A Mobility-based Load Control Scheme at Mobility Anchor Point in Hierarchical Mobile IPv6 Networks

Sangheon Pack, Taekyoung Kwon, and Yanghee Choi
School of Computer Science and Engineering
Seoul National University, Seoul, Korea
Email: {shpack, tk}@mmlab.snu.ac.kr and yhchoi@snu.ac.kr

***Abstract*— In Hierarchical Mobile IPv6 (HMIPv6) networks, the mobility anchor point (MAP) handles binding update (BU) procedures locally to reduce signaling overhead for mobility. However, as the number of mobile nodes (MNs) handled by the MAP increases, the MAP suffers from the overhead not only to handle signaling traffic (e.g., binding update) but also to process data tunneling traffic. Therefore, it is important to control the number of MNs serviced by the MAP, in order to mitigate the burden of the MAP that results in higher latency. In this paper, we propose a mobility-based load control scheme, which consists of two sub-algorithms: (1) threshold-based admission control algorithm and (2) session-to-mobility ratio (SMR) based replacement algorithm. Here the SMR is defined as a ratio of the session arrival rate to the handoff rate. When the number of MNs at a MAP reaches to the full capacity, the MAP replaces an existing MN at the MAP, whose SMR is high, with an MN that just requests binding update. The replaced MN is redirected to its home agent. We analyze the proposed load control scheme using the Markov chain model in terms of the new MN blocking probability and the ongoing MN dropping probability. By combining the threshold-based admission control with the SMR-based replacement, the above probabilities are lowered significantly compared to the threshold-based admission control alone.**

I. Introduction

In IP-based mobile networks, a few mobility agents (e.g., home agent, foreign agent) are used to support mobility. Mobile IPv6 (MIPv6) [1] is a de facto mobility protocol in IPv6 wireless/mobile networks. To mitigate the high signaling overhead in Mobile IPv6 networks when MNs hand off frequently, Hierarchical Mobile IPv6 (HMIPv6) [2] was proposed. 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, so that the signaling traffic for mobility over the whole network will be reduced. However, the MAP can become a single point of bottleneck, leading to a degradation in performance, as it serves more and more mobile nodes (MNs). This is because the MAP not only handles binding update locally, but also performs encapsulation/decapsulation for every data packet destined for MNs. In other words, when a lot of MNs are serviced by a single MAP, the MAP suffers from traffic overload and this results in higher processing latency. In addition, as we expect a proliferation of mobile networking in the near future, how to control the traffic load on the MAP will be one of the most crucial issues. For the purpose of load control at the MAP, it is needed to limit the number of MNs serviced by the MAP.

A number of load control schemes at mobility agents in cellular networks or IP-based mobile networks have been proposed in the literature [3][4][5][6]. In our previous work [7], we applied Fang's admission control algorithm [10] to the MAP, which notably reduces the ongoing MN dropping probability compared to the basic HMIPv6 specification. However, the lower ongoing MN dropping probability is obtained at the cost of higher new MN blocking probability. In this paper, we propose a mobility-based load control scheme, which reduces the new MN blocking probability as well as the ongoing MN dropping probability as follows. The MAP monitors a session-to-mobility ratio (SMR) for each MN, where the SMR is the ratio of the session arrival rate to the handoff rate. If the SMR of an MN is high, it means that the MN is slowly moving and hence is unlikely to request frequent handoffs. In this case, it is better to process the binding update requests at its HA [8]. Based on this observation, if an MN requests a binding update to a MAP when there is no available capacity, the MAP chooses a currently serviced MN whose SMR is greater than δ, replaces the chosen MN with the MN newly requesting BU, and redirects the replaced MN to its HA.

The remainder of this paper is organized as follows. Section II provides an overview of HMIPv6. In Section III, we propose a mobility-based load control scheme for the MAP in HMIPv6 networks. In Section IV, we analyze the proposed load control scheme using the Markov chain model. Numerical results are presented in Section V. Section VI concludes this paper.

II. HMIPv6 Overview

In HMIPv6, 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 is attributed to an RCoA when it receives Router Advertisement (RA) messages 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).

