On Hierarchical Pipeline Paging in Multi-Tier Overlaid Hierarchical Cellular Networks

Yang Xiao, *Senior Member, IEEE,* Hui Chen, *Member, IEEE,* Xiaojiang Du, *Member, IEEE,* Yan Zhang, *Member, IEEE,* Hsiao-Hwa Chen, *Senior Member, IEEE,* and Mohsen Guizani, *Fellow, IEEE*

***Abstract*—We propose a hierarchical pipeline paging (HPP) for multi-tier hierarchical cellular networks, in which different tiers overlay with one another to provide overlapped coverage of cellular service, and each mobile terminal can be paged in any tier of a network. Paging requests (PRs) are queued in different waiting queues, and multiple PRs in each waiting queue are served in a pipeline manner. We study HPP, hierarchical sequential paging (HSP), and hierarchical blanket paging (HBP) schemes analytically in terms of discovery rate, total delay, paging delay, and cost. It is shown that HPP scheme outperforms both HBP and HSP schemes in terms of discovery rate while maintaining the same cost as HSP scheme. The HPP scheme outperforms HSP scheme in terms of total delay and has a lower total delay than HBP scheme when traffic load is high.**

***Index Terms*—Multi-tier network, cellular network, hierarchical pipeline paging.**

I. Introduction

IN a multi-tier hierarchical cellular network, different tiers overlay each other to provide multi-layer coverage of cellular service, and a mobile terminal (MT) can be served and paged in any tier. For example, macrocell-microcell two-tier hierarchical cellular networks are introduced to increase capacity and handle MTs with different mobilities [1], [2]. In the future, networks more than four tiers may be needed when satellite and other networks are also considered. Paging for wireless systems has been well studied in the literature. However, most available schemes concern sequential paging in a single-tier network. In addition, existing systems use a blanket paging scheme, in which when an incoming call arrives, all cells in the local area (LA) are paged. Some sequential paging schemes were proposed to reduce paging cost, such as selective paging (SP) schemes [3] and intelligent paging schemes. In the SP scheme, when an incoming call arrives to an MT, the associated LA is divided into several paging areas (PAs), which are paged one by one until the MT is found. However, in all these schemes the paging process was considered on per user basis for exploring the best paging strategy for a particular user to reduce paging delay or other system cost for single-tier networks. In addition, paging requests (PRs) for different MTs are served in a first-in-first-out (FIFO) manner. Therefore, they introduce extra and unnecessary delay due to the fact that, during each paging cycle, unpaged cells may be idle and unused in terms of paging. In reality, many MTs can be paged simultaneously [4], [5]. Ensemble-paging algorithms [4], [6] were proposed to improve system performance. In [6], the ensemble-paging scheme, adapted from a single MT paging method, provided a group of k most likely cells to be paged for each PR in each paging cycle with known location probabilities of individual MTs. In [4], the ensemble paging was conducted on per cell basis, in which the distance between a cell and the last known cell (where an MT stays) is used as a priority to put a PR into the priority PR queue of the cell. In [8], [10], the authors proposed a two-tier (macrocell-microcell) paging scheme (also on per user basis) in which the macrocell tier is paged first, and then the microcell tier. In [9], the authors proposed a concurrent paging in a single tier network. In summary, most schemes were considered to operate either i) on per user basis to achieve a better performance in terms of cost with/without the paging delay constraint per PR, totally ignoring other PRs in the queue, or ii) in a single tier system.

The contributions of this paper are twofold. First, we propose and study a hierarchical pipeline paging (HPP) for multi-tier hierarchical cellular networks, in which PRs are queued in different waiting queues, and multiple PRs in each waiting queue are served in a pipeline manner. Second, we study and compare the HPP, hierarchical sequential paging (HSP), and hierarchical blanket paging (HBP) schemes analytically in terms of discovery rate, total delay, paging delay, and cost. We provide six Lemmas, which give very useful information. It is noted that both HSP and HBP schemes can be treated as traditional schemes applied to a multi-tier system.

This paper is organized as follows. Section II presents the proposed hierarchical pipeline paging. Section III provides analytical models. Section IV provides performance study of our proposed scheme; and Section V presents the conclusion of our work.

