DYNAMIC CONTROL OF THE ADDRESS BINDING UPDATE FOR MOBILE NODES IN A HIERARCHICAL MOBILE IP NETWORK

Sang-Yong Kim   Hideaki Takagi
University of Tsukuba

Abstract   Hierarchical Mobile Internet Protocol version 6 (HMIPv6) has been proposed by the Internet Engineering Task Force (IETF) as an enhancement of the MIPv6. It is designed to reduce the amount of required signaling and to speed up the hand-off operation for mobile connection by means of hierarchical routing. In the HMIPv6, a router called the Mobile Anchor Point (MAP) serves mobile nodes (MNs) to aid their address binding and update as a local home agent. However, as the number of MNs increases in a MAP domain, the resource of the MAP becomes scarce, which increases the chance of failing in admitting new and hand-off MNs. Therefore some control scheme is needed in order to guarantee the quality of service (QoS) of the network.

    In this paper, we propose a dynamic control of the address binding update for the MNs in an HMIPv6 network. We consider three types of MNs entering the MAP domain, namely, a new MN, a hand-off MN in the sleep mode, and a hand-off MN in the active mode. We impose different costs on rejection of requests from these MNs, since the forced termination of an ongoing communication causes greater pain to a user than blocking a new MN. An optimal admission policy is obtained from a semi-Markov decision process with respect to the number of MNs present in a MAP domain. Based on this optimal policy, we calculate the probabilities of blocking each type of MN. We show numerically that our control reduces the probability of rejecting active mode hand-off MNs at little expense of blocking new MNs. It is also shown that our dynamic control outperforms the static control based on the guard channel scheme.

Keywords: Telecommunication, hierarchical mobile IP network, binding update, dynamic admission control, semi-Markov decision process, value iteration algorithm

1.   Introduction

The continuous growth and sophistication in mobile wireless communication has been demanding successive enhancement of protocol support in networking. First, the Mobile Internet Protocol version 4 (MIPv4) was proposed as a standard by the Internet Engineering Task Force (IETF) in 1996 [6–8]. It has several functional entities such as a mobile node (MN), a home agent (HA), and a foreign agent (FA). The HA is a router on a MN's home network that delivers packets to the MN by maintaining its location information when it is away from home. The FA is a router on a network visited by the MN that provides routing for the MN while registered. Every MN is assigned an IP address called a home address by the HA on the home network where the MN is registered originally. When the MN is away from the home network, it is assigned a care-of address (CoA) temporarily by the FA on a foreign network. The association of a MN's home address with its CoA, along with the remaining lifetime of their association, is called binding. A MN sends Binding Update (BU) to its HA everytime it moves. The location of the MN is tracked with the binding information while it is away from its home network. When a correspondent node (CN) sends a packet to the MN with its home address, the packet is delivered to the MN at that

address. If the MN is away from home, the HA receives the packet and forwards it to the FA on a foreign network which the MN is visiting. This process is called triangle routing.

The triangle routing in MIPv4 causes traffic overhead between the home and foreign networks in a wide area as well as a large latency in the hand-off of MNs locally. To avoid these problems, the Mobile IP version 6 (MIPv6) [9] and then the Hierarchical Mobile IP version 6 (HMIPv6) [16] have been proposed. In the MIPv6, a MN sends BUs to its HA as well as all CNs it communicates with. Once a CN has learned the MN's CoA, it may route the packet directly to the MN bypassing the HA. However, it takes time to send the BU to the HA that is typically far away. The HMIPv6, an enhancement of the MIPv6, is designed to reduce the amount of required signaling and to speed up the hand-off operation for mobile connection by means of hierarchical routing. In the HMIPv6, there is no FA. Instead, a special router called the Mobile Anchor Point (MAP) serves MNs under its coverage as their local HA to aid them in binding and updating IP addresses. It reduces the hand-off related latency, because MNs can register its bound IP address at the MAP more quickly than at a remote HA.

However, as the number of MNs increases in a MAP domain, the MAP suffers from resource starvation. This is caused by the BU operation and the encapsulation/decapsulation of every packet destined to all the MNs present in the MAP domain. The amount of necessary resource depends on the application running on each MN. Therefore, there is the maximum number of MNs, depending on the application, that can be connected for a router. The quality of service (QoS) in a domain which serves a large number of MNs may not keep up with that in other domains with low density of MNs. Therefore, some control scheme is needed in order to guarantee the same level of QoS throughout the network.

Admission control in a mobile communication network can be static (i.e., independent of the instantaneous traffic condition) or dynamic. For the call admission control of voice traffic in a cellular network, a static control based on the guard channel (also called cutoff priority) scheme was first studied by Hong and Rappaport [3]. In this scheme, a certain number of channels are reserved for exclusive use by the hand-off traffic. Their model for a single cell has been extended to handle the hand-off traffic from neighboring cells in homogeneous and inhomogeneous networks [10, 12]. One-dimensional birth-and-death processes are used to model the stochastic process for the number of calls existing in a cell. Dynamic admission control was proposed by Yang and Geraniotis [15], who employed a semi-Markov decision process with the value iteration algorithm to obtain the optimal policy for the admission of voice and data traffic. Ohmikawa et al. [4] have refined their model for voice calls by taking into account hand-off calls in cellular environment. As for the control of binding update for MNs in a hierarchical mobile IP network, Pack et al. [5] consider a static, guard channel based scheme for the call admission control scheme proposed by Fang and Zhang [1]. Vasilache et al. [14] propose a threshold-based load balancing policy that uses a model of data sharing among multiple HAs in a mobile IP network.

In this paper, we present a dynamic control method of the address binding and updating for the MNs in an HMIPv6 network. We consider three types of MNs entering the MAP domain, namely, a new MN, a hand-off MN in the sleep (not in communication) mode, and a hand-off MN in the active (in communication) mode. We impose different cost for blocking them, because the forced termination of ongoing communication causes greater pain to a user than blocking a new MN. An optimal admission policy is obtained from a semi-Markov decision process with respect to the number of MNs present in a MAP domain. Once the optimal policy is obtained, we calculate the probabilities of blocking each type of MN. The hand-off rates of MNs are computed from the generation rate of new MNs.

We show numerically to what extent the control parameters have impact on the admission performance. For comparison purpose, we also analyze the static control method based on the guard channel scheme with a birth-and-death process model.

The rest of the paper is organized as follows. In Section 2, we describe the address binding update operation for MNs in an HMIPv6 network in more detail. We then propose analytic models for studying the performance of the static guard channel scheme and that of the dynamic admission control scheme of binding update. In Section 3, we define the performance measures such as the probability of blocking new MNs, the probability of rejecting hand-off MNs, and the resource utilization at the MAP. The calculation of these measures is shown for both static and dynamic control cases. Section 4 presents the numerical results with discussion, including the comparison among dynamic control cases with different parameters and the comparison with the static control case. Concluding remarks are given in Section 5.

2. System Model

In this section, we first describe the address binding and update operation for MNs in an HMIPv6 network. We assume that our network is homogeneous such that all MAP domains have the same stochastic properties. For the modeling purpose, we introduce three types (new, sleep mode hand-off, and active mode hand-off) MNs appearing in a single MAP domain. Our problem is to control the admission of each type of MNs so as to minimize the cost associated with the rejection of MNs. We first present a threshold based static control method using the guard channel scheme. The static method is modeled by a birth-and-death process. We then propose a dynamic admission control method. We keep track of a sequence of events that each MN experiences by following the state-dependent admission policy. The dynamic admission control method is studied by means of a semi-Markov decision process with the value iteration algorithm for obtaining the optimal policy.

2.1. Binding update in an HMIPv6 network

In a mobile IPv4 network, it is always necessary that every MN registers its location to the HA in the home network. When the MN moves to a foreign network, the location registration takes time, which will result in the deterioration of hand-off quality. The HMIPv6 [16] has been proposed to speed up the location registration. In an HMIPv6 network, a MN which is away from its home network is served by an MAP in a local domain. As long as the MN moves within the local domain, the location registration to the home network is omitted. The default router of each MN is its Access Router (AR). An AR belongs to one or more MAP domains. The AR sends out periodically to the MNs the router advertisement (RA) message that includes the information about the MAP domain.