Fig. 1 depicts the basic operations which are performed in HMIPv6. An MN entering a MAP domain will receive RA messages containing information on one or more local MAPs. Then, the MN selects the most suitable MAP by a number of criteria (distance, mobility, preference, etc.). However, the

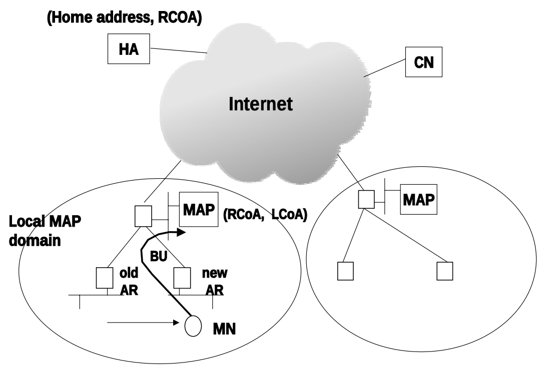

Fig. 1. Basic binding update procedure in HMIPv6

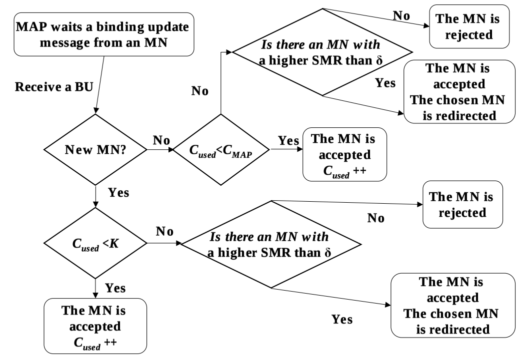

Fig. 2. Mobility-based load control procedure

question of how to select a MAP is the beyond the scope of this paper. We simply assume that a specific MAP selection scheme is used and that the MN sends a BU message to the selected MAP. The MN can bind its current location (i.e., LCoA) with an address on the MAP's subnet (i.e., RCoA). Acting as a local home agent (HA), the MAP will receive all packets on behalf of the MNs it is serving and will decapsulate and forward them to the MN's current address. If the MN changes its current address within a local MAP domain, it only needs to register the new address to the MAP. The RCoA does not change as long as the MN moves within the MAP domain. This makes the MN's mobility transparent to the correspondent nodes (CNs).

III. Mobility-based Load Control Scheme

A. System Model

In this paper, we propose a mobility-based load control scheme based on the following system models.

1 The capacity of a MAP (C_{MAP}) is represented by the maximum number of MNs that it can service.

2 An MN performs a binding update to only one MAP, which is selected by a MAP selection scheme.

3 If the binding update message is rejected by the MAP, the rejected MN can perform another binding update procedure to another MAP or the HA/CNs.

4 MNs are classified into three types: new MN, handoff MN, and refresh MN.

A *New MN*: New MN refers to an MN performing the initial binding update to the MAP (e.g. when an MN is turned on).

B *Handoff MN*: When an MN, which is serviced by a MAP, moves into another MAP domain, the MN sends a binding update to the new MAP. Since the handoff MN is registered in the previous MAP domain, the handoff MN should have a higher priority than the new MN [11].

C *Refresh MN*: In Mobile IPv6 [1], a lifetime is specified in the binding update message, which indicates the valid time of the binding information. After the specified lifetime, the binding information expires and the MN should refresh the binding information. The refresh MN is an MN performing a binding update for the purpose of refreshment.

In the mobility-based load control scheme, the handoff MN and refresh MN are classified into the same class: the ongoing MN class. In other words, the admission control algorithm considers two classes of MNs: *ongoing MNs* and *new MNs*. In addition, the ongoing MN always has a higher priority than the new MN.

B. Threshold-based Admission Control Algorithm

Fig. 2 shows the overall procedure of the mobility-based load control scheme. Here, C_{used} denotes the current number of MNs registered to the MAP. The mobility-based load control scheme consists of the threshold-based admission control algorithm and the SMR-based replacement algorithm. When a BU message arrives at a MAP, the MAP performs the threshold-based admission control algorithm as follows.

First of all, the MAP needs to determine whether the received BU message comes from an ongoing MN or a new MN. To make this possible, the **H flag** is added to the existing BU message [1]. The H flag is set if the MN is in the active state when it sends a BU message. To distinguish active state from idle state, each MN maintains an active state timer [9]. If an MN in idle state sends or receives data, the active timer is initialized and the MN enters into the active state. Every time the MN sends or receives data, the active timer is reset. If the MN does not send or receive data until the active timer expires, the MN returns to the idle state. If the H flag in the BU message is set, the MN is deemed to be an ongoing MN. Otherwise, the MN is regarded as a new MN.