II. Hierarchal Pipeline Paging

Let N denote the number of tiers in a multi-tier cellular network. Assume that a tier-j cell is larger than a tier-k cell if $j < k$, and in a tier-j cell, there are several tier-$(j+1)$ cells, where $0 < j < N$. For example, in a macrocell-microcell-minicell three-tier network, macrocell-tier is tier-1, microcell-tier is tier-2, and minicell-tier is tier-3. In other words, each

Manuscript received June 1, 2008; revised November 8, 2008 and March 8 2009; accepted June 8, 2009. The associate editor coordinating the review of this letter and approving it for publication was Q. Zhang.

Y. Xiao is with Dept. of Computer Science, Univ. of Alabama, USA (e-mail: yangxiao@ieee.org).

H. Chen is with Dept. of Math & Computer Science, Virginia State Univ., USA (e-mail: huichen@ieee.org).

X. Du is with Dept of Computer and Information Sciences, Temple Univ., USA (e-mail: dxj@ieee.org).

Y. Zhang is with Simula Research Laboratory, Norway (e-mail: yanzhang@ieee.org).

H.-H. Chen (corresponding author) is with Dept. of Engineering Science, National Cheng Kung Univ., Taiwan (e-mail: hshwchen@ieee.org).

M. Guizani is with Dept. of Information Science, Kuwait University, Kuwait (e-mail: mguizani@ieee.org).

Digital Object Identifier 10.1109/TWC.2009.080717

LA has many tier-1 cells, each tier-1 cell has several tier-2 cells, each tier-2 cell has several tier-3 cells, etc. Assume that MTs can receive signals from all N tiers. Thus an MT can be paged from any tier.

Let c_j denote the cost of paging a tier-j cell, where $1 \leq j \leq N$. It is clear that $c_1 > c_2 > \cdots > c_N$ holds. Furthermore, assume that each tier-j cell has the same number of k_j tier-$(j+1)$ cells. If $c_j/k_jc_{j+1} = 1$, paging a tier-j cell and paging all tier-$(j+1)$ cells within the tier-j cell have the same total paging cost.

In the proposed HPP scheme, tier-j cells in a LA are grouped into D tier-j Paging Areas (denoted as Tier-j-PAs), where D is a paging delay hard bound, i.e., within D paging cycles, an MT associated with a being-served Paging Request (PR) is found either inside or outside the LA. A particular case is that the coverage area of all tier-j cells within a Tier-j-PA is covered by a Tier-$(j+1)$-PA. It is noted that D is not equal to the total delay constraint, T, which is the summation of the average queuing delay and the average paging delay.

There is approximately the same number of tier-j cells within each Tier-j-PA, where $1 \leq j \leq N$. When an MT is paged in a Tier-j-PA, all tier-j cells in Tier-j-PA are paged simultaneously in corresponding paging channels in the tier-j cells. Each tier implements a Paging Waiting Queue (PWQ), and there are in total N PWQs. For instance, PWQ-j is the PWQ for PRs that will be paged in Tier-j-PAs.

When a paging Request (PR) arrives to an LA, the PR is put into the PWQ-j randomly with probability p_j, where $\sum_{j=1}^{N} p_j = 1$. PRs in the PWQ-j are served using a FIFO pipeline, and multiple MTs in the same PWQ can be paged in parallel as explained next. For each served PR from the PWQ-j, Tier-j-PAs are paged one by one (a Tier-j-PA per paging cycle) until the corresponding MT is found or all Tier-j-PAs are paged. Up to $D \times L$ PRs can be served in the pipeline during one paging cycle in tier-j, where L is the number of paging channels in a tier-j cell. In the next paging cycle, those PRs whose corresponding MTs were found or all Tier-j-PAs had been paged will be removed, new PRs in the PWQ-j if available will be added, and up to $D \times L$ PRs (including both original and newly-added ones) can be served in the pipeline. More specifically, a PR goes through at most D paging cycles. If a PR is found at one paging cycle, it will not go through the next stage, and it will be removed from the pipeline, and another PR, if available in the PWQ-j, fills in the place. If a PR is paged in the current paging cycle in the Tier-j-PA k ($1 \leq k \leq D$) and it will be paged in the next paging cycle (the corresponding MT is not found and not all Tier-j-PAs are paged), it will be scheduled in the next paging cycle to page Tier-j-PA $(k+1)$ if $k+1 \leq D$ or Tier-j-PA 1 otherwise.

