HA binding updates, the number of HMAP binding updates increases. In the weight set 1, the binding updates to the HMAP and the LMAP are identical. Therefore, the increase of the number of HMAP binding updates does not increase the total binding update cost. In addition, the HA binding update cost is much higher than the MAP binding update cost in weight set 4. Hence, reducing the number of HA binding updates is the most critical issue. Consequently, F exhibits a relatively lower binding update cost than V and T, especially for weight sets 1 and 4. However, F shows a similar binding update cost to V in weight sets 2 and 3. This is because weight sets 2 and 3 have relatively high HMAP binding update costs.



(a) Weight set 3

(b) Weight set 4

Fig. 9. Weighted binding update cost: Weight sets 3 and 4.

5.2 Packet Processing Cost

As similar to the binding update cost, the packet processing costs[2] at the HMAP and LMAP should have different values. In general, the HMAP handles a larger number of MNs than the LMAP and the hop distance from the MN to the HMAP is longer than the one to the LMAP. Therefore, the HMAP processing cost is higher than the LMAP processing cost. Based on this observation, we define the *weighted packet processing cost* as follows:

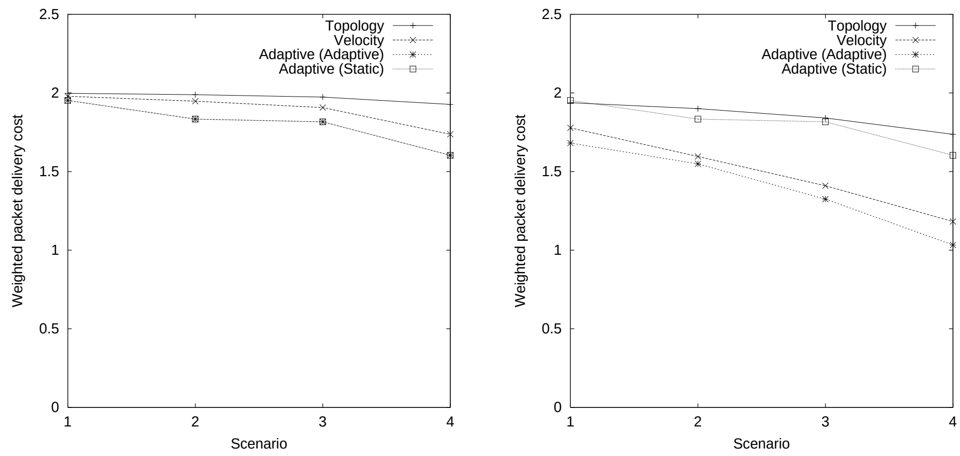
$$PD = \delta \cdot P_{HMAP} + \epsilon \cdot P_{LMAP}$$

where P_{HMAP} and P_{LMAP} are the probabilities that an MN is serviced by the HMAP and LMAP when a session arrives, respectively, i.e. $P_{HMAP} + P_{LMAP} = 1.0$. δ and ϵ are weight values for the HMAP and LMAP processings, respectively ($\delta > \epsilon$).

5.2.1 Effect of Session Arrival Rate

In this simulation, δ and ϵ are set to 2 and 1, respectively. In addition, we calculate the average weighted packet delivery cost per session. Namely, the weighted packet delivery cost of F is 2.0 because it always uses the HMAP, i.e. $P_{HMAP} = 1.0$ and $P_{LMAP} = 0.0$. Figure 10 shows the weighted packet delivery cost in different session arrival rates (λ_S). As described before, T uses the HMAP as a default serving MAP, so that an MN stays at the HMAP for a longer time. This characteristic results in a higher packet delivery cost.

[2] In this work, we consider the packet deliver cost within a foreign network, i.e. from the MAP to the MN.



(a) $\lambda_S = 0.01$ (b) $\lambda_S = 1$

Fig. 10. Effect of session arrival rate.

