산업기술연구(강원대학교 산업기술연구소 논문집), 제25권 B호, 2005.
Journal of Industrial Technology, Kangwon Natl. Univ., Korea, No. 25 B, 2005.

# 최소 평균 페이징 지연을 위한 최적의 페이징 그룹 수에 관한 연구

# A Study of Optimal Group Number to Minimize Average Paging Delay

이 구 연*
Lee, Goo-Yeon

## Abstract

We present a numerical analysis of the optimal group number for minimizing the average paging delay. In the analysis, we consider uniform distributions for location probability conditions and apply M/D/1 queueing model to paging message queues of cells. We also get the lower bounds of group numbers and investigate the minimum transmission capacity under average paging delay constraints. Minimizing the average paging delay is important because it means minimizing the amount of bandwidth used for locating mobile terminals. Therefore, the numerical results of this paper will be very useful in PCS system when designing its signalling capacity due to its simplicity and effectiveness.

Keywords : *PCS, Paging, optimum cell group number, paging delay*

## 1. Introduction

Mobile systems have tradeoffs between registration and paging costs because of the location tracking property. Location registration and paging process cause increasing signalling costs as the demand of wireless services and number of mobile users grow rapidly.

Location registration is the process of reporting the current cell locations. When a mobile terminal enters a new location area consisting of a number of cells, it registers with the system. Therefore, the system is always aware of the current location area of a mobile terminal. However, when a call arrives, the system still needs to find out the cell in which the called terminal resides in the location area.

On the other hand, paging is the process in which a system searches for a mobile terminal by sending paging messages to cells in the last registered location area. There are several schemes about paging. In a centralized paging scheme, a called terminal number is broadcast to all cells. The scheme is efficient when the paging load is low compared to the system's signalling capacity. However, when the load is not low, the scheme is inefficient in the use of radio bandwidth for signalling. In other techniques, the basic unit of paging is a cell or a group of cells, where the problem is to design an efficient search algorithm such that the

* 강원대학교 전기전자정보통신공학부 교수, 공학박사

relevant costs are minimized.

Paging cost is usually measured in terms of cells to be paged before the called terminal is found [1],[2],[3]. Since the number of paging messages are concerned with the number of cells to be paged, the paging costs also include radio channel occupancy which affects delays in finding mobile terminals. It is necessary to minimize paging delay in order to efficiently manage mobile networks. An effect of minimizing the amount of paging messages used for locating mobile terminals is that the remaining bandwidth could be used for other purposes including user data transmission.

Many paging schemes have been proposed to improve the efficiency of bandwidth utilization, which reduced the paging costs based on location probabilities calculated using different methods [4], [5], [6], [7]. However, managing location probabilities of individual users is not a good method in that investigating each user's mobility characteristics and managing moving patterns individually are not practical in mobile communication systems.

Multi-step paging schemes were suggested to satisfy the delay bounds while reducing the paging costs [8], [9], [10], [11], [12], [13], [14]. In each step, a group of cells called paging area (PA) is searched in one polling cycle which is the time elapsed from when a paging message is sent to a PA until the response is received. In the schemes, the paging delays are represented by the required polling cycles which is the number of PAs to be searched before the called terminal is found. A polling cycle consists of two factors. One is the queueing time in paging message queues of cells and the other is the fixed waiting period which is determined based on the maximum round trip time of cells. If no response arrives in the fixed waiting period after sending a paging message in a cell, then the system concludes that the called terminal is not in the cell. Considering polling cycles to be paging delays is applicable when the fixed waiting period overwhelms the queueing time. Therefore, if the fixed waiting period is not large and the waiting time in the paging message queues is not negligible, we should count on the queueing delay.

In this paper, we present a simple multi-step paging scheme based on sequential searching PAs and analyze the optimal group number minimizing the average paging delay according to paging load. We consider uniform distributions for location probability conditions. We apply M/D/1 queueing model to the paging message queues, which reflects the practical situations. In the search algorithm, the cells are divided into equal-sized groups in advance for the queue-length synchronization of cells in a group. The results of the analysis are given in closed forms.