Next, we introduce HSP and HBP schemes. The HSP works in a similar way as the HPP scheme, except that PRs are served one by one, i.e., the HSP is a per-user scheme. In other words, only one PR can be served in each tier. The definitions on PWQ-j, Tier-j-PA, and D in the HPP scheme also apply to the HSP scheme. The HBP is similar to the HSP scheme except that $D = 1$ holds and Tier-j-PA is the same as the LA. In other words, all tier-j cells in the LA are paged at the same time.

III. Analytical Models

The performance metrics are defined as follows. Discovery rate (D_R) is defined as the departure rate of PRs on the average for all tiers, while a PR's departure means that the PR's service is completed. Cost (C) is defined as the number of cells paged per PR on the average. Paging delay ($\overline{D}$) (also called service time) is defined as the duration between a PR's time of being served and the PR's departure time on the average, where the PR's departure time is defined as the time when either the corresponding MT is found or all PAs have been paged for this PR. Total delay (T) is defined as the average time duration between a PR's arrival time and the PR's departure time. It is noted again that the total delay includes both queuing and paging delays. We assume that a PR is generated by a particular user. In addition, let us define the domain of a paging scheme as the traffic load range that allows the total delay of the scheme to have a finite value when the system reaches steady state, i.e., when the maximum discovery rate is achieved. The concept of the domain provides us with how large of a traffic load the scheme can handle.

It is assumed that the PR arrival rate follows a Poisson distribution and the paging service time (per paging cycle) follows an exponential distribution. Thus the service time is independent of the PR arrival process. Assume further that the number of paging channels per base station (i.e., Node-B in 3G) is one, i.e., $L = 1$.

Let N_j denote the numbers of tier-j cells in an LA. We assume that the queue length is infinite. Denote λ as the arrival rate, $1/\mu$ ($\lambda < \mu$) as the mean service time, and σ as the service time variance. The total delay for an M/G/1 queuing system [7] is $T_{M/G/1}(\frac{1}{\mu}, \sigma^2) = \frac{1}{\mu} + \frac{\frac{\lambda}{\mu}(1+\sigma^2\mu^2)}{2(\mu-\lambda)}$. It is noted that D is the hard bound on the number of paging cycles, while $1/\mu$ is the average time in each cycle.

A. Hierarchical Pipeline Paging

Paging request (PR) rates corresponding to the PWQ-j is $p_j\lambda$, where $j = 1, \cdots, N$. Since PRs' arrivals to the LA follow a Poisson process with rate λ, PRs' arrivals to the PWQ-j also follow a Poisson process with rate $p_j\lambda$, based on multiple-type classification of Poisson process.

Let T_j denote the average total delay of PRs in PWQ-j, including both queuing and paging delays. Let $D_{R,j}$ denote the discovery rate for PRs in PWQ-j, P_i denote the probability that a PR will be found in the i-th trial, x denote the random variable representing the number of paging cycles needed for an MT, and m denote the number of paging cycles needed on the average. Therefore, we have $P_i = \frac{1}{D}$, for $i = 1, 2, \cdots, D$ and $m = \sum_{i=1}^{D} iP_i = \frac{D+1}{2}$.

Given a value m_1 of the random variable x, the paging scheme is equivalent to that a PR needs to go through one queue and m_1 paging stages. In other words, a PR needs to go through one queue (PWQ-j) and m paging stages since m is the mean of x. We can use an $M/E_m/D$ queue to model the pipeline scheme, in which D m-stage Erlangian servers E_m work in parallel, where the m-stage Erlangian server E_m has the mean $\frac{m}{\mu}$ and the variance $\sigma^2 = 1/[m(\mu/m)^2] = m/\mu^2$. However, the total delay for an $M/E_m/D$ queue is still an open problem in queuing theory. Instead, we adopt D $M/E_m/1$

queues with a splitting arrival rate to approximately model the average performance. In fact, we expect that an $M/E_m/D$ queue has a little better performance than D $M/E_m/1$ queues with a splitting arrival rate in terms of the discovery rate and the total delay. Therefore, we can approximate the following metrics as

$$T_j = \begin{cases} \frac{m}{\mu} + \frac{p_j \lambda m(m+1)}{2\mu(D\mu - p_j\lambda m)}, & p_j\lambda m < D\mu \\ \infty, & p_j\lambda m \geq D\mu \end{cases} \quad (1)$$

$$D_{R,j} = \begin{cases} p_j\lambda, & p_j\lambda m < D\mu \\ D\mu/m = 2D\mu/(D+1), & p_j\lambda m \geq D\mu \end{cases} \quad (2)$$

$T = \sum_{j=1}^{N} p_j T_j$, $C = \{\sum_{j=1}^{N} D_{R,j} c_j \sum_{i=1}^{D} \frac{1}{D} \sum_{k=1}^{i} \frac{N_j}{D}\}/\{\sum_{j=1}^{N} D_{R,j}\}$, $\overline{D} = \frac{m}{\mu}$, and $D_R = \sum_{j=1}^{N} D_{R,j}$.

