An Optimal Location Management Scheme for Minimizing Signaling Cost in Mobile IP

Jiang Xie and Ian F. Akyildiz
Broadband and Wireless Networking Laboratory
School of Electrical and Computer Engineering
Georgia Institute of Technology, Atlanta, GA 30332

Abstract—**Mobile IP is a solution for mobility on the global Internet. However, it does not extend well to highly mobile users. Mobile IP regional registration is proposed to reduce the number of location updates to the home network, and reduce the signaling delay. This paper introduces an optimal regional location management mechanism for Mobile IP that results in the minimum signaling cost. A novel discrete analytical model is developed which captures the mobility and packet arrival pattern of a mobile terminal. This model does not impose any restrictions on the shape and the geographic location of Internet subnets. Given the average total location update and packet delivery cost, an iterative algorithm is then used to determine the optimal size of regional networks. Analytical results are also obtained to demonstrate how the optimal value changes under various parameters.**

I. INTRODUCTION

Mobile IP is a standard proposed by the Internet Engineering Task Force (IETF) [1] [2]. It is a solution for mobility on the global Internet. However, Mobile IP does not extend well to highly mobile users. When a mobile node (MN) moves from one subnet to another one, it must send a location update to its home agent (HA) even though the MN is not communicated with others. These location updates incur the latency of messages traveling to the possibly distant home network [3]. They also add significant signaling traffic if the number of MNs increases [4].

Mobile IP regional registration is proposed for MNs to register locally so that the number of signaling messages to the home network and the signaling delay are reduced [5]. The general model of operation is illustrated in Fig. 1. When an MN moves from one regional network to another one, it performs a home registration with its HA. During the home registration, the HA registers the publicly routable address of another mobility agent called gateway foreign agent (GFA). When an MN changes foreign agent (FA) within a regional network, it does not need to register with its HA. Instead, it performs a regional registration to the GFA to update its FA care-of address. During the communications, packets sent to an MN are addressed to the HA first. The HA encapsulates these packets and sends them to the registered GFA of the MN. The GFA checks its visitor list and forwards the packets to the corresponding FA in the visiting subnet of the MN. The FA further relays the packets to the MN.

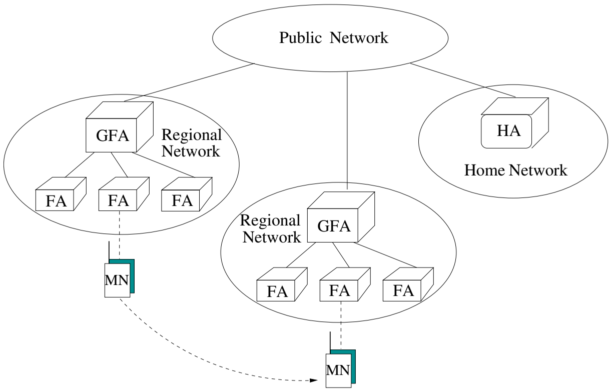

Fig. 1. A general operation model of regional registration.

However, it is not clear that how many FAs should be beneath a GFA within a regional network. The size of regional networks is very critical for the system performance. A small number of FAs within a regional network will lead to excessive location updates to the home network and consequently cannot provide the full benefit of regional registration. A large number of FAs will generate a high traffic load on GFAs, which results in a high cost of packet delivery [4]. To minimize the signaling traffic, it is desirable to find the optimal number of FAs beneath a GFA.

Previous researchers proposed some methods for calculating the optimal location area (LA) size in Personal Communication Systems (PCS) to reach the minimum costs for location update and terminal paging [7]. However, there are some differences between the analysis of location management schemes for Mobile IP and those in PCS. First, the cellular network is geographic-oriented. We may use two-dimensional geometric shape to represent the coverage area of each cell, such as rectangular and hexagonal cells [8]-[10]. But Internet is more spatial-oriented. We cannot use any geometric shape to accurately abstract a subnet. Second, in PCS, since each cell is indicated by a geometric shape, we may easily calculate the geographic distance between two cells [11]. On the other hand, the distance between two end points in Internet has nothing to do with the geographic location of these two points. Their distance is usually counted by the number of hops packets travel. Third, when an incoming call arrives, the cellular network locates the terminal by simultaneously paging all cells within an LA. However,

---

This work is supported by National Science Foundation (NSF) under Grant No. CCR-99-88532

0-7803-7400-2/02/$17.00 © 2002 IEEE

current Mobile IP does not support paging. But because of the triangular routing, packet delivery introduces extra processing and transmission costs. So there is packet delivery cost instead of paging cost.