The rest of this paper is organized as follows. In Section 2, we present our model and definitions. In Section 3, we analyze the model and present discussions of the results. We conclude in Section 4.

## 2. Model and Definitions

In the search procedure considered in this paper, the system determines the optimal group number according to paging load and divides cells into equal-sized cell groups. And the system operates as the following steps.

Step 1: When a call arrives, go to step 2.
Step 2: Paging messages for the call are inserted into paging message queues of cells in a randomly selected group which is not yet paged for the call.
Step 3: When the paging message gets to the head of the queue, it is transmitted. Then the system waits for the response. Note that while waiting, the system goes on transmitting other paging messages in the queue.
Step 4: If the response from the called terminal arrives within the fixed waiting period, $T_w$, then the paging for the call completes. Otherwise, go to step 2. This procedure continues until all groups are searched.

We let $N_G$ be the cell group number. We denote the number of calls arriving at the system a unit time by $C$ which becomes the paging load to the system. We also let $H$ be the maximum number of paging messages which a cell can transmit a unit time, that is, the transmission capacity for paging messages of a cell. We assume that the transmission capacities of all cells are the same. We let $t_r$ be the random variable representing the time elapsed between sending a paging message and receiving the corresponding response in a cell and let $T_r$ be the mean value of $t_r$.

If the paging load, $C$, is much smaller than the transmission capacity of each cell, $H$, flooding finds the called terminal quickly with more messages than polling. Polling finds the terminal using less paging messages because cells are searched one by one until finding the terminal. Furthermore, if $C$ is larger than $H$, flooding makes the wireless system diverge and break down. In the case, cell grouping is necessary for the system to work. We have a simple criterion about paging method.

If $C$ is much smaller than $H$, flooding( $N_G$ = 1 ) is used.
Else if $C < 2 \cdot H$, cells should be partitioned into groups and paging messages are sent to groups in sequence.
Else, the system is overloaded and it should be reconfigured.

The paging message queue of a cell is shown in Fig. 1. Since it is common to assume that call interval to a mobile terminal is exponentially distributed and if there are a number of mobile terminals in the system, the input arrival pattern to the paging message queue of a cell could be considered Poisson distributed. A cell can transmit $H$ paging messages a unit time. Assuming that the messages have constant length, the service time of the queue is deterministic. Then the queue can be considered to be an M/D/1 system with arrival rate $\lambda_c$ and service time $\frac{1}{H}$.

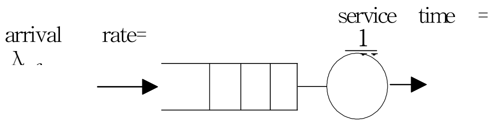


Fig. 1. M/D/1 queueing model of the paging message queue of a cell

## 3. Analysis and Results

We let $N$ be the number of cells in the system. We assume that there are no movements of mobile terminals while paging durations for the simplicity of analysis. We will get the average number of total paging messages transmitted in the system a unit time, which is denoted by $K$. $K$ has the same value as the average number of total paging messages generated for $C$ calls. For a call, the probability that the called terminal is found at the n'th ($n \le N_G$) group is $\frac{1}{N_G}$ since the group is selected at random. In the case, the number of paging messages transmitted to find out the location of the terminal is $n \cdot \frac{N}{N_G}$. Since there are $C$ call arrivals a unit time, $K$ is given by

$$K = C \sum_{n=1}^{N_G} \left[ \frac{1}{N_G} \cdot n \cdot \frac{N}{N_G} \right] = \frac{CN}{N_G^2} \cdot \frac{N_G(N_G+1)}{2} = \frac{CN(N_G+1)}{2N_G}. \quad (1)$$

Because mobile terminals are assumed to be uniformly distributed among all cells in the system and cell groups to page are selected at random, the average number of paging messages transmitted in a cell a unit time is $\frac{K}{N}$ which becomes the input arrival rate to the paging message queue of a cell. Therefore we have $\lambda_c = \frac{K}{N}$. From M/D/1 queueing theory, the average system time, $T_c$, which is the sum of queueing time and service time of a queue is given by

$$T_c = \frac{2-\rho}{2(1-\rho)} \overline{x} \quad (2)$$

where $\rho$ is the utilization factor and $\overline{x}$ is the mean service time.