Recall that the maximum number of MNs that can be serviced by a MAP is limited to C_{MAP}. Let K denote the threshold value. When the number of MNs serviced by the MAP is less than K, both a new MN and an ongoing MN will be accepted. Otherwise, to give higher priority to ongoing

MNs, only an ongoing MN will be accepted, which lowers the ongoing MN dropping probability [10].

C. SMR-based Replacement Algorithm

When an MN with Internet connectivity is hardly requesting handoffs, it is better for its home agent to process the MN's binding update request rather than the MAP [8]. This is mainly because if the MAP processes this BU request, there is additional packet encapsulation/decapsulation processing overhead at the MAP.

Based on this observation, we proposed the SMR-based replacement algorithm, in order to reduce the MN blocking/dropping probabilities. This algorithm is triggered in two cases. One is when a new MN requests binding update when there is no capacity for new MNs (i.e., $C_{used} \geq K$). The other is when an ongoing MN requests binding update when there is no capacity for ongoing MNs (i.e., $C_{used} = C_{MAP}$). If the SMR-based replacement algorithm is triggered, an MN with a higher SMR than δ is chosen and the chosen MN is replaced by the MN that newly requests binding update. In this way, the MN that will be rejected by the threshold-based admission control algorithm can be accepted.

Note that the replaced MN is not forced to be terminated. The replaced MN in fact, is redirected to its home agent. This redirection is realized as follows. Once the MN to be replaced is chosen, the MAP sends the BNACK message to the chosen MN. The BNACK message contains the status code 130 with the reason “Insufficient resources” [1]. Then, the MN will send binding update message to its HA. For the MN whose binding update request is accepted, the MAP will sends the BACK message.

IV. Performance Analysis

In this section, we evaluate the mobility-based load control scheme using Markov chain models in terms of a new MN blocking probability (P_{NB}) and an ongoing MN dropping probability (P_{OD}).

A. Assumptions

To develop the Markov chain models for the performance evaluation, we make the following assumptions.

- A.1 The arrival processes for the ongoing and new MNs follow Poisson distributions with rate of λ_O and λ_N, respectively.
- A.2 The residence times of the ongoing and new MNs follow exponential distributions with mean of $1/\mu_O$ and $1/\mu_N$, respectively.
- A.3 Let P_δ be the probability that the SMR of an MN is larger than δ, which is a threshold value showing for HMIPv6 lower cost. If the SMR of each MN is assumed to be ideally and identically distributed, $P(C, \delta)$, which is the probability that there are no MNs with a larger SMR than δ among C MNs, is as follows:

$$P(C,\delta) = \begin{pmatrix} C \\ 0 \end{pmatrix} \cdot P_\delta^0 \cdot (1-P_\delta)^C$$

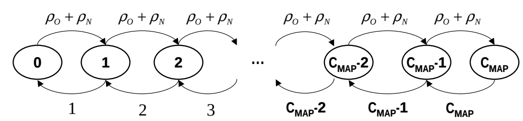

Fig. 3. Markov chain model in the case of the no load control scheme

B. Markov Chain Model

1) No Load Control Scheme: For comparison purposes, we compare the proposed scheme with the case without any load control. To find the steady state distribution of the no load control scheme, a two-dimensional Markov chain is required. However, in this paper, we use an approximated method introduced in [10] for the simplicity of analysis. In the approximated method, the service rate is normalized as unity. Therefore, in the approximated Markov chain model, the total offered load to the MAP is the sum of the ongoing MN offered load ($\rho_O = \frac{\lambda_O}{\mu_O}$) and the new MN offered load ($\rho_N = \frac{\lambda_N}{\mu_N}$), as follows:

$$\rho_T = \rho_O + \rho_N$$

Fig. 3 shows the approximated one-dimensional Markov chain model of the no load control scheme.