Figure 1 depicts the basic structure of an HMIPv6 network. According to [16], when a MN moves into a new MAP domain, it receives the RA message containing the global IP address of the MAP. The MN then coins two care-of addresses: the Regional Care-of Address (RCoA) and the Local Care-of Address (LCoA). The RCoA is an address on the MAP's subnet formed by combining the MN's interface identifier and the subnet prefix of the MAP's global address. The LCoA is the on-link CoA on a MN's interface. The MN sends the binding of RCoA and LCoA to the MAP, and the binding of home address and RCoA to the HA and CNs it communicates with. The BU information along with its lifetime is saved in the binding cache at the MAP, HA and CNs. The lifetime of the BU information is the time interval during which the RCoA is valid.

If the MN changes its current address within a local MAP domain, it only needs to
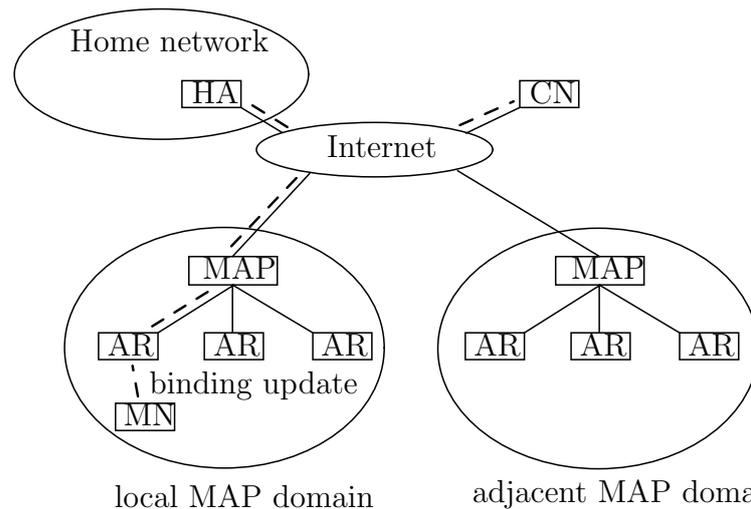
Figure 1: Basic structure of a hierarchical mobile IPv6 network

register the new address with the MAP. This is called local binding update. Hence, only the RCoA needs to be registered with the HA and the CNs. The RCoA does not change as long as the MN moves within a MAP domain. This makes the MN's mobility transparent to the HA and the CNs it is communicating with.

2.2. Three modes of mobile nodes

We focus on a single MAP domain. For the modeling purpose, we assume that there are three types of MNs entering the MAP domain, namely, a new MN, a sleep mode hand-off MN (called "a sleep MN" for short), and an active mode hand-off MN (called "an active MN" for short). Table 1 shows the distinction of new, sleep and active modes of MNs.

Table 1: Comparison of new, sleep and active modes of MNs

State of MN when entering a MAP domain	New MN	Sleep MN	Active MN
Power	Off	On	On
Update of BU information	No	Yes	Yes
TCP session	No	Yes	Yes
Exchange of user data (datagram)	No	No	Yes
Pain for a rejected MN	None	Little because of no effects on the user data exchange	Big because user data exchange is terminated

Suppose that there is a MN in the MAP domain with its power off. The user of that MN turns on its power when he wants to communicate with a CN. After a while, the MN sends the BU information to its HA. This is considered to be the arrival of a new MN in the MAP domain. The TCP session is then established between the MN and a CN. When an active MN enters a MAP domain, it is exchanging user data (datagram). When a sleep MN enters a MAP domain, it is not exchanging user data although the TCP session has been established. The MAP knows whether each MN in its domain is exchanging user data with
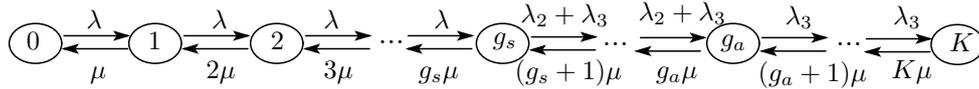
Figure 2: State transition rate diagram for a Markov chain representing the guard channel scheme

CN, because all the data go through the MAP. When a MN moves from a MAP domain to another by hand-off, the new MAP receives the information about the hand-off MN from the old MAP in order to avoid packet loss. In this way, the new MAP gets to know whether the hand-off MN is in the active mode or in the sleep mode.

2.3. Static admission control (guard channel scheme)

Let us first consider a guard channel based control method which is static in nature. We assume that the MAP can accommodate up to K MNs for binding updates, where K is specified in the router implementing the MAP function. The critical element in the guard channel scheme is the channel reservation. It reserves a fixed number of channels for MNs with high priority. Priority is given to active MNs over sleep MNs, and to sleep MNs over new MNs. In this case, two thresholds, denoted by g_s and g_a, are predefined. When the number of MNs exceeds the first threshold g_s, sleep and active MNs are admitted while new MNs are blocked. After the second threshold g_a, only active MNs are admitted; new and sleep MNs are blocked.

The BU information has a lifetime in the binding cache at the MAP, and each MN has a residence time in the MAP domain. The lifetime of the BU information is the time interval during which the information on the association of the home address and RCoA remains valid in the binding cache at the HA and MAP. The value for the lifetime of BU information is set by the MN in the header of the BU message. It is assumed for the mathematical tractability of our model that this lifetime has exponential distribution with mean $1/\mu_1$, which implies that it expires at rate μ_1. The residence time of a MN is the time interval during which the MN physically stays in a MAP domain. This is also assumed to be exponentially distributed with mean $1/\mu_2$, which implies that MNs go out of the MAP domain at rate μ_2. Thus the expiration of the BU information and the departure of the MN from the MAP domain are two independent processes. Since these events are assumed to occur with exponentially distributed inter-event times at rates μ_1 and μ_2, respectively, in the two processes, it follows that in the superposed process they occur with exponentially distributed inter-event times at rate $\mu = \mu_1 + \mu_2$.

We also assume that new, sleep, and active MNs are generated according to independent Poisson processes with rate λ_1, λ_2, and λ_3, respectively. The independence assumption for the arrivals of new and hand-off MNs has been justified in the past literature [3], [11, p.266].

On the above assumptions, we can say that the number of MNs existing in the MAP domain changes as a continuous time Markov chain, more specifically as a one-dimensional birth-and-death process. Figure 2 depicts the state transition rate diagram for the Markov chain with state space $\{0, 1, 2, \cdots K\}$. Assuming that the Markov chain is ergodic, let p_k be the steady state probability that there are k MNs in the MAP domain, where $0 \le k \le K$. Then we have the set of balance equations:

$$
\begin{aligned}
\lambda \cdot p_k &= (k+1)\mu \cdot p_{k+1} & 0 \le k \le g_s - 1 \\
(\lambda_2 + \lambda_3) \cdot p_k &= (k+1)\mu \cdot p_{k+1} & g_s \le k \le g_a - 1 \\
\lambda_3 \cdot p_k &= (k+1)\mu \cdot p_{k+1} & g_a \le k \le K - 1,
\end{aligned} \tag{2.1}
$$

and the normalization condition:

$$\sum_{k=0}^{K} p_k = 1. \tag{2.2}$$

Hence, the steady state probabilities are given by

$$p_k = \begin{cases} \dfrac{1}{k!}\left(\dfrac{\lambda}{\mu}\right)^k \cdot p_0 & 1 \le k \le g_s \\[2ex] \dfrac{\lambda^{g_s}(\lambda_2 + \lambda_3)^{k-g_s}}{k!\mu^k} \cdot p_0 & g_s + 1 \le k \le g_a \\[2ex] \dfrac{\lambda^{g_s}(\lambda_2 + \lambda_3)^{g_a-g_s}\lambda_3^{k-g_a}}{k!\mu^k} \cdot p_0 & g_a + 1 \le k \le K, \end{cases} \tag{2.3}$$

where

$$\frac{1}{p_0} = \sum_{k=0}^{g_s} \frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k + \sum_{k=g_s+1}^{g_a} \frac{\lambda^{g_s}(\lambda_2+\lambda_3)^{k-g_s}}{k!\mu^k} + \sum_{k=g_a+1}^{K} \frac{\lambda^{g_s}(\lambda_2+\lambda_3)^{g_a-g_s}\lambda_3^{k-g_a}}{k!\mu^k}. \tag{2.4}$$

We now introduce a cost structure. Situations that ongoing communications are forcibly terminated generally displease the user more severely than the case when he is not in communication or powered off. Thus we impose different rejection costs, γ_1, γ_2, and γ_3 for a new, sleep, and active MN, respectively, such that $\gamma_1 \le \gamma_2 \le \gamma_3$. The cost quantifies the relative strength of pain that a rejected MN feels. It is true in both new and sleep modes of MN that there is no transmission of user data (datagram). However, a sleep MN has registered its location in the binding cache at the MAP, HA, and CN, and it has established the TCP session with CN, thus it is ready to exchange user data. On the other hand, a new MN has to complete such preparation before starting to exchange user data. Therefore, we assume that the damage to the user in the sleep mode is bigger than that in the new mode when the arrival is rejected.