From the definitions above, we have for the paging message queue

$$\rho = \frac{K}{HN} \text{ and } \bar{x} = \frac{1}{H}. \quad (3)$$

Then we have

$$T_c = \frac{2 - \frac{K}{HN}}{2(1 - \frac{K}{HN})} \cdot \frac{1}{H} = \frac{2HN - K}{2H^2N - 2HK}. \quad (4)$$

Inserting (1) into (4), we get

$$T_c = \frac{N_G(4H - C) - C}{N_G(4H^2 - 2HC) - 2HC}. \quad (5)$$

In (5), we note that $T_c$ is not a function of $N$.

For $T_c$ not to diverge, $\rho = \frac{K}{HN}$ should be less than 1. Inserting (1) into $\rho$ of (3), we have

$$\rho = \frac{C(N_G + 1)}{2N_G H} < 1 \quad (6)$$

Since $2H - C > 0$, we get from (6)

$$N_G > \frac{C}{2H - C} = \frac{\frac{C}{H}}{2 - \frac{C}{H}} \quad (7)$$

A. Average Paging Delay

We consider the average paging delay a call suffers from. The probability that a called terminal is found at the n'th($n \le N_G$) group is $\frac{1}{N_G}$. In the case, the average paging delay is $nT_c + (n-1)T_w + T_r$. That is because the call suffers from $n$ queuing delays, $(n-1)$ timeouts due to no responses and one paging response time. Therefore, the average paging delay a call suffers from, $D$, is given by

$$D = \frac{1}{N_G} \sum_{n=1}^{N_G} [nT_C + (n-1)T_w + T_r] = \frac{N_G + 1}{2} T_C + \frac{N_G - 1}{2} T_w + T_r \quad (8)$$

Inserting (5) into (8) results in

$$D = \frac{N_G + 1}{2} \cdot \frac{N_G(4H - C) - C}{N_G(4H^2 - 2HC) - 2HC} + \frac{N_G - 1}{2} T_w + T_r \quad (9)$$

From (9), some graphs are presented to investigate the behavior of $D$ by varying $N_G$. They are shown in Figs. 2 and 3.

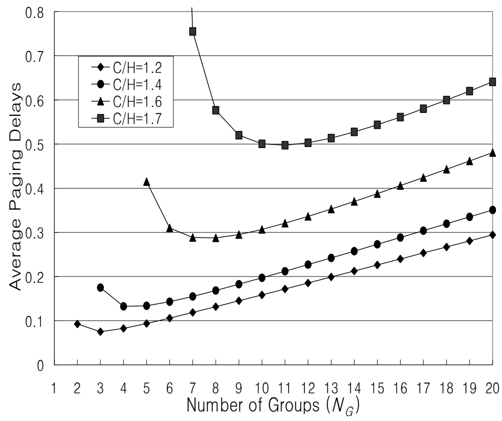


Fig. 2. Average paging delay when $T_w$=0.01, $H$=100, $T_r$=0.5 $T_w$ and $C/H$=1.2, 1.4, 1.6 and 1.7.