The transition rates are as follows:

$$\begin{aligned} q(i,i+1) = &\quad \rho_O + \rho_N \qquad (0 \leq i < C_{MAP}) \\ q(i+1,i) = &\quad i+1 \qquad\qquad (0 \leq i < C_{MAP}) \end{aligned} \tag{1}$$

where $q(i, i+1)$ and $q(i+1, i)$ are the transition rate from state i to state $i+1$ and the transition rate from state $i+1$ to state i, respectively. In this Markov chain, state i refers to the number of MNs registered. The steady state probability (p_k) can be obtained from Eq. (2) and the new MN blocking probability and the ongoing MN dropping probability are obtained from Eq. (3).

$$p_k = \frac{(\rho_O + \rho_N)^k / k!}{\sum_{n=0}^{C_{MAP}} (\rho_O + \rho_N)^n / n!} \qquad (0 \leq k \leq C_{MAP}) \tag{2}$$

$$P_{NB} = P_{OD} = p_{C_{MAP}} = \frac{(\rho_O + \rho_N)^{C_{MAP}} / C_{MAP}!}{\sum_{n=0}^{C_{MAP}} (\rho_O + \rho_N)^n / n!} \tag{3}$$

2) Mobility-based Load Control Scheme: In the mobility-based load control scheme, the total number of MNs in a MAP domain is used to make a decision as to whether a new MN is accepted or not. Let K be the threshold value in the threshold-based admission control algorithm. To find the dropping and blocking probabilities in the mobility-based

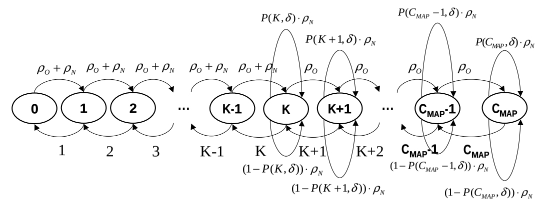

Fig. 4. Markov chain model in the case of the mobility-based load control scheme

load control scheme, the above-mentioned approximated one-dimensional Markov chain model is considered [10]. Fig. 4 depicts the approximated Markov chain model. State i represents the number of MNs in a MAP domain. $q(i,i')$ denotes the transition rate from state i to state i'.

$$\begin{aligned} q(i,i+1) &= \rho_O + \rho_N & (0 \le i < K) \\ q(i+1,i) &= i+1 & (0 \le i < K) \\ q(i,i+1) &= \rho_O & (K \le i < C_{MAP}) \\ q(i+1,i) &= i+1 & (K \le i < C_{MAP}) \end{aligned} \quad (4)$$

According to the Markov chain model depicted in Fig. 4, p_k can be obtained from Eq. (5). Then, the new MN blocking probability and the ongoing MN dropping probability in the mobility-based load control scheme are given by Eqs. (6) and (7), respectively. Compared to the threshold-based load control scheme [7] [1], the additional term, $P(k,\delta)$, is added when the current MAP load is equal to or larger than K (for new MNs) or equal to C_{MAP} (for ongoing MNs).

$$p_k = \frac{(\rho_O+\rho_N)^k/k!}{\sum_{n=0}^{K}(\rho_O+\rho_N)^n/n! + \sum_{n=K+1}^{C_{MAP}}(\rho_O)^{n-K}\cdot(\rho_O+\rho_N)^K/n!} \quad (0 \le k \le K)$$
$$p_k = \frac{(\rho_O)^{k-K}\cdot(\rho_O+\rho_N)^K/k!}{\sum_{n=0}^{K}(\rho_O+\rho_N)^n/n! + \sum_{n=K+1}^{C_{MAP}}(\rho_O)^{n-K}\cdot(\rho_O+\rho_N)^K/n!} \quad (K < k \le C_{MAP}) \quad (5)$$

$$P_{NB} = \sum_{k=K}^{C_{MAP}} P(k,\delta)\cdot p_k \quad (6)$$

$$P_{OD} = P(C_{MAP},\delta)\cdot p_{C_{MAP}} \quad (7)$$

V. Numerical Results

For numerical analysis, C_{MAP} is assumed to be 50. In terms of the threshold value (K), 40 and 45, are evaluated. δ is set to 1000. The arrival rates of new MN and ongoing MN are set to 1/20 and 1/20 (1/sec), respectively. The residence time is changed from 20s to 1180s, in order to vary the offered load.

[1]Hereafter, the threshold-based load control scheme refers to a load control scheme utilizing only threshold-based admission control algorithm.