In this paper, we introduce a new optimal location management scheme. We propose a mathematical model to capture the mobility and packet arrival patterns of each terminal. This model does not have any constraint on the shape and the geographic location of subnets. It is a general model which is applicable for all types of subnets. The distance unit in our model is the number of hops packets travel. Based on this model, we obtain the average location update and packet delivery costs. We use an iterative method to determine the optimal number of FAs under a GFA. We provide numerical results that demonstrate how the optimal value changes under various parameters.

This paper is organized as follows. In Section II, the mobility model is described and a method for deriving the total signaling cost is introduced. Then, in Section III, the iterative method for obtaining the optimal number of FAs beneath a GFA is provided. In Section IV, analytical results are presented, followed by the conclusions in Section V.

II. SIGNALING COST FUNCTION

In the following discussion, we suppose that the Mobile IP regional registration protocol [5] supports one level of foreign agent hierarchy beneath the GFA. The performance metric is the total signaling cost for location update and packet delivery. We do not take the periodic binding updates that an MN sends to its HA or FA to refresh their cache into account.

A. Location Update Cost

Similar to [6], we define the following parameters for location update:

C_{hg}	The transmission cost of location update between the HA and the GFA.
C_{gf}	The transmission cost of location update between the GFA and the FA.
C_{fm}	The transmission cost of location update over the wireless link between the FA and the MN.
a_h	The processing cost of location update at the HA.
a_g	The processing cost of location update at the GFA.
a_f	The processing cost of location update at the FA.

Fig. 2 and Fig. 3 illustrate the signaling message flows for location registration with the home network and regional registration within a regional network, respectively.

According to these message flows, the home registration cost and the regional registration cost for each location update can be calculated as:

$$C_{Uh} = 2a_f + 2a_g + a_h + 2C_{hg} + 2C_{gf} + 2C_{fm} \quad (1)$$

$$C_{Ur} = 2a_f + a_g + 2C_{gf} + 2C_{fm} \quad (2)$$

Let l_{hg} be the average distance between the HA and the GFA in terms of the number of hops packets travel, and l_{gf} be the average distance between the GFA and the FA. We assume the transmission cost is proportional to the distance between the source and the destination mobility agents and the proportionality constant is δ_U. Thus C_{hg} and C_{gf} can be expressed as $C_{hg} = l_{hg}\delta_U$ and $C_{gf} = l_{gf}\delta_U$. Since the transmission cost of the wireless link is generally higher than that of the wired link [6], we assume that the transmission cost over the wireless link is ρ times higher than the unit distance wireline transmission cost. The transmission cost between the FA and the MN can be written as $C_{fm} = \rho\delta_U$. Then the home registration and regional registration costs for each location update can be expressed as:

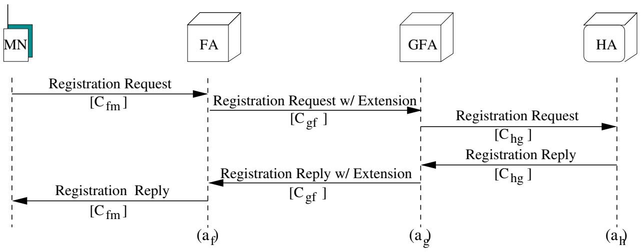

Fig. 2. Process of home location registration.

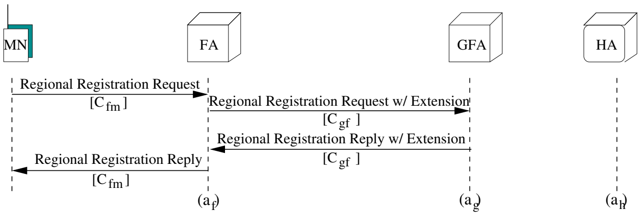

Fig. 3. Process of regional location registration.

$$C_{Uh} = 2a_f + 2a_g + a_h + 2(l_{hg} + l_{gf} + \rho)\delta_U \quad (3)$$

$$C_{Ur} = 2a_f + a_g + 2(l_{gf} + \rho)\delta_U \quad (4)$$

Assume each MN may move randomly between N subnets and there are k subnets within a regional network. We propose a discrete system to model the movements of each MN. In our model, MNs may visit a subnet more than once and it may also move back and forth between two subnets. We call the action each MN moving out of a subnet "*a movement*". Define a random variable M so that each MN moves out of a regional network at movement M. At movement 1, MNs may reside in either subnet $1,\ 2,\ \cdots$ or N. At movement 2, MNs may move to any of the N subnets. We assume MNs will move out to the other $N-1$ subnets with equal probability $\frac{1}{N-1}$. Fig. 4 shows an example of our discrete system in which $N = 5$ and $k = 3$. In the figure, each node represents a subnet and node 2, 3, and 4 belong to one regional network. The MN moves out of the regional network at movement 5.