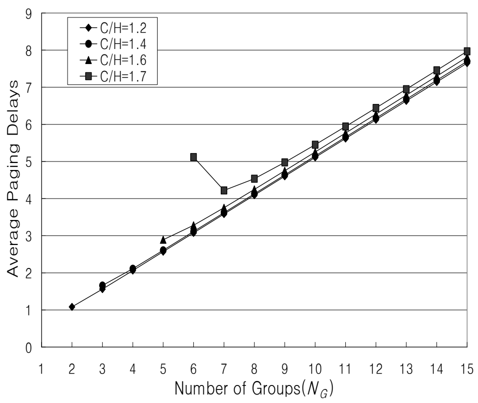


Fig. 3. Average paging delay when $T_w$=1, $H$=100, $T_r$=0.5 $T_w$ and $C/H$=1.2, 1.4, 1.6 and 1.7.

In the figures, $T_r$ is assumed to have a half value of $T_w$. The assumption is reasonable because $T_w$ is determined based on the maximum round-trip time from the base stations of cells to the boundaries of the cells and called terminals which locate uniformly among the cells respond to paging messages immediately. In Fig. 2, we see that the average paging delays increase as $C/H$ increases and the optimal group numbers also increase. It is because larger $C/H$ means higher call arrival rate compared to the transmission capacity and the system needs to divide cells into more groups to avoid system divergence.

Fig. 2 shows that when $T_w$ has a small value of 0.01, the queueing delay in the paging message queue is not negligible and rather dominant in the average paging delay. In the case, there exist minimum values of the average paging delays and the analysis for the optimal group number is applicable. However, in Fig. 3, we see that when $T_w$ has a larger value of 1, there is a tendency that $T_w$ is dominant in the average paging delay. In the case, the average paging delays almost depend on the number of polling cycles, where considering polling cycles to be paging delays is effective. The optimal group numbers of the case are the smallest group numbers for the system not to diverge. For example, the optimal group numbers with $C/H$=1.2, 1.4 and 1.6 in Fig. 3 are 2, 3 and 5, respectively. We also see in Fig. 3 that when $C/H$ is 1.7, the queueing delay becomes dominant again and we have the minimum value of the average paging delays. Generally, when $C/H$ approaches the value of 2, the analysis is applicable regardless of the value of $T_w$.

## B. Lower Bounds of Group Numbers

From (7), we can get the lower bounds of group numbers as varying $C/H$. The lower bounds are the smallest group numbers for the system not to diverge. Fig. 4 shows the lower bounds. From the figure, we see that the lower bounds of group numbers grow rapidly when $C/H$ approaches 2. Therefore, for the manageable group numbers, $C/H$ should be maintained under some moderate values. If the call arrival rate increases greatly than planned and the value of $C/H$ approaches 2, then the system should increase the transmission capacity so that $C/H$ has a moderate value.

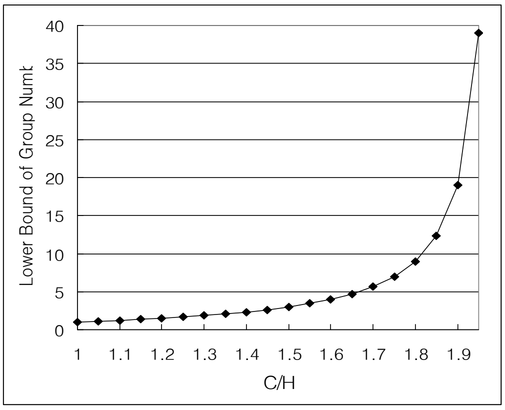


Fig. 4. Lower bounds of group numbers

## C. Optimal Group Number for Minimizing Average Paging Delay

(9) can be transformed into the following equation through tedious calculation.