B. HBP and HSP

We adopt an $M/M/1$ queue to model the paging system for the HBP scheme, and obtain performance metrics for the HBP scheme as: $C = \{\sum_{j=1}^{N} D_{R,j} c_j N_j\}/\{\sum_{j=1}^{N} D_{R,j}\}$, $\overline{D} = 1/\mu$,

$$T_j = \begin{cases} \frac{1}{\mu - p_j\lambda}, & p_j\lambda < \mu \\ \infty, & p_j\lambda \geq \mu \end{cases} \quad (3)$$

$$D_{R,j} = \begin{cases} p_j\lambda, & p_j\lambda < \mu \\ \mu, & p_j\lambda \geq \mu \end{cases} \quad (4)$$

We can model the HSP scheme with an $M/E_m/1$ queuing model. The m-stage Erlangian server E_m has the mean m/μ and the variance $\sigma^2 = m/\mu^2$. Therefore, we have the performance metrics for the HSP scheme as: $\overline{D} = m/\mu$, $C = \{\sum_{j=1}^{N} D_{R,j} c_j \sum_{i=1}^{D} \frac{1}{D} \sum_{k=1}^{i} \frac{N_j}{D}\}/\{\sum_{j=1}^{N} D_{R,j}\}$, and

$$T_j = \begin{cases} \frac{m}{\mu} + \frac{p_j \lambda m(m+1)}{2\mu(\mu - p_j\lambda m)}, & p_j\lambda m < \mu \\ \infty, & p_j\lambda m \geq \mu \end{cases} \quad (5)$$

$$D_{R,j} = \begin{cases} p_j\lambda, & p_j\lambda m < \mu \\ \mu/m, & p_j\lambda m \geq \mu \end{cases} \quad (6)$$

It should be noted that the equations for T, C, $D_{R,j}$, and D_R for HPP can also apply to the HSP and HBP schemes.

C. Comparisons of HPP, HBP and HSP

We have the following Lemmas, which will help us better understand the paging schemes and provide some guidelines to choose good parameters. Proofs of these Lemmas are omitted due to the limited space in this letter.

Lemma 1: The HSP scheme has a lower cost than the HBP scheme when $0 < \lambda < \frac{\mu}{\max_j\{p_j\}m}$ or $\lambda \geq \frac{\mu}{\min_j\{p_j\}}$, but it performs worse than the HBP scheme in terms of discovery rate and total delay no matter what value of λ is taken. Furthermore, when the traffic load is heavy (or $\frac{\lambda}{\mu} \geq \frac{1}{\min_j\{p_j\}}$), the discovery rate of the HBP scheme is m times that of HSP scheme where $m = (D+1)/2$.

Lemma 2: The HPP scheme outperforms the HSP scheme in terms of discovery rate and total delay. The HPP scheme maintains the same cost as the HSP scheme when $0 < \lambda < \frac{\mu}{\max_j\{p_j\}m}$ or $\lambda \geq \frac{D\mu}{\min_j\{p_j\}m}$. Furthermore, when the traffic load is high (or $\frac{\lambda}{\mu} \geq \frac{D}{\min_j\{p_j\}m}$), the discovery rate of the HPP scheme is D times that of the HSP scheme.

Lemma 3: The HPP scheme outperforms the HBP scheme in terms of discovery rate. The HPP scheme keeps the same cost as the HSP scheme when $0 < \lambda < \frac{\mu}{\max_j\{p_j\}m}$ or $\lambda \geq \frac{D\mu}{\min_j\{p_j\}m}$. The HPP scheme outperforms the HBP scheme in terms of total delay when the traffic load is high (or $\lambda \geq \frac{\mu}{\max_j\{p_j\}}$). Furthermore, when the traffic load is high (or $\frac{\lambda}{\mu} \geq \frac{D}{\min_j\{p_j\}m}$), the discovery rate of the HPP scheme is D/m times that of the HBP scheme.