The probability that each MN moves out of a regional network, i.e., the probability of performing a home registration at movement m is:

$$P_h^m = \frac{N-k}{N-1} \cdot \left(\frac{k-1}{N-1}\right)^{m-2}, \quad \text{where } 2 \le m < \infty \quad (5)$$

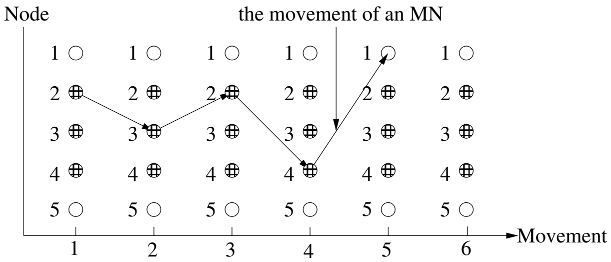

Fig. 4. Discrete system mobility model of an MN.

It can be shown that the expectation of M is:

$$E[M] = \sum_{m=2}^{\infty} mP_h^m = 1 + \frac{N-1}{N-k} \tag{6}$$

Assume within a regional network, the average time each MN stays in each subnet before making a movement is T_f. Therefore, the location update cost per unit time is:

$$C_{LU} = \frac{E[M]C_{Ur} + C_{Uh}}{E[M]T_f} \tag{7}$$

Note that our method does not impose any restrictions on the shape and the geographic location of subnets. It is a general model which is applicable to arbitrary subnets.

B. Packet Delivery Cost

Because of the triangular routing, there are extra costs for packet delivery under Mobile IP regional registration. The packet delivery cost includes the transmission and processing cost to route a tunneled packet from the HA to the registered GFA, and further forward to the serving FA of an MN. Assume

T_{hg}	The transmission cost of packet delivery between the HA and the GFA.
T_{gf}	The transmission cost of packet delivery between the GFA and the FA.
v_h	The processing cost of packet delivery at the HA.
v_g	The processing cost of packet delivery at the GFA.

The cost of each packet delivery procedure can be expressed as:

$$C_{PD} = v_h + v_g + T_{hg} + T_{gf} \tag{8}$$

We assume the transmission cost of delivering data packets is proportional to the distance between the sending and the receiving mobility agents with the proportionality constant δ_D. Then $T_{hg} = l_{hg}\delta_D$ and $T_{gf} = l_{gf}\delta_D$.

The load on a GFA for processing and routing packets to each FA depends on k, the number of FAs under a GFA. If k is large, the complexity of the visitor list lookup and IP routing table lookup in the GFA is high. In addition, since the total bandwidth of the network is limited, if the traffic to a GFA is heavy, the transmission delay and the number of retransmissions cannot be bounded. These factors will result in a high processing cost at the GFAs. Assume on average there are ω MNs in a subnet. Then, the total number of MNs a GFA serves in a regional network is ωk on average. Therefore, the complexity of the GFA visitor list lookup is proportional to ωk. Since IP routing table lookup is based on the *longest prefix matching* and most implementations use the traditional *Patricia trie* [12], the complexity of IP address lookup is proportional to the logarithm of the length of the routing table k [13]. We define the packet processing cost function at the GFA as:

$$v_g = \zeta k \cdot \lambda_a \left(\alpha\omega k + \beta log(k)\right) \tag{9}$$

where λ_a is the average packet arrival rate for each MN, α and β are weighting factors of visitor list and routing table lookups, and ζ is a constant which captures the bandwidth allocation cost at the GFA. The larger the ζ is, the more negative effects an MN experiences from not enough network bandwidth available.