In Figure 10, A with static and adaptive SMR threshold values is evaluated. The static SMR threshold is calculated assuming that the average session arrival rate is set to 0.01. Consequently, A shows the lowest packet delivery cost when the session arrival rate is 0.01, while A exhibits the second highest packet delivery cost when the session arrival rate is 1, in the case of the static SMR threshold. This is because the calculated SMR is not appropriate to unexpected high session arrival rate and it results in a higher packet processing cost, if the session arrival rate is equal to 1. This result reveals that the performance of A is highly dependent on the session arrival rate in the static SMR threshold case. On the other hand, Figure 10 also depicts the weighted packet delivery cost when the adaptive SMR threshold is used. In this case, the SMR threshold is updated according to the change of session arrival rate. As a result, A shows the lowest packet delivery cost for all session arrival rates. Unlike to A, V and T show similar trends as the session arrival rate varies because they do not consider the session activity.

5.3 MAP load

Figure 11 plots the relative load at the HMAP. The relative HMAP load represents the probability that an MN is at the HMAP service. Therefore, the relative HMAP load of F is 1.0. Scenario 1 has more fast MNs than other scenarios, so that more MNs are serviced by the HMAP. On the other hand, as the number of slow MNs increases, the HMAP load decreases.

In scenario 1, 2, and 3, A shows the lowest HMAP load, whereas V shows the lowest HMAP load in scenario 4. However, the lowest HMAP load does not mean the best performance in terms of load balancing. This is because as the HMAP load decreases, the LMAP load increases. The most desirable scheme

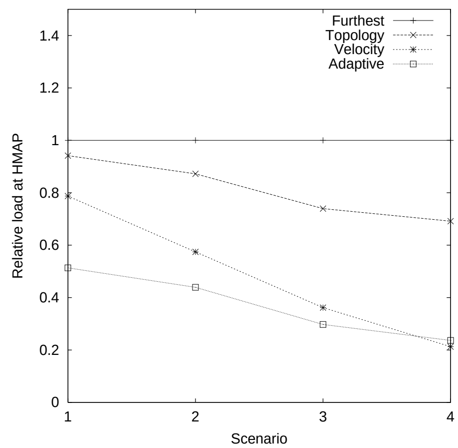


Fig. 11. Relative load at the HMAP.

may distribute the MAP load to the HMAP and LMAP as evenly as possible, i.e. the relative HMAP load is 0.5. Consequently, in terms of load balancing, A and V are better solutions than F and T.

6 Conclusion

In Hierarchical Mobile IPv6 networks, a MAP localizes mobility management by dealing with binding update messages and performs data tunneling by processing encapsulation/decapsulation. Considering this burden on the MAP, we believe that there should be multiple MAPs depending on the size of a foreign network. In this case, it is important to select the most appropriate MAP with respect to MNs. In this paper, we proposed an adaptive MAP selection scheme and conducted a comparative study for four MAP selection schemes: the distance-based, velocity-based, topology-based, and adaptive MAP selection scheme. The simulation results indicate that the adaptive MAP selection scheme has less binding update and packet delivery costs than other selection schemes. In addition, the adaptive MAP selection scheme achieves a well balanced distribution of the traffic load among MAPs. Although the adaptive MAP selection scheme does not require any complexity regarding to velocity measurement, it results in a degree of computation overhead at the side of MNs, since MNs should monitor the session arrival and mobility rates. However, this overhead can be mitigated in the near future via advances of mobile device technologies.

Acknowledgments

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, 2004, Korea.

A Cost Function Model

As similar to our previous works [9,12], we formulate the total cost (C_T) incurred in the HMIPv6 networks, which consists of the binding update cost (C_{BU}) and packet delivery cost (C_{PD}) .

A.1 Binding Update Cost

Let N_D and N_C be the number of MAP domain crossings (i.e. inter-MAP handoffs) and subnet crossings (i.e. intra-MAP handoffs), respectively. Then, the binding update cost per session can be expressed as follows:

$$C_{BU} = E(N_D) \cdot U_{HA} + E(N_C) \cdot U_{MAP} \tag{A.1}$$

where U_{HA} and U_{MAP} are unit binding update costs to the HA and to the MAP, respectively.