Then the average cost with thresholds g_s and g_a in the static control method is given by

$$C(g_s, g_a) = \frac{1}{\lambda}\left[\lambda_1\gamma_1 \sum_{k=g_s}^{g_a-1} p_k + \lambda_2\gamma_2 \sum_{k=g_a}^{K-1} p_k\right]. \tag{2.5}$$

Therefore, we can determine the optimal values for thresholds g_s and g_a such that they minimize $C(g_s, g_a)$.

2.4. Dynamic admission control (semi-Markov process)

We next consider a dynamic control method for the address binding update of MNs. Let us again focus on a single MAP domain, which can accommodate up to K MNs for binding updates. We assume three types (new, sleep, and active) of MNs as defined in Section 2.2. The model and parameters of the stochastic behavior of MNs in the MAP domain are the same as for the static control method described in Section 2.3. If a new, sleep, or active MN arrives and the number of existing MNs is less than K, the MN is either admitted in the MAP domain or it is rejected according to the admission control policy.

In order to find the optimal policy, we employ a semi-Markov decision process in a fashion similar to [4, 15]. The process is observed when a binding update message of an MN expires its lifetime, a registered MN goes out of the MAP domain, a new MN is placed, a sleep MN arrives, or an active MN arrives. After observing one of these events, decision is made according to the policy and the corresponding cost is incurred as a consequence.
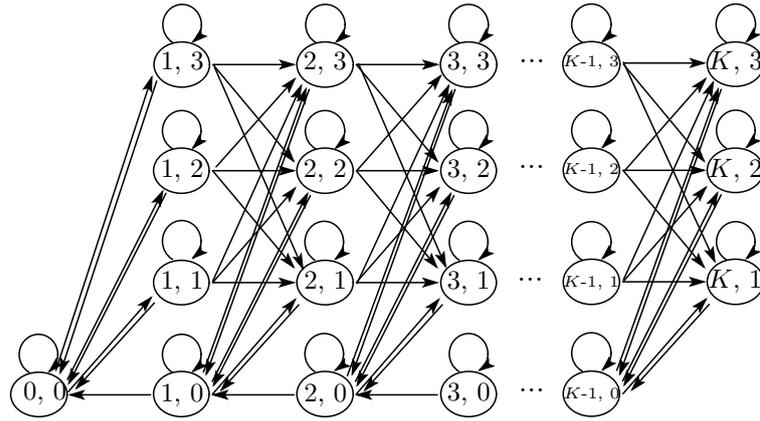
Figure 3: State transition diagram at decision epochs embedded in a Markov chain representing the dynamic admission control

We identify a set of Markovian decision epochs embedded in the continuous time process such that, if we specify the state at a decision epoch and provide information thereafter, we know the state at the next decision epoch. The state is defined by the pair (k, j), where k denotes the number of MNs existing in the MAP domain and j denotes how the state is entered. There are five cases of the state transition from state (k, j) at a decision epoch $(0 \le k \le K)$. First, if a new MN is admitted in the MAP domain, the chain moves to state $(k + 1, 1)$ $(0 \le k \le K - 1)$. Second, if a sleep MN is admitted, the chain moves to state $(k + 1, 2)$ $(0 \le k \le K - 1)$. Third, if an active MN is admitted, the chain moves to state $(k + 1, 3)$ $(0 \le k \le K - 1)$. Fourth, if a new, a sleep, or an active MN is placed but rejected, the chain stays in state (k, j) $(0 \le k \le K)$. Fifth, if the lifetime of the BU information expires or if a MN leaves the MAP domain as hand-off, the chain moves to state $(k - 1, 0)$ $(1 \le k \le K)$. Note that the last case is a fictitious decision epoch because no decision is made actually. Figure 3 depicts these transitions in the Markov chain embedded at decision epochs. The set of possible states in the process is denoted by

$$\mathcal{I} = \{\boldsymbol{x} = (k, j) \mid 0 \le k \le K, 0 \le j \le 3\} \backslash \{\boldsymbol{x} = (0, 1), (0, 2), (0, 3), (K, 0)\}. \qquad (2.6)$$

The probabilities of transition from state (k, j) to state $(k - 1, 0)$, $(k + 1, 1)$, $(k + 1, 2)$, $(k + 1, 3)$, and (k, j) are denoted by $q_{k,k-1}$, $q_{k,k+1}^{(1)}$, $q_{k,k+1}^{(2)}$, $q_{k,k+1}^{(3)}$, and $q_{k,k}$, respectively, such that $q_{k,k-1} + q_{k,k+1}^{(1)} + q_{k,k+1}^{(2)} + q_{k,k+1}^{(3)} + q_{k,k} = 1$. They are given in terms of $\mu = \mu_1 + \mu_2$, λ_1, λ_2, and λ_3 (let $\lambda = \lambda_1 + \lambda_2 + \lambda_3$) as well as the action parameters $a_{(k,1)}$, $a_{(k,2)}$, and $a_{(k,3)}$ of the dynamic admission control as follows:

$$q_{k,k-1} = \frac{k\mu}{\lambda + k\mu} \qquad\qquad 1 \le k \le K \qquad (2.7)$$

$$q_{k,k+1}^{(j)} = \frac{\lambda_j a_{(k,j)}}{\lambda + k\mu} \qquad\qquad 1 \le j \le 3, \ 0 \le k \le K - 1 \qquad (2.8)$$

$$q_{k,k} = \frac{\lambda - \lambda_1 a_{(k,1)} - \lambda_2 a_{(k,2)} - \lambda_3 a_{(k,3)}}{\lambda + k\mu} \qquad\qquad 0 \le k \le K, \qquad (2.9)$$

where

$$
a_{(k,1)} = \begin{cases} 0 & \text{if a new MN is rejected} \\ 1 & \text{if a new MN is admitted} \end{cases}
$$

$$
a_{(k,2)} = \begin{cases} 0 & \text{if a sleep MN is rejected} \\ 1 & \text{if a sleep MN is admitted} \end{cases} \tag{2.10}
$$

$$
a_{(k,3)} = \begin{cases} 0 & \text{if an active MN is rejected} \\ 1 & \text{if an active MN is admitted} \end{cases}
$$

when k MNs are served in the MAP domain ($0 \le k \le K - 1$), and

$$
a_{(K,1)} = a_{(K,2)} = a_{(K,3)} = 0. \tag{2.11}
$$

Note that $q_{k,k'} = q_{k,k'}^{(1)} = q_{k,k'}^{(2)} = q_{k,k'}^{(3)} = 0$ for $|k - k'| > 1$.