$$D=\left(\sqrt{A_1(N_G-r)}-\sqrt{\frac{A_2}{N_G-r}}\right)^2+D_{\min} \tag{10}$$

where

$$r = \frac{C}{2H-C} \quad A_1 = \frac{4H-C}{4H(2H-C)} + \frac{T_w}{2},$$
$$A_2 = \frac{HC}{(2H-C)^3} \tag{11}$$

and

$$D_{\min} = \frac{\sqrt{C(4H-C)+2CT_wH(2H-C)}}{(2H-C)^2}$$

$$+ \frac{2H+T_w(3HC-C^2-2H^2)}{(2H-C)^2} + T_r \tag{12}$$

Note that $N_G - r \geq 0$ from (7) and $A_1, A_2$ are positive since $2H-C \geq 0$. Therefore $D$ has the minimum value $D_{\min}$ when

$$\sqrt{A_1(N_G - r)} = \sqrt{\frac{A_2}{N_G - r}} \tag{13}$$

which results in

$$N_{G_{\min}} = N_G|_{when\, D=D_{\min}} =$$
$$= \frac{C}{2H-C} + \frac{2H}{2H-C}\sqrt{\frac{C}{4H-C+4T_wH^2-2T_wHC}} \tag{14}$$

$N_{G_{\min}}$ is the theoretical optimal group number for minimizing the average paging delay. We use the term 'theoretical' since $N_{G_{\min}}$ is the number which makes $D$ minimal but it may not be an integer. Figs. 5 and 6 show the theoretical optimal group numbers and their minimum delays when $T_w$=0, 0.01, 0.1 and 1 as $C/H$ varies.

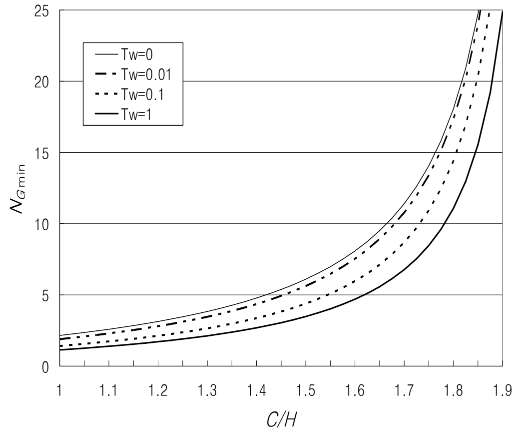


Fig. 5. $N_{G_{\min}}$ when $H$=100, $T_r$=0.5 $T_w$ and $T_w$=0, 0.01, 0.1 and 1.

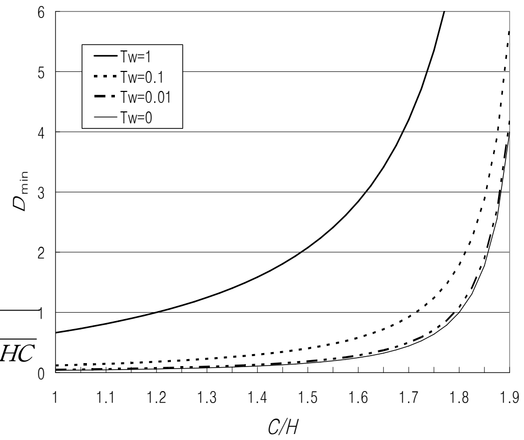


Fig. 6. $D_{\min}$ when $H$=100, $T_r$=0.5 $T_w$ and $T_w$=0, 0.01, 0.1 and 1.

In Figs. 5 and 6, the case of $T_w$=0 is inserted to see the limit values as $T_w$ goes to zero. However the case of $T_w$=0 is not a real situation because there needs some time to receive the paging response. In Fig. 5, we see that as $T_w$ gets large, $N_{G_{\min}}$ gets smaller. It is because as $T_w$ goes large, the contribution of queueing delay to paging delay decreases. Therefore, $N_{G_{\min}}$ approaches the values of Fig.

4 as $T_w$ goes infinite. As we expect, in Fig. 6, we see that $D_{\min}$ increases as $T_w$ increases. The case of $T_w$=0 in Fig. 6 is the pure queueing delay in the paging message queues. When determining $T_w$, we should select an appropriate value so that the corresponding value of $D_{\min}$ is not too distant from $D_{\min}$ of $T_w$=0.