The processing cost function at the HA can be defined as: $v_h = \eta\lambda_a$, where η is a packet delivery processing cost constant at the HA. Then the total packet delivery cost per unit time is:

$$C_{PD} = \eta\lambda_a + \zeta k \cdot \lambda_a \left(\alpha\omega k + \beta log(k)\right) + (l_{hg} + l_{gf})\delta_D \tag{10}$$

C. Total Signaling Cost

Based on the above analysis, we may get the overall average signaling cost function from (7) (10):

$$C_{TOT}(k, \lambda_a, T_f) = C_{LU} + C_{PD} \tag{11}$$

III. OPTIMAL VALUE

The optimal number of FAs beneath a GFA, k_{opt}, is defined as the value of k that minimizes the cost function derived in Section II. Because k can only be an integer, the cost function is not a continuous function of k. Therefore, it is not appropriate to take derivatives with respect to k of the cost function to get the minimum. We use an iterative algorithm. Note that iterative algorithm may result in a local minimum. Solutions to solving the local minimum problem were discussed in [11]. Similar to the algorithm proposed in [7], we define the cost difference function between the system with number k and the system with number $k - 1 (k \geq 2)$, i.e.,

$$\Delta(k, \lambda_a, T_f) = C_{TOT}(k, \lambda_a, T_f) - C_{TOT}(k-1, \lambda_a, T_f) \tag{12}$$

Given Δ, the algorithm to find the optimal value of k is defined as follows:

$$k_{opt}(\lambda_a, T_f) = \begin{cases} 1, \text{ if } \Delta(2, \lambda_a, T_f) > 0 \\ max\{k : \Delta(k, \lambda_a, T_f) \leq 0\}, \text{ otherwise} \end{cases} \tag{13}$$

The optimal regional network size k_{opt} is a designed value. It is computed before the communications based on the average packet arrival rate λ_a and average subnet residence time T_f over all users. Our algorithm also needs to know the number of hops between the HA and the GFA, l_{hg}, and the number of hops between the GFA and the FA, l_{gf}. If each MN has dedicated paths for transmitting signaling messages from FAs to GFAs and HAs, then l_{hg} and l_{gf} are fixed numbers. If not, signaling packets may take different paths each time according to the traffic load and routing algorithms at each mobility agent. Thus, l_{hg} and l_{gf} vary within a certain range.

IV. ANALYTICAL RESULTS

In this section, we demonstrate some numerical results. Table I lists some of the parameters used in our performance analysis. Since the total number of subnets that MNs may access through wireless channels is limited, we assume $N = 30$. For our numerical evaluation, we assume l_{hg} and l_{gf} are fixed numbers. We set $l_{hg} = 25$ and $l_{gf} = 10$.

A. The Impact of Residence Time on the Total Signaling Cost

First, we show how the total signaling cost changes with the size of the regional networks under different average residence time, when the average packet arrival rate is fixed.

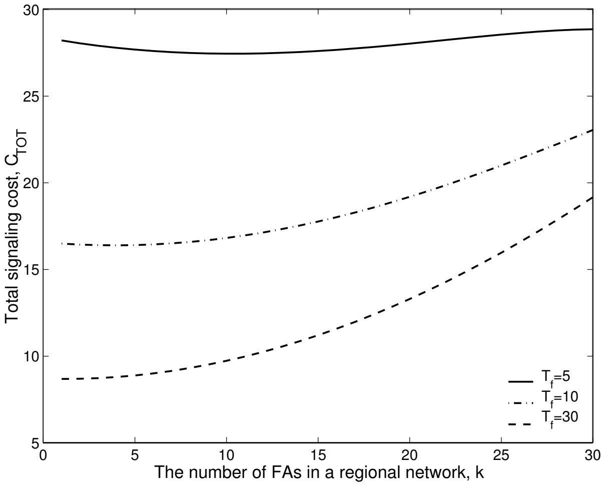

Fig. 5. Total signaling cost versus the size of regional networks.

Fig. 5 plots the total signaling cost as a function of k, when average packet arrival rate $\lambda_a = 0.3$. As shown in the figure, the total signaling cost increases as the average residence time T_f decreases. The total signaling cost is the highest when $T_f = 5$. This is because when T_f is small, the mobility rate is high. MNs are more "active" roaming between subnets. For the same regional network size, high mobility rate leads to frequent location updates and thus high signaling cost. It is also observed that when T_f is small, the minimum signaling cost can be reached when the regional network size is large. As T_f increases, the system requires a smaller regional network size to achieve the optimal performance. This result is intuitive. When mobility rate is high, the cost for location update dominates. Systems with larger regional networks may reduce the number of home registrations and provide the benefit of regional registration.