Assuming that the Markov chain is ergodic, let $p_{k,j}$ be the steady state probability for the chain to be in state $(k, j) \in \mathcal{I}$. Then we have the set of balance equations:

$$
p_{k,j} = \begin{cases} \left(\sum_{i=0}^{3} p_{k+1,i} \right) \cdot q_{k+1,k} + p_{k,0} \cdot q_{k,k} & j = 0 \\ \left(\sum_{i=0}^{3} p_{k-1,i} \right) \cdot q_{k-1,k}^{(j)} + p_{k,j} \cdot q_{k,k} & 1 \le j \le 3, \end{cases} \tag{2.12}
$$

and the normalization condition:

$$
\sum_{(k,j) \in \mathcal{I}} p_{k,j} = 1. \tag{2.13}
$$

Let $\widetilde{p}_{k,j}$ be the probability of state (k, j) at an arbitrary point in the continuous time domain. Then, from the theory of semi-Markov processes [2, Section 9-1], we have

$$
\widetilde{p}_{k,j} = \frac{p_{k,j} \cdot \eta_k}{\eta} \qquad (k, j) \in \mathcal{I}, \tag{2.14}
$$

where

$$
\eta_k = \frac{1}{\lambda + k\mu} \tag{2.15}
$$

is the mean sojourn time in state (k, j), and η is the average time interval between the successive points of state transitions, given by

$$
\eta = \sum_{(k,j) \in \mathcal{I}} p_{k,j} \cdot \eta_k. \tag{2.16}
$$

Furthermore, the marginal probability that there are k MNs in the MAP domain at an arbitrary time is given by

$$
p_k = \sum_{j=0}^{3} \widetilde{p}_{k,j} \quad 0 \le k \le K. \tag{2.17}
$$

2.5. Value iteration algorithm for the semi-Markov decision process

The value iteration algorithm [13, Section 3.4] is applied to our semi-Markov decision process for determining the optimal policy. To do so, we convert the semi-Markov decision model into a discrete time Markov decision model such that the average costs per unit time over the long term of each stationary policy are the same in both models. This is done by a method called uniformization. The uniformization is a technique to transform the original continuous time Markov chain with non-identical state transition rates into an equivalent continuous time Markov chain in which the state transition epochs are generated by a Poisson process at a uniform rate and the state transitions are described by a discrete time Markov chain that allows for fictitious transitions from a state to itself. In extending this uniformization technique to our semi-Markov decision process, we choose the uniform rate $1/\tau$ of the Poisson process as

$$0 < \tau < \min_{\boldsymbol{x} \in \mathcal{I}} \tau_{\boldsymbol{x}} = \frac{1}{\lambda + K\mu}, \tag{2.18}$$

where $\tau_{\boldsymbol{x}} = \eta_k = 1/(\lambda + k\mu)$ is the expected time until the next decision epoch in state $\boldsymbol{x} = (k, j)$. Then we consider that the decision epochs at state $\boldsymbol{x}$ occur at a uniform rate $1/\tau$ rather than at the state-dependent rate $1/\tau_{\boldsymbol{x}}$, but only a fraction $\tau/\tau_{\boldsymbol{x}}$ are real transitions out of state $\boldsymbol{x}$ and the remaining fraction $1 - \tau/\tau_{\boldsymbol{x}}$ leave the process in state $\boldsymbol{x}$. Therefore, the transition probabilities in the corresponding discrete time model is given by

$$\bar{P}_{\boldsymbol{x},\boldsymbol{x}'}(a_{\boldsymbol{x}}) = \begin{cases} \dfrac{\tau}{\tau_{\boldsymbol{x}}} P_{\boldsymbol{x},\boldsymbol{x}'}(a_{\boldsymbol{x}}) & \boldsymbol{x}' \neq \boldsymbol{x} \\[2mm] \dfrac{\tau}{\tau_{\boldsymbol{x}}} P_{\boldsymbol{x},\boldsymbol{x}}(a_{\boldsymbol{x}}) + \left\{1 - \dfrac{\tau}{\tau_{\boldsymbol{x}}}\right\} & \boldsymbol{x}' = \boldsymbol{x} \end{cases}, \tag{2.19}$$

where

$$P_{\boldsymbol{x},\boldsymbol{x}'}(a_{\boldsymbol{x}}) = \begin{cases} q_{k,k-1} & \boldsymbol{x}' = (k-1, 0) \\[2mm] q_{k,k+1}^{(j)} & \boldsymbol{x}' = (k+1, j) \qquad 1 \leq j \leq 3 \\[2mm] q_{k,k} & \boldsymbol{x}' = \boldsymbol{x} \\[2mm] 0 & \text{otherwise} \end{cases} \tag{2.20}$$

is the transition probability from state $\boldsymbol{x} = (k, j)$ to state $\boldsymbol{x}'$ if action $a_{\boldsymbol{x}}$ is chosen in the original semi-Markov process.

If the current state is $\boldsymbol{x} = (k, j)$ and action $a_{\boldsymbol{x}}$ is chosen, the average cost per unit time until the next decision epoch is given by $C(a_{\boldsymbol{x}})/\tau_{\boldsymbol{x}}$, where

$$C(a_{\boldsymbol{x}}) = (1 - a_{\boldsymbol{x}})\gamma_j \quad \boldsymbol{x} = (k, j), \tag{2.21}$$

which represents the instantaneous cost if action $a_{\boldsymbol{x}}$ is chosen at state $\boldsymbol{x} = (k, j)$. The pain experienced by a rejected user at state $\boldsymbol{x} = (k, j)$ is denoted by γ_j, where we let $\gamma_0 = 0$ and $\gamma_1 \leq \gamma_2 \leq \gamma_3$ as justified in Section 2.3.

The quantity $V_n(\boldsymbol{x})$ is introduced as the minimal total expected cost with n steps left to the time horizon when the current state is $\boldsymbol{x}$ in the discrete time process. Since the goal is to minimize the average cost per unit time over the long term, we must go backward in the time axis until the one-step difference $V_n(\boldsymbol{x}) - V_{n-1}(\boldsymbol{x})$ converges.

The value iteration algorithm for finding the optimal policy in the discrete time Markov decision model is described as follows:

1. Choose $V_0(\boldsymbol{x})$ such that $0 \leq V_0(\boldsymbol{x}) \leq \dfrac{1}{\tau_{\boldsymbol{x}}} \min\limits_{a_{\boldsymbol{x}} \in \{0,1\}} \{C(a_{\boldsymbol{x}})\}$ for all $\boldsymbol{x} \in \mathcal{I}$. Set $n := 1$.

2. Compute for $\boldsymbol{x} \in \mathcal{I}$

$$
V_n(\boldsymbol{x}) = \min_{a_{\boldsymbol{x}} \in \{0,1\}} \left[\frac{C(a_{\boldsymbol{x}})}{\tau_{\boldsymbol{x}}} + \frac{\tau}{\tau_{\boldsymbol{x}}} \sum_{\boldsymbol{x}' \in \mathcal{I}} P_{\boldsymbol{x},\boldsymbol{x}'}(a_{\boldsymbol{x}}) V_{n-1}(\boldsymbol{x}') + \left\{ 1 - \frac{\tau}{\tau_{\boldsymbol{x}}} \right\} V_{n-1}(\boldsymbol{x}) \right]
$$

   by determining the action $a_{\boldsymbol{x}}$ that minimizes the quantity in the brackets.

3. Compute the upper and lower bounds of the one-step difference by

$$
\begin{aligned}
M_n &= \max_{\boldsymbol{x} \in \mathcal{I}} \{V_n(\boldsymbol{x}) - V_{n-1}(\boldsymbol{x})\} \\
m_n &= \min_{\boldsymbol{x} \in \mathcal{I}} \{V_n(\boldsymbol{x}) - V_{n-1}(\boldsymbol{x})\}
\end{aligned}
\tag{2.22}
$$

4. If $0 \leq M_n - m_n \leq \varepsilon m_n$, then stop the algorithm with the optimal set of actions for all $\boldsymbol{x}$. Otherwise $n := n + 1$ and go to step 2.

Here, ε is a prespecified small positive constant (tolerance number) for stopping the iteration.

According to [13, p.269], the value iteration algorithm converges if the Markov chain is aperiodic for each policy. In the above, by choosing τ as in (2.18), we have $\bar{P}_{\boldsymbol{x},\boldsymbol{x}}(a_{\boldsymbol{x}}) > 0$ for all $\boldsymbol{x}$ and $a_{\boldsymbol{x}}$. Then the Markov chain is aperiodic, and the algorithm converges.

Once the optimal policy $\{a_{\boldsymbol{x}}; \boldsymbol{x} \in \mathcal{I}\}$ is determined, we can obtain the steady state probabilities $p_{k,j}$, then $\widetilde{p}_{k,j}; (k,j) \in \mathcal{I}$, and finally $p_k; 0 \leq k \leq K$ as shown in Section 2.4.

3. Performance Measures

In this section, we define some performance measures to compare the dynamic admission control method with various ratios, denoted by $\gamma_1{:}\gamma_2{:}\gamma_3$, of the pain experienced by rejected new MNs, sleep MNs, and active MNs. We also compare it with the static control method based on the guard channel scheme and with the case of no admission control. We will use the probability of blocking new MNs, the probability of blocking sleep MNs, and the probability of forced termination of active MNs. Furthermore, we consider the MAP's resource utilization as a measure of effective resource usage for each method.