### D. Minimum Transmission Capacity Under Average Paging Delay Constraints

We investigate the minimum transmission capacity under the average paging delay constraint which is denoted by $D_a$. That is, we investigate the minimum $H$, $H_{\min}$, and the corresponding $N_G$ when $C$ and $D_a$ are given.

From (9), $H_{\min}$ can be obtained from the following inequality

$$D_a \geq \frac{N_G+1}{2} \cdot \frac{N_G(4H-C)-C}{N_G(4H^2-2HC)-2HC} + \frac{N_G-1}{2} T_w + T_r.$$
(15)

(15) is a second-order inequality about $H$. We let $H_m(N_G)$ be the minimum $H$ which satisfies (15) when $N_G$ is given. Applying the conditions $H > \frac{C}{2}$ and $N_G(4H^2-2HC)-2HC > 0$ to (15), $H_m(N_G)$ can be obtained. $H_m(N_G)$ is a function of $N_G$. Fig. 7 shows $H_m(N_G)$ as $N_G$ varies. From the figure, we can get the minimum value of $H_m(N_G)$ which becomes $H_{\min}$ and the corresponding value of $N_G$ .

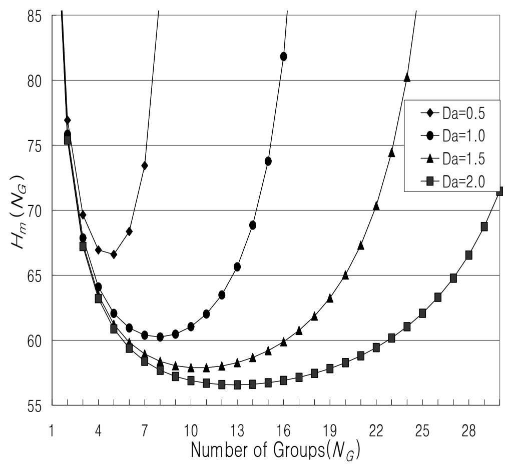


Fig. 7. $H_m(N_G)$ when $C$=100, $T_w$=0.01, $T_r$=0.5 $T_w$ and $D_a$=0.5, 1.0, 1.5 and 2.0.

From the figure, we see that when $D_a$ goes large, $H_{\min}$ goes small, as we expect. However, $H_{\min}$ can not get across $\frac{C}{2}$ and approaches it as $D_a$ goes infinite. Therefore, to minimize the transmission capacity, we do not have to set $D_a$ to an unnecessarily big value. That is, when determining $D_a$, we should select an appropriate value so that the corresponding $H_{\min}$ is not too distant from $\frac{C}{2}$. From the results, we are able to design the system efficiently under average paging delay constraints by assigning the minimum bandwidth to the signalling channel for paging messages. Thus we can use the saved bandwidth for other purposes.

The numerical results in this paper are independent on $N$. There can be two questions about the results. One is the possibility that $N_{G_{\min}}$ is not an integer. In the case, one of the two integers near $N_{G_{\min}}$ will be the optimal group number, which can be easily decided by inserting the two integers into (9) and comparing the calculated values of $D$. Of course, if $N_{G_{\min}}$ is an integer, it is the optimal group number. The other is the possibility that

$N$ is not a multiple number of the optimal group number. If $N$ is a multiple number of the optimal group number, cells in the system are divided into the optimal number of equal-sized groups. However, in the case that $N$ is not a multiple number of the optimal group number, the situation can be solved by dividing cells into the optimal number of groups of similar sizes with difference of maximum one. It is acceptable because if $N$ does not have too small values, the minimum average paging delay will be near the theoretical value.