Let t_S, t_C, and t_D be the inter-session time, subnet residence time, and MAP domain residence time, respectively. Then, the subnet crossing probability (P_C) and the MAP domain crossing probability (P_D) per session are defined as follows:

$$P_C = \Pr(t_S > t_C)$$

$$P_D = \Pr(t_S > t_D)$$

Then, the probability density functions of N_D and N_C are derived from these subnet and domain crossing probabilities are given by Eqs. (A.2) and (A.3), respectively [19].

$$\Pr(N_C = n) = P_C^n \cdot (1 - P_C) \tag{A.2}$$

$$\Pr(N_D = n) = P_D^n \cdot (1 - P_D) \tag{A.3}$$

Then, the average number of subnet crossings and MAP domain crossings are given by Eqs. (A.4) and (A.5).

$$E(N_C) = \sum_{n=0}^{\infty} n \cdot P_C^n \cdot (1 - P_C) \tag{A.4}$$

$$E(N_D) = \sum_{n=0}^{\infty} n \cdot P_D^n \cdot (1 - P_D) \tag{A.5}$$

If t_S, t_C, and t_D are assumed to follow an exponential distribution with rate of λ_S, μ_C, and μ_D, respectively, P_C and P_D can be calculated as follows because of memoryless property of the exponential distribution [18].

$$\begin{aligned} P_C = \Pr(t_S > t_C) &= \int_0^{\infty} \Pr(t_S > \tau) \cdot \mu_C e^{-\lambda_C \tau} d\tau \\ &= \int_0^{\infty} e^{-\lambda_S \tau} \cdot \mu_C e^{-\mu_C \tau} d\tau = \frac{\mu_C}{\mu_C + \lambda_S} \end{aligned}$$

$$\begin{aligned} P_D = \Pr(t_S > t_D) &= \int_0^{\infty} \Pr(t_S > \tau) \cdot \mu_D e^{-\mu_D \tau} d\tau \\ &= \int_0^{\infty} e^{-\lambda_S \tau} \cdot \mu_D e^{-\mu_D \tau} d\tau = \frac{\mu_D}{\mu_D + \lambda_S} \end{aligned}$$

Then, the average number of subnet crossings and domain crossings are as follows:

$$\begin{aligned} E(N_C) &= \sum_{n=0}^{\infty} n \cdot \left(\frac{\mu_C}{\mu_C + \lambda_S}\right)^n \cdot \left(1 - \frac{\mu_C}{\mu_C + \lambda_S}\right) \\ &= \left(1 - \frac{\mu_C}{\mu_C + \lambda_S}\right) \cdot \frac{\frac{\mu_C}{\mu_C + \lambda_S}}{\left(1 - \frac{\mu_C}{\mu_C + \lambda_S}\right)^2} = \frac{\mu_C}{\lambda_S} \end{aligned}$$

$$\begin{aligned} E(N_D) &= \sum_{n=0}^{\infty} n \cdot \left(\frac{\mu_D}{\mu_D + \lambda_S}\right)^n \cdot \left(1 - \frac{\mu_D}{\mu_D + \lambda_S}\right) \\ &= \left(1 - \frac{\mu_D}{\mu_D + \lambda_S}\right) \cdot \frac{\frac{\mu_D}{\mu_D + \lambda_S}}{\left(1 - \frac{\mu_D}{\mu_D + \lambda_S}\right)^2} = \frac{\mu_D}{\lambda_S} \end{aligned}$$

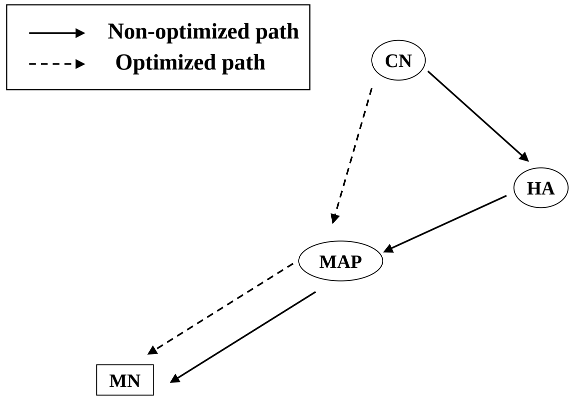


Fig. A.1. Packet delivery in HMIPv6 networks

Consequently, the binding update cost per session is as follows.

$$C_{BU} = \frac{\mu_D}{\lambda_S} \cdot U_{HA} + \frac{\mu_C}{\lambda_S} \cdot U_{MAP} \tag{A.6}$$

In this model, the SMR is defined as λ_S/μ_C. The domain crossing rate is inversely proportional to the root of the number of subnets (n) in the MAP domain and proportional to the subnet crossing rate [20]. That is, $\mu_D = \mu_C/\sqrt{n}$. U_{HA} and U_{MAP} are dependent on only the hop distance from the MN (i.e. they are regardless of mobility and session activity). Therefore, Eq. (A.7) is derived from Eq. (A.6).