3.1. Blocking and forced termination probabilities

There are three cases that irritate mobile users. The first case is that the registration request of a new user arriving at a MAP domain is rejected because the resource of the MAP is not sufficient, which results in the blocking of a new MN. The second case is that the binding update of a user who tries to cross the MAP boundary while not in communication is rejected, which is the blocking of a sleep MN. The last case is that the ongoing communication of a mobile user is forcibly terminated by the failed binding update when crossing the MAP boundary, which is the forced termination of an active MN. Therefore, we consider the probability of blocking new MNs, the probability of blocking sleep MNs, and the probability of forced termination of active MNs as performance measures from user's viewpoint.

In the guard channel scheme considered in Section 2.3, if a new MN is placed and there are g_s or more MNs in the MAP domain, it is blocked. Similarly, if a sleep MN arrives when there are g_a or more MNs in the MAP domain, it is blocked. Forced termination of an active MN occurs only if the MAP's capacity is full when it arrives. Therefore, the action in the guard channel scheme can be expressed as

$$
a_{(k,1)} = \begin{cases} 1 & 0 \leq k \leq g_s - 1 \\ 0 & g_s \leq k \leq K, \end{cases}
\tag{3.1}
$$

$$a_{(k,2)} = \begin{cases} 1 & 0 \le k \le g_a - 1 \\ 0 & g_a \le k \le K, \end{cases} \tag{3.2}$$

and

$$a_{(k,3)} = \begin{cases} 1 & 0 \le k \le K - 1 \\ 0 & k = K. \end{cases} \tag{3.3}$$

In both guard channel scheme and dynamic admission control, it is assumed that MNs arrive according to Poisson processes independently. Let P_{bn} and P_{bs} denote the probabilities of blocking new and sleep MNs, respectively. Let P_{fa} be the probability of forced termination of active MNs. Using the PASTA (Poisson Arrivals See Time Averages) property [2, Section 11-2], we have

$$P_{bn} = \sum_{k=0}^{K} [1 - a_{(k,1)}]p_k, \tag{3.4}$$

$$P_{bs} = \sum_{k=0}^{K} [1 - a_{(k,2)}]p_k, \tag{3.5}$$

and

$$P_{fa} = \sum_{k=0}^{K} [1 - a_{(k,3)}]p_k, \tag{3.6}$$

where $p_k; 0 \le k \le K$ is given in (2.3) for the guard channel scheme and in (2.17) for the dynamic control scheme.

3.2. Resource utilization at mobility anchor point

Another performance measure is the resource utilization U at a MAP. It is given as the ratio of the average number N of MNs present in the MAP domain to the capacity K of the MAP:

$$U = \frac{N}{K}, \tag{3.7}$$

where

$$N = \sum_{k=1}^{K} k \cdot p_k. \tag{3.8}$$

From the viewpoint of the efficiency in using the resource at the MAP, it is nice to have high value of U. However, as the utilization U increases, the blocking probabilities usually grow, resulting in the user's dissatisfaction.

4. Numerical Experiments

Numerical experiments have been carried out in order to evaluate the impact of each admission control on the performance. We first show a method for calculating the arrival rates of sleep and active MNs from that of new MNs. We then plot the values of performance measure against the new MN arrival rate. We compare and discuss the numerical results for various cases.

4.1. Calculation of the hand-off MN arrival rate

In the modeling and performance analysis in Sections 2 and 3, we have assumed that new, sleep, and active MNs arrive in the MAP domain independently. However, it is realistic to consider that hand-off MNs are generated when MNs leaving the adjacent domain come in. We present a technique for estimating the arrival rates of sleep and active MNs when the arrival rate of new MNs is given.

Suppose that the network consists of independent and statistically identical MAP domains. Then the inbound hand-off rate can be considered equal to the outbound rate to adjacent MAP domains. Given the steady state probabilities for the number of MNs existing in the MAP domain, the inbound hand-off rate λ_H from neighboring MAP domains can be determined by the fixed-point method [10, 12]. Then, the value of λ_H can be calculated by the following algorithm:

1. Initialize $\lambda_H = 0$.
2. Calculate $p_k; 0 \leq k \leq K$ as shown in Section 2.
3. Compute

$$\lambda_H = \sum_{k=1}^{K} (k\mu_2) \cdot p_k. \tag{4.1}$$

4. Repeat steps 2 and 3 until λ_H converges.

Next, we obtain the arrival rates λ_2 and λ_3 for the sleep and active MNs. To do so, let us assume that each MN alternates periods of active and sleep modes. The average durations of active and sleep mode periods are denoted by $1/\mu_a$ and $1/\mu_s$, respectively. Both durations are assumed to be exponentially distributed. The probability p_{active} that a MN is active at an arbitrary point in time is given by the theory of alternating renewal processes [2, Section 5-6] as

$$p_{active} = \frac{\mu_s}{\mu_a + \mu_s}. \tag{4.2}$$

Therefore, we get

$$\lambda_2 = (1 - p_{active}) \cdot \lambda_H \tag{4.3}$$

and

$$\lambda_3 = p_{active} \cdot \lambda_H. \tag{4.4}$$

In our numerical experiments, we have assumed that the average active period of a MN is $1/\mu_a = 2$ (minutes) and that the average sleep period of a MN is $1/\mu_s = 3$ (minutes). Therefore, the probability that a MN is active at an arbitrary point in time is $p_{active} = 0.4$.

For the dynamic admission control case, step 2 in the above algorithm is conducted with the interim values of λ_2 and λ_3 along with the mean inter-event time

$$\tau = \frac{1}{\lambda + 2K\mu} \tag{4.5}$$

for the Poisson process in the value iteration algorithm mentioned in Section 2.5. The choice in (4.5) satisfies (2.18), which guarantees the convergence of the value iteration algorithm. Actually we have chosen $\varepsilon = 10^{-3}$ in step 4 for stopping the iteration.

Finally we note that the inbound hand-off rate λ_H, calculated by (4.1), is proportional to the average number N of MNs present in the MAP domain, which is given in (3.8).

## 4.2. Performance results and discussion

It has been assumed throughout our experiments that the average lifetime of a binding update is $1/\mu_1 = 3$ (minutes) and that the average residence time in a MAP domain is $1/\mu_2 = 5$ (minutes). The capacity of the MAP is assumed to be $K = 20$. We have chosen this value as a representative number of the maximum simultaneous connections in some actual routers of intermediate size. We first examine the effectiveness of our dynamic admission control by means of the blocking probability of new MNs, the blocking probability of sleep MNs, and the forced termination probability of active MNs as well as the MAP's resource utilization.
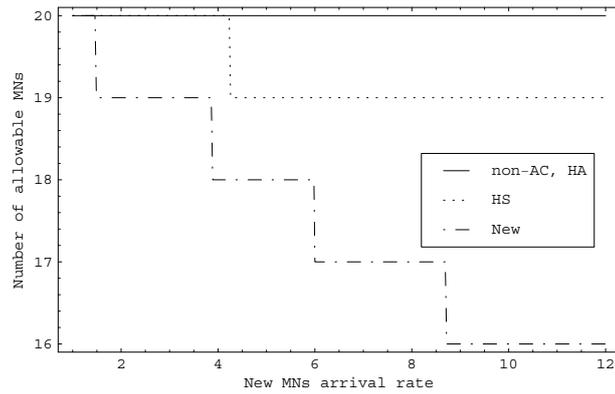
(a) $\gamma_1{:}\gamma_2{:}\gamma_3 = 1{:}1{:}5$
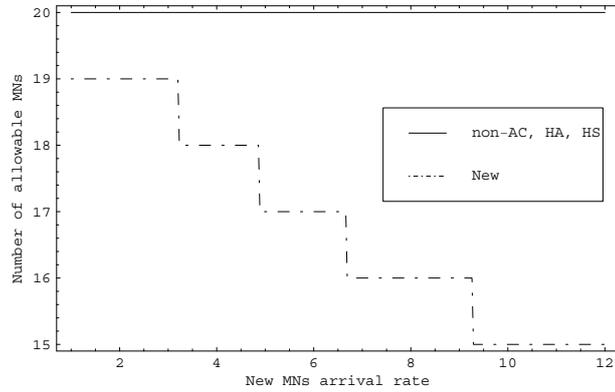
(b) $\gamma_1{:}\gamma_2{:}\gamma_3 = 1{:}5{:}5$
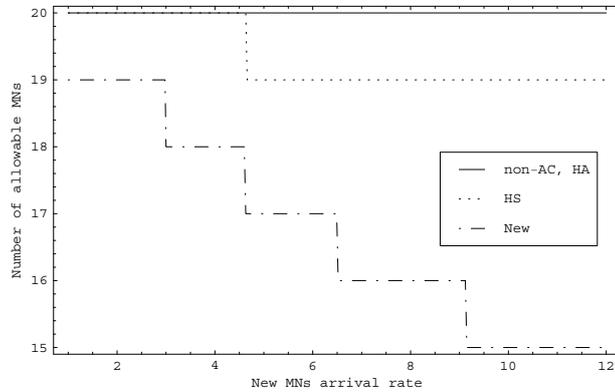
(c) $\gamma_1{:}\gamma_2{:}\gamma_3 = 1{:}2{:}10$