Lemma 4: The maximal discovery rate of the HPP scheme is 200% of the discovery rate of the HBP scheme. The discovery rate of the HPP scheme is a strictly increasing function of D when the traffic load is high (i.e., $\lambda \geq \frac{D\mu}{\max\{p_j\}m}$). When D equals to 2, 4, and 6, and the traffic is high enough (or $\lambda \geq \frac{D\mu}{\min\{p_j\}m}$), the discovery rate of the HPP scheme is 133%, 160%, and 171% of the discovery rate of the HBP scheme, respectively. When $D > 5$, the discovery rate of the HPP scheme already increases to at least 171% of that for the HBP scheme.

Lemma 5: When traffic load is high (or $\lambda \geq \frac{D\mu}{\max_j\{p_j\}m}$), the minimum cost (lower bound) of the HPP scheme or the HSP scheme is 50% of the cost of the HBP scheme. The cost of the HPP or HSP scheme is a strictly decreasing function of D. When D equals to 2, 4, and 6, the cost of the HPP or HSP scheme is 75%, 62.5%, and 58.3% of the cost of the HBP scheme, respectively. When $D > 5$, cost of the HPP or HSP scheme is already reduced to at least 58.3% of that for the HBP scheme.

Lemma 6: We can obtain $\text{Domain}^{HBP} = \{\lambda \mid \lambda/\mu < \frac{1}{\max_j\{p_j\}}\}$, $\text{Domain}^{HSP} = \{\lambda \mid \lambda/\mu < \frac{1}{\max_j\{p_j\}m}\}$, and $\text{Domain}^{HPP} = \{\lambda \mid \lambda/\mu < \frac{D}{\max_j\{p_j\}m}\}$. Furthermore, we have $\text{Domain}^{HSP} \subset \text{Domain}^{HBP} \subset \text{Domain}^{HPP}$.

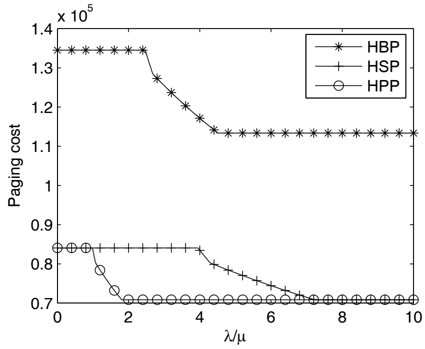
IV. Performance Evaluation

In this section, we evaluate performance for the proposed schemes. We choose a three-tier network, macrocell-microcell-minicell, as an example, i.e., $N = 3$. The number of macrocell-tier cells is 50. We further use the following parameters in this section unless stated otherwise: $k_1 = 19$, $k_2 = 7$, $c_1 = 6325$, $c_2 = 18$, $c_3 = 1$, $D = 4$, $T = 10$, $\mu = 1$, and $\lambda = 1.4$. It is noted that λ/μ can be treated as the paging traffic load.

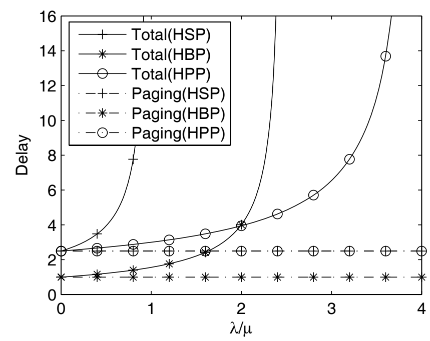
A. Comparisons of HPP, HBP and HSP

We have used the following parameters in our comparisons, $p_1 = 0.4055$, $p_2 = 0.2215$ and $p_3 = 1 - p_1 - p_2$, which were chosen randomly and the results are very typical among those values we chose.