B. The Impact of Packet Arrival on the Total Signaling Cost

Next, we see how the total signaling cost changes with the size of the regional networks under different average packet arrival rate, when the average residence time is fixed.

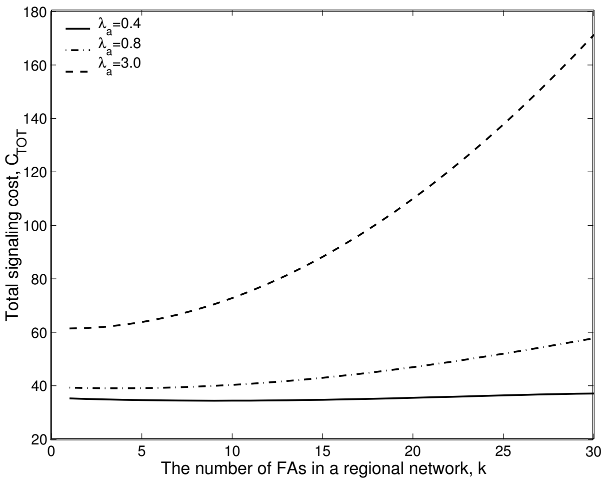

Fig. 6. Total signaling cost versus the size of regional networks.

Fig. 6 shows the total signaling cost as a function of k, when average residence time $T_f = 4$. Fig. 6 indicates that the total signaling cost increases as the average packet arrival rate λ_a increases. The total signaling cost is the highest when $\lambda_a = 3.0$. This is because when λ_a is large, the traffic load to GFAs and HAs is high, which results in a high signaling cost. Note that when λ_a is large, the optimal performance can be reached when the system has a small size of regional networks. As λ_a decreases, the minimum signaling cost can be achieved with larger regional networks. When the traffic to mobility agents is heavy, the packet delivery cost is high. There is a cost advantage if the number of FAs in a regional network is small. Similarly, when the packet arrival rate is small, the location update cost is relatively high. There is a cost advantage if location update to the home network is performed less frequently.

C. The Impact of Residence Time on the Optimal Value

Then, we investigate how the optimal regional network size varies with the average residence time.

Fig. 7 shows the optimal value k_{opt} as a function of the average residence time T_f, when the average packet arrival rate $\lambda_a = 0.1, 1$, and 10. Note that similar results to the Fig. 6 are observed from Fig. 7, i.e., when λ_a is large, under the same average residence time T_f, the optimal regional network size k_{opt} is small. From Fig. 7 we may also see that for all the three curves, the optimal value k_{opt} decreases as T_f increases. This conclusion is the same as what we observed from Fig. 5.

D. The Impact of Packet Arrival Rate on the Optimal Value

Finally, we study the impact of packet arrival rate on the optimal value.

TABLE I

PERFORMANCE ANALYSIS PARAMETERS

Pkt Process Cost			Distance Cost Unit		Wireless Multiple	# of MNs/subnet	Weight		Pkt Delivery Const.	
a_h	a_g	a_f	δ_U	δ_D	ρ	ω	α	β	ζ	η
30.0	20.0	15.0	0.2	0.05	10	15	0.3	0.7	0.01	10.0

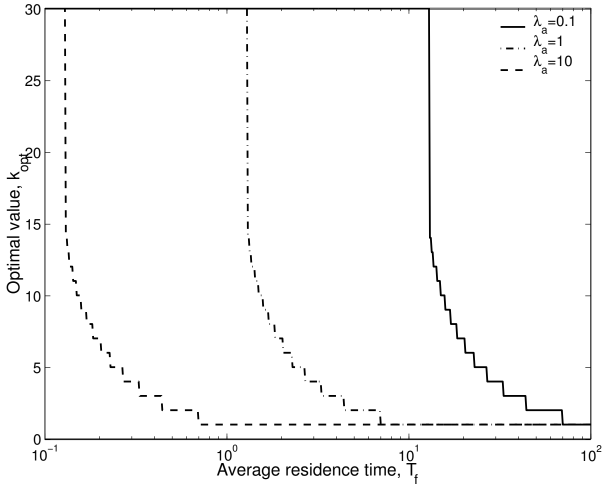

Fig. 7. Optimal regional network size versus average residence time.