Figure 4: Maximum number of MNs admitted in a MAP domain by dynamic control

Figures 4(a)–(c) show the maximum number of new, sleep, and active MNs that can be admitted in a MAP domain as a result of optimization for the dynamic admission control with different cost ratios $\gamma_1{:}\gamma_2{:}\gamma_3$. In the legend of these figures, "non-AC" means that no admission control is adopted, "HA", "HS", and "New" refer to active MNs, sleep MNs, and new MNs, respectively. We first note that active MNs should always be admitted up to the capacity. Fewer new and sleep MNs can be admitted as the arrival rate λ_1 of new MNs increases. However, in Figure 4(b), sleep MNs are not rejected until the MAP's resource is exhausted, because its rejection cost is the same as for active MNs. No rejection of MNs occurs under the MAP's capacity in the non-AC case.
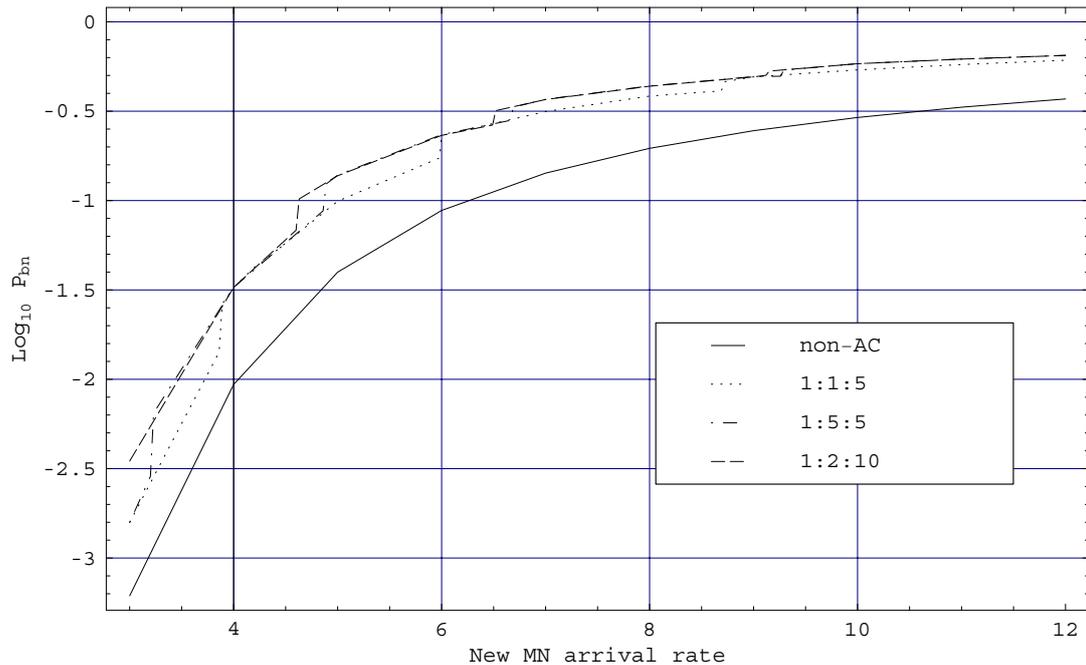
Figure 5: Blocking probability of new MNs by dynamic control with cost ratio $\gamma_1{:}\gamma_2{:}\gamma_3$

Figure 5 shows the blocking probability P_{bn} of new MNs as a function of the new MN arrival rate λ_1 for the optimal policy shown in Figures 4(a)–(c). Since our dynamic admission control restricts new MNs for the benefit of hand-off MNs, it is reasonable that the larger the ratio $\gamma_1{:}\gamma_2{:}\gamma_3$, the larger the blocking probability of new MNs. For example, at $\lambda_1 = 8$ the values of P_{bn} after implementation of the dynamic admission control with $\gamma_1{:}\gamma_2{:}\gamma_3 = 1{:}5{:}5$ and $1{:}2{:}10$ are 1.9 and 2.2 times larger, respectively, than that under the non-AC.

Similarly, Figure 6 shows the blocking probability P_{bs} of sleep MNs as a function of the new MN arrival rate λ_1. As sleep MNs have advantage over new MNs, P_{bs} is smaller under the dynamic admission control with larger ratio of $\gamma_1{:}\gamma_2{:}\gamma_3$. For example, at $\lambda_1 = 10$ the blocking probabilities when the dynamic admission control is adopted with $\gamma_1{:}\gamma_2{:}\gamma_3 = 1{:}5{:}5$ and $1{:}2{:}10$ are $1/100$ and $3/100$ times, respectively, that under the non-AC.