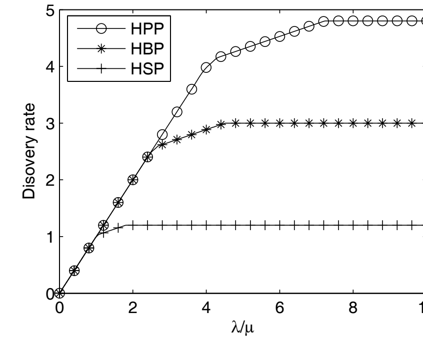
Figs. 1(a)–1(c) compare the HBP, HSP, and HPP schemes in terms of four performance metrics over the paging load. Fig. 1(a) shows that the HPP scheme has either the same cost as or less cost than the HSP scheme, and they are in general better than the HBP scheme in terms of the cost. We observe that the paging costs of all three schemes decrease a little bit in some cases due to the effects of load balancing, but they become constant eventually. The HPP scheme offers the lowest cost. Fig. 1(b) shows that the HPP scheme has the same paging delay as the HSP scheme, and they perform worse than HBP



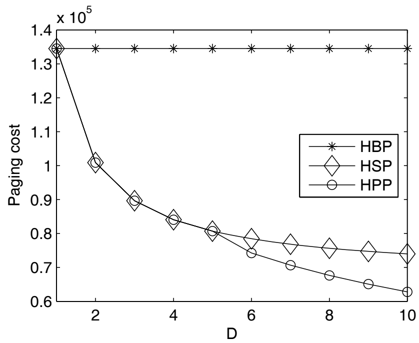
(a) Paging cost vs. λ/μ



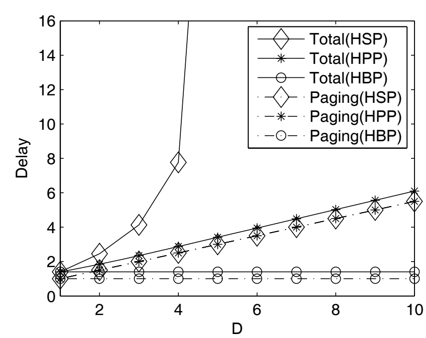
(b) Total and paging delay vs. λ/μ



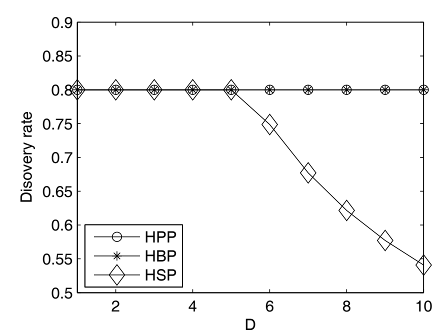
(c) Discovery rate vs. λ/μ



(d) Paging cost vs. D



(e) Total and paging delay vs. D



(f) Discovery rate vs. D

Fig. 1. Comparisons of metrics versus load: 1(a)–1(c) and comparisons of metrics versus D: 1(d)–1(f)

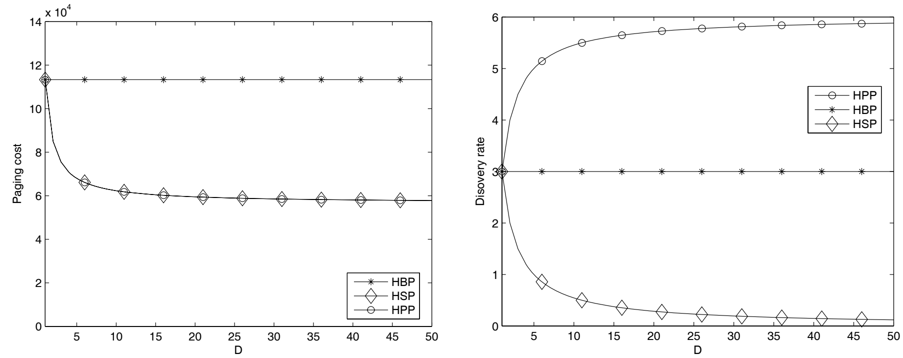
scheme in terms of the paging delay. Fig. 1(b) also shows that the HPP scheme has a shorter total delay than HSP scheme, and has a shorter total delay than HBP scheme when paging load is high. However, the paging delay is not as important as the total delay in determining the overall performance. Total delay shown in Fig. 1(b) is the summation of paging delay and queueing delay. Thus, the difference between the total delay and the paging delay is the queueing delay. Thus, we can easily infer from Fig. 1(b) that the queueing delay of either scheme increases quickly when the corresponding paging delay remains a constant, as paging load increases. Fig. 1(c) shows that the HPP scheme has the highest discovery rate and the HSP scheme has the lowest discovery rate among the three schemes.