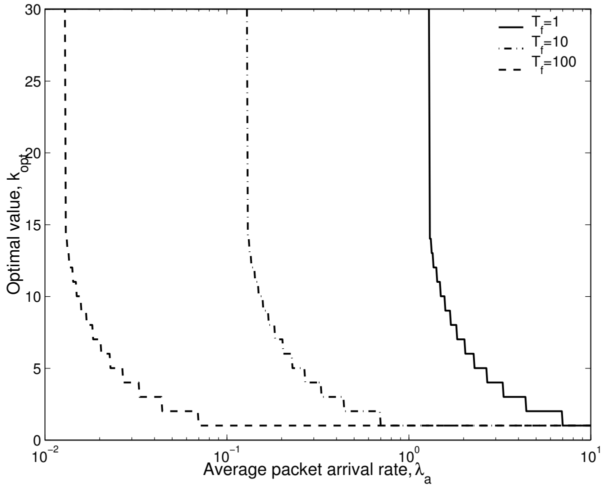

Fig. 8. Optimal regional network size versus average packet arrival rate.

Fig. 8 plots the optimal value k_{opt} as a function of the average packet arrival rate λ_a, when the average residence time $T_f = 1, 10$, and 100. Similar results can be also observed as in Fig. 5, i.e., when T_f is small, under the same average packet arrival rate λ_a, the optimal regional network size is large. Also note that the optimal value k_{opt} decreases as λ_a increases. This is because when the packet delivery cost dominates, the saving in packet delivery becomes significant. The saving can be attributed to the smaller regional network size.

V. CONCLUSION

In this paper, we introduced an optimal regional location management mechanism for Mobile IP which results in the minimum consumption of network resource. The mobility of MNs is modeled by a novel discrete analytical model. Based on this model, we obtained the average signaling cost function of location update and packet delivery. Given this average total cost function, we determined the optimal number of FAs in a regional network by using an iterative algorithm. Our analytical model does not have any constraint on the shape and the geographic location of Internet subnets. Analytical results demonstrated that the optimal regional network size decreases as the user mobility rate decreases, or the traffic load increases.

REFERENCES

[1] C. E. Perkins, "IP mobility support," Request for Comments (RFC) 2002-2006.

[2] C. E. Perkins, "Mobile IP", *IEEE Communication Magazine*, pp. 84-99, May 1997.

[3] R. Caceres and V. N. Padmanabhan, "Fast and scalable handoffs for wireless internetworks," in *Proc. ACM Mobicom'96*, pp.56-66, 1996

[4] C. Castelluccia, "Extending Mobile IP with adaptive individual paging: a performance analysis", in *Proc. IEEE Symposium on Computer and Communications*, pp. 113-118, 2000.

[5] E. Gustafsson, A. Jonsson, and C. E. Perkins, "Mobile IP regional registration (work in progress)," Internet Draft, Internet Engineering Task Force, draft-ietf-mobileip-reg-tunnel-05.txt, September 2001.

[6] Y. Wang, W. Chen, and J. S. Ho, "Performance analysis of Mobile IP extended with routing agents," *Technical Report 97-CSE-13*, Southern Methodist University, 1997.

[7] H. Xie, S. Tabbane, and D. J. Goodman, "Dynamic location area management and performance analysis," in *Proc. 43rd IEEE Vehicular Technology Conference*, pp. 536-539, 1993.

[8] W. Wang and I. F. Akyildiz, "Intersystem location update and paging schemes for multitier wireless networks," in *Proc. ACM MobiCom'2000*, pp.99-109, 2000.

[9] Y-B Lin, W-R Lai, and R-J Chen, "Performance analysis for dual band PCS networks," *IEEE Transactions on Computers*, vol. 49, no. 2, pp. 148-159, February 2000.

[10] I. F. Akyildiz, Y-B Lin, W-R Lai, and R-J Chen, "A new random walk model for PCS networks," *IEEE Journal on Selected Areas in Communications (JSAC): Wireless Series*, vol. 18, no. 7, pp. 1254-1261, July 2000.

[11] J. S. Ho and I. F. Akyildiz, "Mobile user location update and paging under delay constraints," *ACM-Baltzer Journal of Wireless Networks (WINET)*, vol. 1, no. 4, pp.413-425, December 1995.

[12] B. Lampson, V. Srinivasan, and G. Varghese, "IP lookups using multiway and multicolumn search," *IEEE/ACM Transactions on Networking*, vol. 7, no. 3, pp.324-334, June 1999.

[13] H-Y Tzeng, and T. Przygienda, "On fast address-lookup algorithms," *IEEE Journal on Selected Areas in Communications (JSAC)*, vol. 17, no. 6, pp.1067-1082, June 1999.