In Figure 7, we plot the forced termination probability P_{fa} of active MNs as a function of the new MN arrival rate λ_1. As active MNs have great advantage over new and sleep MNs under the dynamic admission control with large ratio $\gamma_1{:}\gamma_2{:}\gamma_3$, the forced termination probabilities are significantly smaller at high load conditions. This is the highlight of our dynamic control method. We may remark that the blocking probability in Figure 6 and the forced termination probability in Figure 7 drop abruptly, e.g., at $\lambda_1 = 7$ and $\lambda_1 = 10$ for
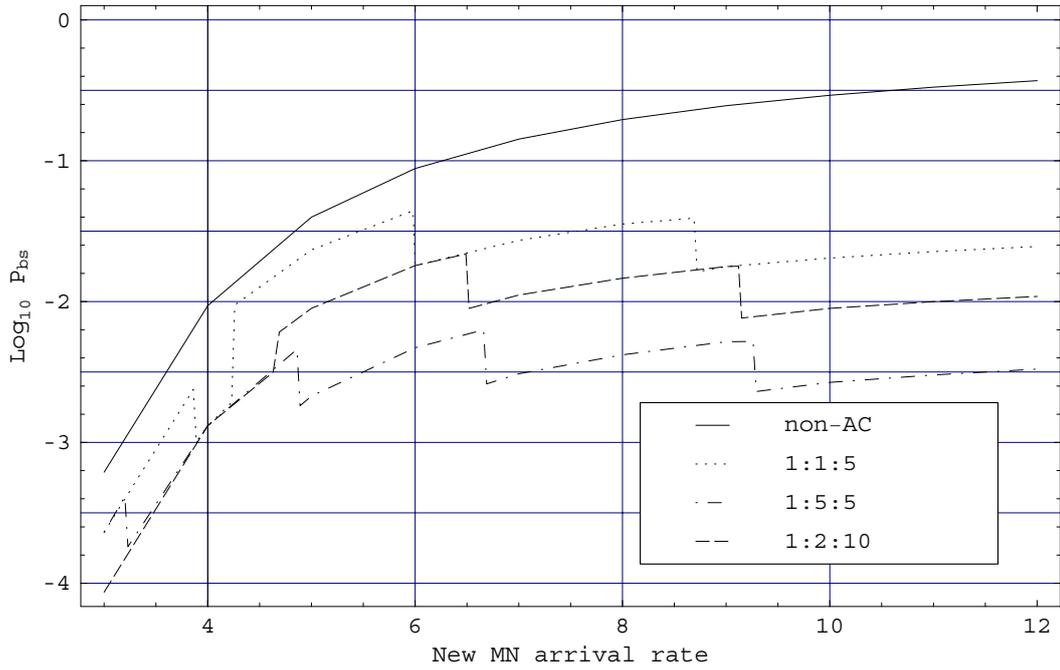
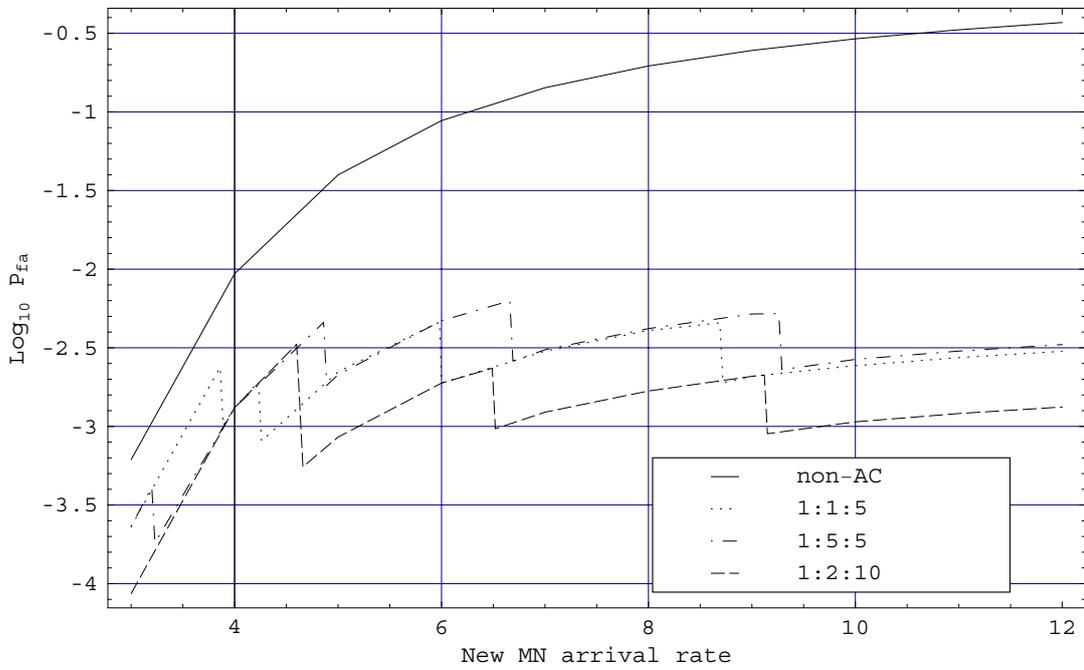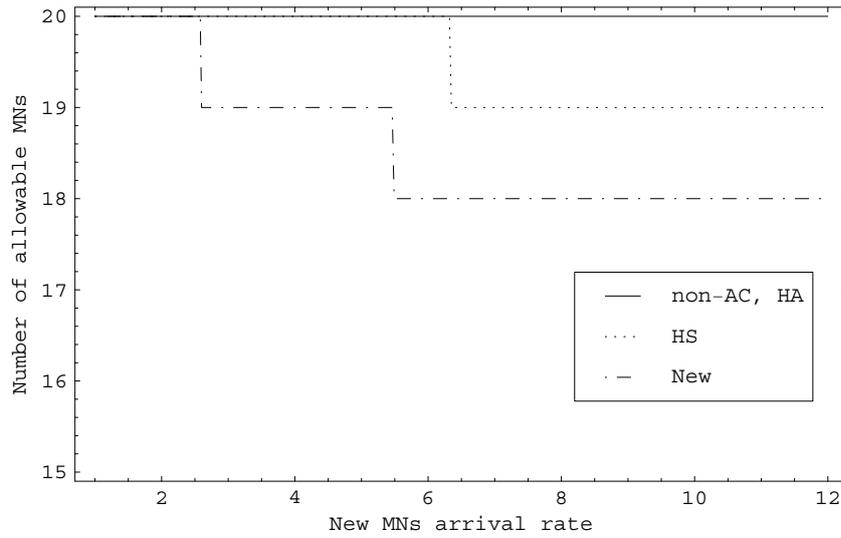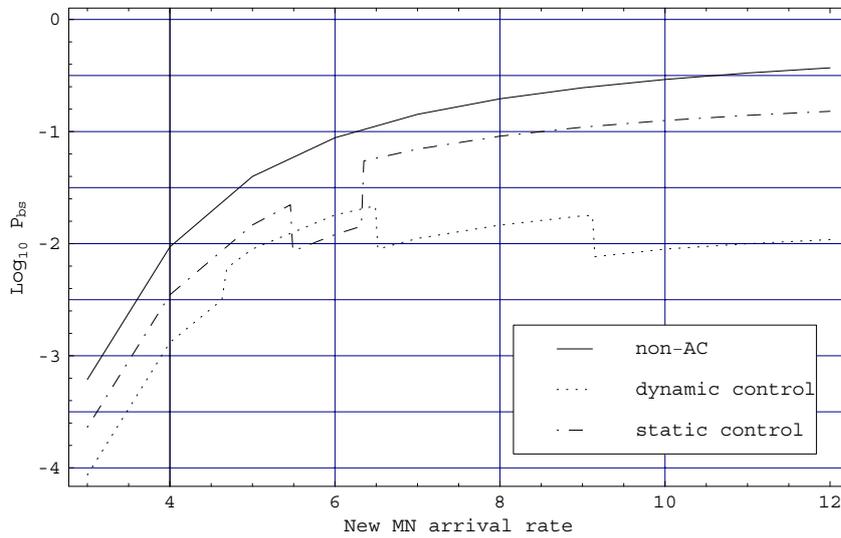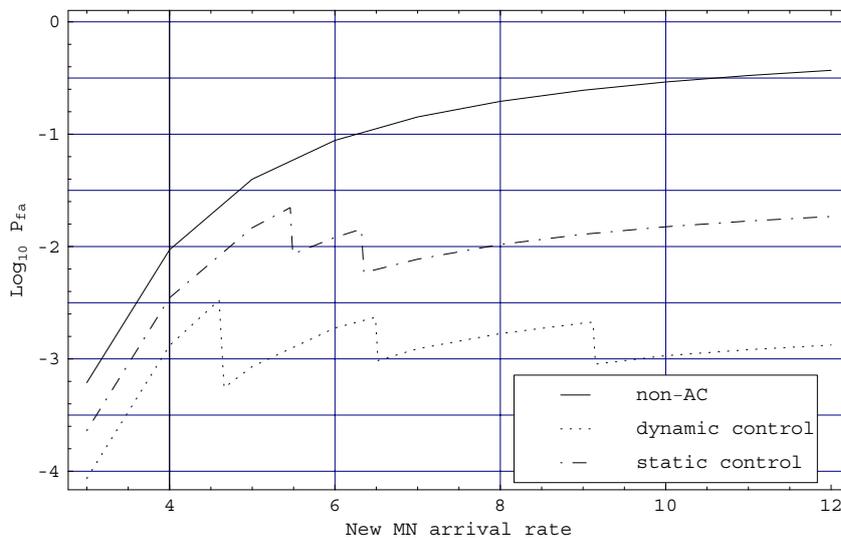Figure 6: Blocking probability of sleep MNs by dynamic control