$$C_{BU} = \frac{1}{SMR} \cdot (\frac{1}{\sqrt{n}} \cdot U_{HA} + U_{MAP}) = g(SMR, D) \tag{A.7}$$

A.2 Packet Delivery Cost

Since HMIPv6 supports route optimization, there are two types of paths as shown in Figure A.1: non-optimized and optimized paths. In HMIPv6 networks, a CN check its binding cache before it sends some packets. If there is a matching binding entry for the destination MN, the CN sends the packets to the MN's current location, i.e. the MN's RCoA in HMIPv6 networks (optimized path). Otherwise, the CN sends the packets to the MN's HA and the HA tunnels the received packets to the MN's current location (non-optimized path). After receiving the tunneled packets, the MN sends a binding update message to the CN. Then, the CN directly sends the subsequent packets to the MN's RCoA.

Consequently, the packet delivery cost per session is given by Eq. (A.8).

$$C_{PD} = \eta \cdot L_S \cdot P_N + (1 - \eta) \cdot L_S \cdot P_O \tag{A.8}$$

where η is the ratio of packets transiting the HA before the completion of the BU procedure. L_S is the average session size. P_N and P_O denote the packet delivery cost in the case of non-optimized path and optimized path, respectively.

In IP networks, the packet delivery cost consists of the transmission cost and processing cost. The transmission cost is the cost incurred in the packet delivery through a network, which is proportional to the hop distance of the packet delivery path. On the other hand, the processing cost is associated with the packet processing, e.g. routing table lookup, mapping table lookup, etc. Therefore, P_N and P_O can be calculated as follows.

$$P_N = T \cdot D_N + \Theta_{HA} + \Theta_{MAP}$$

$$P_O = T \cdot D_O + \Theta_{MAP}$$

where T, D_N, and D_O are the unit transmission cost, the hop distance of non-optimized path, and the hop distance of optimized path, respectively. Θ_{HA} and Θ_{MAP} denote the processing cost at the HA and MAP, respectively. Since the processing cost is proportional to the number of MNs served by the mobility agent, the processing cost of the higher level MAP is higher than that of the lower level MAP. This is because the higher level MAP services a larger number of MNs than the lower level MAP. Roughly, we can assume that the hop distances between the CN and the HA, the CN and the MAP, and the HA and the MAP are fixed. Then, D_N and D_O are functions of the hop distance (D) between the MAP and the MN. Other parameters (e.g. T, Θ_{HA}, Θ_{MAP}, η, etc.) are assumed as input parameters. Consequently, C_{PD} is a function of D (i.e. $C_{PD} = h(D)$).

As mentioned before, the total cost per session (C_T) is the sum of C_{BU} and C_{PD}. As shown in Eq. (A.9), the total cost is a function of SMR and hop distance from the MN to the MAP.

$$C_T = C_{BU} + C_{PD} = g(SMR, D) + h(D) = f(SMR, D) \tag{A.9}$$

Biography

Sangheon Pack. He received 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 the School of Computer Science and Engineering at the Seoul National University, Korea. He is a member of the IEEE and ACM. During 2002-2005, he was 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, wireless multimedia transmission, and QoS provision issues in the next-generation wireless/mobile networks.

Minji Nam. She received her B.S. and M.S degrees in Computer Science and Engineering from Seoul National University in 2003 and 2005, respectively. From 2005, she has worked on Portable Internet Development Team for Korea Telecom. Her research interests are mobile networks, portable internet technology (IEEE 802.16) and Mobile IPv6.

Taekyoung Kwon. He 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 He received 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 Superieure 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 director of Data Communication Section, and Protocol Engineering Center. He was research student at Centre National d'Etude des Telecommunications (CNET), Issy-les-Moulineaux, during 1981-1984. He was also 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 director of Computer Network Research Center in Institute of Computer Technology (ICT). He was editor-in-chief of Korea Information Science Society journals. He was chairman of the Special Interest Group on Information Networking. He has been associate dean of research affairs at Seoul National University. He was president of Open Systems and Internet Association of Korea. His research interest lies in the field of multimedia systems and high-speed networking.