## 5. Conclusion

In this paper, we described a search procedure and presented a simple criterion about paging method. In the criterion, the call arrival rate should be less than two times the transmission capacity of cells. If the call arrival rate is not too small, cells should be partitioned into groups. We present a numerical analysis of the optimal group number for minimizing the average paging delay according to paging load. In the analysis, we consider uniform distributions for location probability conditions and apply M/D/1 queueing model to the paging message queues of cells. From the results of the analysis, we see that the optimal group numbers increase as $C/H$ increases. We also get the lower bounds of group numbers for the system not to diverge. And we investigate the minimum transmission capacity under average paging delay constraints. From the results, we can see that we do not have to set the average paging delay constraints to an unnecessarily big value, but should select an appropriate value so that the corresponding minimum transmission capacity is not too distant from the half value of call arrival rate. Minimizing the average paging delay is important because it means minimizing the amount of bandwidth used for locating mobile terminals. Therefore, the numerical results of this paper will be very useful in PCS system when designing its signalling capacity due to its simplicity and effectiveness.

## References

[1] B-N. Amotz, I. Kessler and M. Sidi, "Mobile Users: To Update or not to Update?," *IEEE INFOCOM'94*, Vol. 2, pp. 570-576, June 1994.

[2] I.F. Akyildiz, J. McNair, J.S.M. Ho, H. Uzunalioglu and W. Wang, "Mobility Management in Next-Generation Wireless Systems," *Proceedings of the IEEE*, Vol.87, pp 1347-1384, August 1999.

[3] G.Y. Lee and Y. Lee, "Numerical Analysis of Optimum Timer Value for Time-Based Location Registration Scheme," *IEEE Communications Letters*, Vol. 6, No. 10, pp. 431-433, October 2002.

[4] T. Liu, P. Bahl and I. Chlamtac, "Mobility Modeling, Location Tracking, and Trajectory Prediction in Wireless ATM Networks," *IEEE Journal on Selected Areas in Communications(JSAC)*, Vol. 16, NO. 6, pp. 922-936, August 1998.

[5] T.Y.C. Woo, T.F.L. Porta, J. Golestani and N. Agarwal, "Update and Search Algorithms for Wireless Two-Way Messaging: Design and Performance," *IEEE INFOCOM'98*, pp. 737-747, March 1998.

[6] Z. Naor and H. Levy, "Minimizing the Wireless Cost of Tracking Mobile Users: An Adaptive Threshold Scheme," *IEEE INFOCOM'98*, pp. 720-727, March 1998.

[7] L.P. Araujo and J.R.B.de Marca, "A Comparative Analysis of Paging and Location Update Strategies for PCS Networks," *IEEE ICC'98*, Vol. 3. pp. 1395-1399, June 1998.

[8] W. Wang, I.F. Akyildiz and G.L. Stuber, "An Optimal Paging Scheme for Minimizing Signalling Costs Under Delay Bounds," *IEEE Communications Letters*, Vol. 5, No. 2, pp. 43-45, February 2001.

[9] I.F. Akyildiz, H.S.M. Ho, and Y-B. Lin, "Movement-Based Location Update and Selective Paging for Wireless Networks," *IEEE/ACM Transactions on Networking*, Vol. 4, No. 4, pp. 629-638, August 1996.

[10] C. Rose, "State-Based Paging/Registration: A Greedy Technique," *IEEE Transaction Vehicular Technology*, Vol. 48, No. 1, pp. 166-173, January 1999.

[11] G. Wan and E. Lin, "Cost Reduction in Location Management Using Semi-Realtime Movement Information," *ACM-Baltzer Journal of Wireless Networks (WINET),*

Vol. 5, No. 4, pp. 245-256, 1999.

[12] A. Abutaleb and V.O.K. Li, "Paging Strategy Optimization in Personal Communication System," *ACM-Baltzer Journal of Wireless Networks(WINET)*, Vol. 3, pp. 195-204, August 1997.

[13] C. Rose and R. Yates, "Minimizing the Average Cost of Paging Under Delay Constraints," *ACM-Baltzer Journal of Wireless Networks (WINET)*, Vol. 1, pp. 211-219, February 1995.

[14] H.C. Lee and J. Sun, “Mobile Location Tracking by Optimal Paging Zone Partitioning,” *IEEE ICUPC'97*, Vol. 1, pp. 168-172, October 1997.