Figure 7: Forced termination probability of active MNs by dynamic control

(a) Maximum number of MNs admitted in a MAP domain by static control

(b) Blocking probability of sleep MNs

(c) Forced termination probability of active MNs

Figure 8: Performance comparison of static and dynamic control (γ_1:γ_2:γ_3 = 1:2:10)
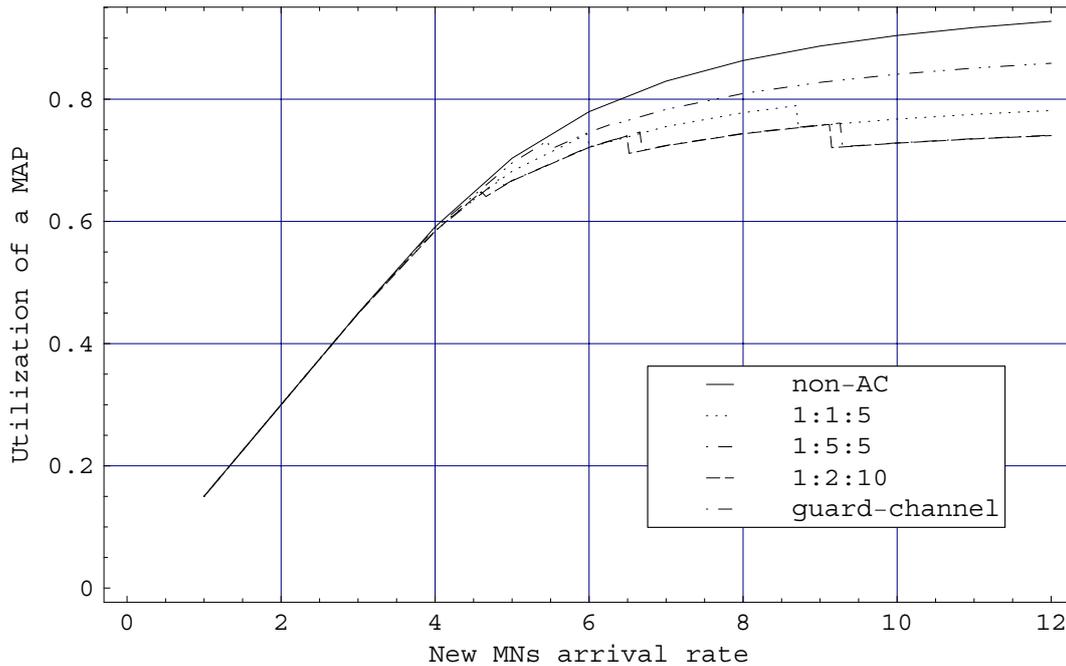
Figure 9: Resource utilization of a MAP

γ_1:γ_2:γ_3 = 1:5:5 and 1:2:10. The reason is that the number of MNs allowed in the MAP domain is decremented by one at those arrival rates as observed in Figures 4(b) and 4(c).

We now compare the performance of static and dynamic control when the cost ratio is γ_1:γ_2:γ_3 = 1:2:10. Figure 8(a) shows the maximum number of new, sleep, and active MNs that can be admitted in a MAP domain for the static control with optimal threshold values g_s and g_a. Comparing this with Figure 4(c) for the dynamic control, we observe that the dynamic control suppresses the admission of new MNs much more severely than static control at high load conditions. This is because the dynamic control tries to block the admission of new and sleep MNs in anticipation of accepting as many as active MNs in the future. Figures 8(b) and 8(c) compare the blocking probability P_{bs} of sleep MNs and the forced termination probability P_{fa} of active MNs, respectively, between the dynamic control and the static control each with its own optimal action policy. We see that P_{bs} and P_{fa} are much smaller with the dynamic control, although both are smaller than the case of no admission control. For example, at $\lambda_1 = 10$, the P_{bs} with the dynamic control is 7.1% and the P_{fa} with dynamic control is 8.3%, respectively, of those under the static control. From these numerical results, we may conclude that the dynamic control is by far more effective than the static control (although the former takes much more computational time that the latter).

Figure 9 shows the resource utilization of a MAP. The utilization values under the dynamic admission control with γ_1:γ_2:γ_3 = 1:1:5, 1:5:5, and 1:2:10 are 84.3%, 79.9%, and 79.8%, respectively, of the value under the non-AC at $\lambda_1 = 12$. In these cases, although the resource of a MAP is less utilized, the control leads to a certain reduction in the probability of blocking sleep MNs and the probability of forced termination of active MNs as shown in Figures 6 and 7, respectively. The MAP's resource utilization under static control is 92.6% of the non-AC value at $\lambda_1 = 12$, which is higher than those under dynamic control. From Figure 9 we can also say that the hand-off rate λ_H is higher for the static guard channel

scheme than the dynamic admission control cases with larger ratio $\gamma_1{:}\gamma_2{:}\gamma_3$.

5. Conclusion

In this paper, we have proposed static and dynamic admission control methods for the address binding update of MNs in an HMIPv6 network. We have developed admission control models using various stochastic processes such as a one-dimensional birth-and-death process, a two-dimensional Markov chain, a semi-Markov decision process, an alternating renewal process, and uniformization. We have conducted performance analysis of these models in terms of the blocking probability for the resource in the MAP domain. We have shown some numerical results leading to the comparison among dynamic control cases with different parameters and the comparison with the static control method based on the guard channel scheme as well as with the case of no admission control. According to the numerical experiments, our dynamic admission control reduces the blocking probability of sleep MNs and significantly the forced termination probability of active MNs at little expense of blocking new MNs. The static control is less effective than the dynamic control in terms of these performance measures.

Inherently in the optimization using Markov decision processes, the computational time may be an obstacle to apply our dynamic admission control method to real systems. However, our results can be used as a benchmark in opposition to those obtained by using quick approximate optimizing methods.

The mechanism of MAP selection [16] has not been considered in this paper. A MN entering a MAP domain receives router advertisement (RA) message containing information on one or more local MAPs. After receiving the RA message, the MN configures its LCoA and RCoA. If the MN receives several RA messages, it selects one of MAPs. This procedure is called MAP selection. If a MN is able to find a MAP which does not suffer from traffic congestion, the performance of the dynamic admission control may be enhanced. Estimate of the performance gain obtained by the MAP selection is one of our future works.

References

[1] Y. Fang and Y. Zhang: Call admission control schemes and performance analysis in wireless mobile networks. *IEEE Transactions on Vehicular Technology*, **51** (2002), 371–382.

[2] D.P. Heymann and M.J. Sobel: *Stochastic Models in Operations Research, Volume I: Stochastic Processes and Operating Characteristics* (McGraw-Hill Book Company, New York, 1982).

[3] D.H. Hong and S.S. Rappaport: Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures. *IEEE Transactions on Vehicular Technology*, **VT-35** (1986), 77–92.

[4] M. Ohmikawa, H. Takagi, and S.-Y. Kim: Optimal call admission control for voice traffic in cellular mobile communication networks. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, **E88-A** (2005), 1809–1815.

[5] S. Pack, T. Kwon, and Y. Choi: A mobility-based load control scheme at mobility anchor point in hierarchical mobile IPv6 networks. *IEEE Global Telecommunications Conference (GLOBECOM) 2004*, **6** (2004), 3431–3435.

[6] C.E. Perkins: IP mobility support. *Request for Comments (RFC) 2002* (1996).

[7] C.E. Perkins: Mobile IP. *IEEE Communications Magazine*, **35** (1997), 84–99.

[8] C.E. Perkins: *Mobile IP: Design Principles and Practices* (Reading, Addison-Wesley, Massachusetts, 1998).

[9] C.E. Perkins and D. B. Johnson: Mobility support in IPv6. *Proceedings of the Second Annual International Conference on Mobile Computing and Networking (MobiCom'96)* (1996), 27–37.

[10] K. Sakamaki and H. Takagi: Evaluation of call loss and forced termination probabilities in cellular communication systems. *The Transactions of the Institute of Electronics, Information and Communication Engineers B-II*, **J80-B-II** (1997), 231–238 (in Japanese).

[11] M. Schwartz: *Mobile Wireless Communications* (Cambridge University Press, Cambridge, United Kingdom, 2005).

[12] H. Takagi, K. Sakamaki, and T. Miyashiro: Call loss and forced termination probabilities in cellular radio communication networks with non-uniform traffic conditions. *IEICE Transactions on Communications*, **E82-B** (1999), 1496–1504.

[13] H.C. Tijms: *Stochastic Models: An Algorithmic Approach* (John Wiley & Sons, Chichester, England, 1994).

[14] A. Vasilache, J. Li, and H. Kameda: Threshold-based load balancing for multiple home agents in mobile IP networks. *Telecommunication Systems*, **22** (2003), 11–31.

[15] W-B. Yang and E. Geraniotis: Admission policies for integrated voice and data traffic in CDMA packet radio networks. *IEEE Journal on Selected Areas in Communications*, **12** (1994), 654–664.

[16] H. Soliman, C. Catelluccia, K.E. Malki, and L. Bellier: Hierarchical mobile IPv6 mobility management (HMIPv6). *IETF Network Working Group* (2004). http://www.ietf.org/internet-drafts/draft-ietf-mipshop-hmipv6-03.txt

Hideaki Takagi
Graduate School of Systems and Information Engineering
University of Tsukuba
1-1-1 Tennoudai, Tsukuba-shi
Ibaraki 305-8573, Japan
E-mail: `takagi@sk.tsukuba.ac.jp`