References

[1] D. Johnson, C. Perkins, and J. Arkko, "Mobility Support in IPv6," IETF RFC 3775, June 2003.

[2] H. Soliman, C. Castelluccia, K. Malki, and L. Bellier, "Hierarchical Mobile IPv6 Mobility Management (HMIPv6)," Internet Draft, IETF, draft-ietf-mipshop-hmipv6-03.txt (work in progress), October 2004.

[3] Y. Lin and I. Chlamtac, "Heterogeneous Personal Communications Services: Integration of PCS Systems," *IEEE Communications Magazine*, vol. 34, no. 9, Sep. 1996.

[4] I. Akyildiz, J. Mcnair, J. Ho, H. Uzunalioglu, and W. Wang, "Mobility Management in Next-Generation Wireless Systems," *Proc. IEEE*, vol. 87, no. 8, August 1999, pp. 1347-1384.

[5] K. Kawano, K. Kinoshita, and K. Murakami, "Multilevel Hierarchical Distributed IP Mobility Management Scheme for Wide Area Networks," in *Proc. IEEE ICCCN*, 2002.

[6] K. Kawano, K. Kinoshita, and K. Murakami, "A Mobility-Based Terminal Management in IPv6 Networks," *IEICE Transactions on Communications*, vol. E85-B, n. 10, Oct. 2002.

[7] Y. Xu, H. C. J. Lee, and V. L. L. Thing, "A Local Mobility Agent Selection Algorithm for Mobile Networks," in *Proc. IEEE ICC*, May 2003.

[8] V. Thing, H. Lee, and Y. Xu, "Designs and Analysis of Local Mobility Agents Discovery, Selection and Failure Detection for Mobile IPv6," in *Proc. IEEE MWCN*, Sep. 2002.

[9] S. Pack and Y. Choi, "A Study on Performance of Hierarchical Mobile IPv6 in IP-based Cellular Networks," *IEICE Trans. Commun.*, vol. E87-B, no. 3, March 2004.

[10] J. Xie and I. Akyildiz, "A Distributed Dynamic Regional Location Management Scheme for Mobile IP," *IEEE Transactions on Mobile Computing*, vol. 1, no. 3, July 2002.

[11] W. Ma and Y. Fang, "Dynamic Hierarchical Mobiltiy Management Strategy for Mobile IP Networks," *IEEE Journal on Selected Areas on Communications*, vol. 22, no. 4, May 2004, pp. 664-676.

[12] S. Pack, T. Kwon, and Y. Choi, "A Comparative Study of Mobility Anchor Point Selection Schemes in Hierarchical Mobile IPv6 Networks," in *Proc. ACM MobiWac 2004*, September 2004.

[13] Y. Lin and V. Mak, "Eliminating the Boundary Effect of a Large-Scale Personal Communication Service Network Simulation," *ACM Transactions on Modeling and Computer Simulation*, vol. 4, no. 2, April 1994.

[14] I. Akyildiz, Y. Lin, W. Lai, R. Chen, "A new random walk model for PCS networks," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 7, July 2000.

[15] Y. Fang, I. Chlamtac and Y. Lin, "Portable movement modeling for PCS Networks," *IEEE Transactions on Vehicular Technology*, vol.49, no.4, pp.1356-1363, July 2000.

[16] Y. Fang and I. Chlamtac, "Analytical Generalized Results for The Handoff Probability in Wireless Networks," *IEEE Transactions on Communications*, vol.50, no.3, 396-399, March 2002.

[17] A. Balachandran, G. M. Voelker, P. Bahl, and P. V. Rangan, "Characterizing User Behavior and Network Performance in a Public Wireless LAN," *ACM SIGMETRICS Performance Evaluation Review*, vol. 30, no. 1, pp. 195-205, June 2002.

[18] L. Kleinrock, "Queuing Systems Volume 1: Theory," *John Wiley & Sons*, 1975.

[19] Y. Xiao, Y. Pan, and J. Li, "Design and Analysis of Location Management for 3G Celluar Networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 15, no. 4, April 2004, pp. 339-349.

[20] K. Wang and J. Huey , "A cost effective distributed location management strategy for wireless networks," *ACM/Kluwer Wireless Networks*, vol. 5, no. 4, 1999, pp. 287-297.