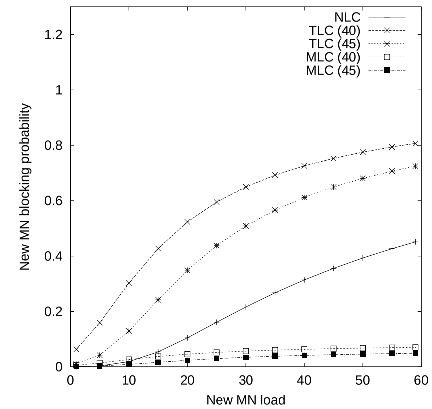

Fig. 5. New MN blocking probability: New MN traffic load

The inter-session arrival time and the MAP domain residence time are denoted by t_S and t_R, and are assumed to follow exponential distributions with mean values of $1/\lambda_S$ and $1/\mu_R$, respectively. Therefore, the SMR is equal to t_R/t_S. P_δ and $P(K,\delta)$ are calculated as follows:

$$P_\delta = \Pr(t_R > \delta\cdot t_S) = \int_0^\infty e^{-\mu_R\delta\tau}\cdot\lambda_S e^{-\lambda_S\tau}d\tau = \frac{\lambda_S}{\lambda_S+\delta\mu_R}$$

$$P(K,\delta) = \left(\frac{\delta\mu_R}{\lambda_S+\delta\mu_R}\right)^K$$

A. New MN Blocking Probability

Figs. 5 and 6 show the new MN blocking probability as the new MN load and the ongoing MN load increase, respectively. As shown in Fig. 5, the new MN blocking probability of the threshold-based load control (TLC) scheme alone is higher than that of the no load control (NLC) scheme. Also, as the threshold value decreases, the blocking probability increases. However, the difference in blocking probabilities between the two schemes does not significantly change as the new MN load (or ongoing MN load) increases. When the mobility-based load control (MLC) scheme is used, the MN blocking probability is significantly reduced as compared to the threshold-based load control scheme. Furthermore, the mobility-based load control scheme shows lower new MN blocking probability than no load control scheme. This is because the mobility-based load control scheme can accept more MNs even though there is no MAP capacity by using the SMR-based replacement algorithm.

B. Ongoing MN Dropping Probability

Figs. 7 and 8 show the ongoing MN dropping probability as the new MN load and the ongoing MN load increase, respectively. Since the threshold-based and mobility-based schemes both reserve a certain amount of MAP capacity for

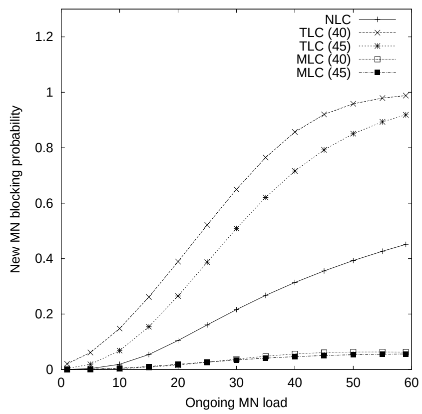

Fig. 6. New MN dropping probability: Ongoing MN traffic load

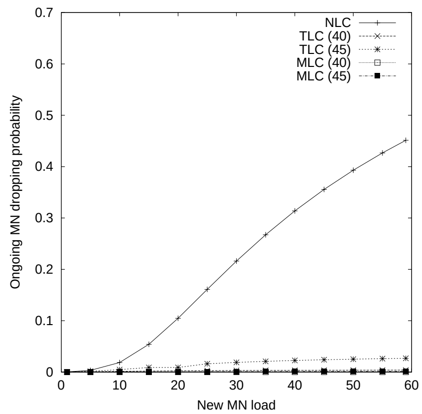

Fig. 7. Ongoing MN blocking probability: New MN traffic load

ongoing MNs, the ongoing MN dropping probabilities for two schemes are too low, regardless of new MN load, as shown in Fig. 7. However, if the threshold-based load control scheme alone is used, the ongoing MN dropping probability increases as the ongoing MN load increases (refer Fig. 8). However, it is possible to reduce the ongoing MN dropping probability by using mobility-based load control scheme with SMR-based replacement algorithm, even though the ongoing MN load is high.

VI. Conclusion

In this paper, we proposed a mobility-based load control scheme, which consists of the threshold-based admission control algorithm and the SMR-based replacement algorithm. For the purpose of load control at the MAP, the MAP first performs the threshold-based admission control algorithm, which limits the number of new MNs at the MAP to a specific threshold value (K) to give higher priority to ongoing MNs. In addition, the SMR-based replacement algorithm is applied when there is no remaining MAP capacity for new or ongoing MNs. In the replacement algorithm, an MN with a higher SMR than δ

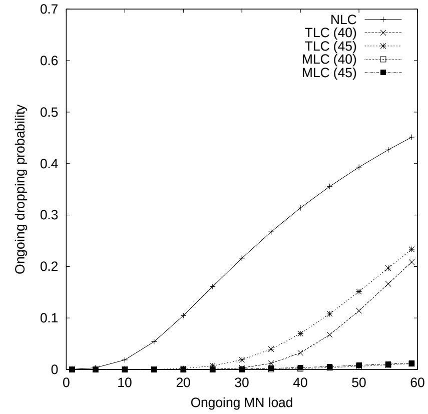

Fig. 8. Ongoing MN dropping probability: Ongoing MN traffic load

is chosen and replaced by newly registering MN. By combining the threshold-based admission control algorithm with the SMR-based replacement algorithm, the new MN blocking and ongoing MN dropping probabilities are lowered significantly compared to the threshold-based admission control alone. As a future work, we will validate the analytical results with simulations. Also, we will seek to analyze the amount of binding update traffic over the whole network.

VII. Acknowledgments

This work was supported in part by the BK 21 project of the Ministry of Education and in part by the NRL project of the Ministry of Science and Technology, 2004, Korea.

References

[1] D. Johnson, C. Perkins, and J. Arkko, “Mobility Support in IPv6,” Internet draft (work in progress), draft-ietf-mobileip-ipv6-24.txt, June 2003.
[2] H. Soliman, C. Castelluccia, K. E. Malki, and L. Bellier, “Hierarchical Mobile IPv6 mobility management (HMIPv6),” Internet draft (work in progress), draft-ietf-mipshop-hmipv6-00.txt, June 2003.
[3] Y. Lin, “Overflow Control for Cellular Mobility Database,” IEEE Trans. on Vehicular Technology, Vol. 49, No. 2, March 2000.
[4] H. Hung, Y. Lin, N. Peng, and S. Yang, “Resolving Mobile Database Overflow With Most Idle Replacement,” IEEE Journal on Selected Areas in Communications, Vol. 19, No.10, Oct. 2001.
[5] J. P. Jue and D. Ghosal, “Design and Analysis of A Replicated Server Architecture for Supporting IP-Host Mobility,” Cluster Computing, Vol. 1, No.2, 1998, pp. 249-260.
[6] A. Vasilache, J. Li, and H. Kameda, “Threshold-Based Load Balancing for Multiple Home Agents in Mobile IP Networks,” Telecommunication Systems, Vol. 22, 2003.
[7] S. Pack, B. Lee, and Y. Choi, “Load Control Scheme at Local Mobility Agent in Mobile IPv6 Networks,” in Proc. World Wireless Congress (WWC) 2004, May 2004.
[8] S. Pack and Y. Choi, “A Study on Performance of Hierarchical Mobile IPv6 in IP-based Cellular Networks,” IEICE Transactions on Communications, Vol. E87-B, No. 3, March 2004.
[9] X. Zhang, J. G. Castellanos, and A. T. Capbell, “P-MIP: Paging Extensions for Mobile IP,” ACM Mobile Networks and Applications, Vol. 7, No. 2, pp. 127-141, 2002.
[10] Y. Fang and Y. Zhang, “Call Admission Control Schemes and Performance Analysis in Wireless Mobile Networks,” IEEE Transactions on Vehicular Technology, Vol. 51, No. 2, March 2002.
[11] Y. B. Lin, S. Mohan, and A. Noerpel, “Queueing priority channel assignment strategies for handoff and initial access for a PCS network,” IEEE Trans. Veh. Technol., Vol. 43, No. 3, pp. 704-712, 1994.
[12] L. Kleinrock, “Queuing Systems Volume 1: Theory,” John Wiley & Sons, 1975.