Figs 1(d)–1(f) compare HBP, HSP, and HPP schemes in terms of four performance metrics versus D. Fig. 1(d) shows that HPP scheme has the same cost as HSP scheme when D is relatively small, and HPP has a lower cost than HSP scheme when D is large due to load balancing. They are better than the HBP scheme in terms of the cost. Fig 1(e) shows that HPP scheme has the same paging delay as HSP scheme, and they are worse than HBP scheme in terms of the paging delay. Fig. 1(e) illustrates that HPP scheme has a shorter total delay than HSP scheme, and has a longer total delay than HBP scheme since the paging load is low. We can also infer from Fig. 1(e) that, the queueing delays of the HBP and HPP schemes are almost invariant while that of the HSP scheme increases sharply. The reason is that the capacity of the HSP scheme reduces while that of the BP scheme is a constant and that of the HPP scheme increases, as D increases. Fig. 1(f) indicates that HSP scheme has the worst discovery rate among the three schemes, while HPP and HBP schemes have the same discovery rate since the paging load is low ($\lambda/mu = 1.4$).

Fig. 2 compares HBP, HSP, and HPP schemes in terms of the maximal discovery rate and the corresponding paging cost over D. Fig. 2(b) demonstrates that HPP scheme has the highest discovery rate, and HSP has the lowest discovery rate. When the paging delay constraint D is large enough, HPP scheme achieves almost 200% of discovery rate of that of HBP scheme. Fig. 2(a) shows that HPP and HSP schemes give the same cost. If the paging delay constraint D is large enough, HPP scheme achieves almost 50% of the cost for HBP scheme, whereas the discovery rate of HSP scheme is far much lower than that of HBP scheme.

V. Conclusion

We have proposed and studied a HPP scheme for multi-tier hierarchical cellular networks. We established analytical models for HPP, HSP, and HBP schemes in terms of discovery rate, total delay, paging delay, and cost. Analytical results were adopted to evaluate performance of the schemes. The following observations are made from this work: 1) Maximizing D value can minimize the paging cost of the HPP scheme with a total delay constraint. 2) The HPP scheme outperforms both HBP and HSP schemes in terms of discovery rate while maintaining at least the same cost as the HSP scheme. In some cases, the HPP scheme has a better paging cost than the HSP scheme. 3) The HPP scheme outperforms the HSP scheme in terms of total delay and has a shorter total delay than the HBP scheme. 4) When the paging delay constraint D is large enough, the HPP scheme achieves almost 200% of discovery rate and 50% of cost of the HBP scheme, whereas



(a) Paging cost vs. D (b) Discovery rate vs. D

Fig. 2. Maximum discovery rate and corresponding paging cost vs. D

the discovery rate of the HSP scheme is far much lower than that of the HBP scheme. 5) The HSP scheme has the lowest discovery rate of all three schemes. 6) The paging delay is not as important as the total delay. 7) The HSP scheme can accommodate the smallest number of Paging Requests (PRs), whereas the HPP scheme can accommodate the largest number of PRs in the system.

It is believed that hierarchical pipeline paging is promising since not only can it improve system performance, but also it is very easy to implement in real applications. Our future study is to optimize the load balance among different queues in the HPP scheme.

REFERENCES

[1] L. Hu and S. S. Rappaport, "Personal communication systems using multiple hierarchical cellular overlays," *IEEE J. Select. Areas Commun.*, vol. 13, no. 2, pp. 406-415, Feb. 1995.

[2] L.-C. Wang, G. L. Stuber, and C.-T. Lea, "Architecture design, frequency planning, and performance analysis for a macrocell/microcell overlaying system," *IEEE Trans. Veh. Technol.*, vol. 46, no. 4, pp. 836-848, Nov. 1997.

[3] W. Wang, I. F. Akyildiz, and G. Stuber, "An optimal partition algorithm for minimization of paging costs," *IEEE Commun. Lett.*, vol. 5, no. 2, pp. 42-45, Feb. 2001.

[4] S. L. Su and K. T. Chen, "Two-stage ensemble paging strategy for cellular communication systems," *IEEE Commun. Lett.*, vol. 6, no. 10, pp. 425-427, Oct. 2002.

[5] Y. Xiao, "A Parallel shuffled paging strategy under delay bounds in wireless systems," *IEEE Commun. Lett.*, vol. 7, no. 8, pp. 367-369, Aug. 2003.

[6] C. Rose and R. Yates, "Ensemble polling strategies for increased paging capacity in mobile communication networks," *Wireless Networks* vol. 2, pp. 159-167, 1997.

[7] L. Kleinrock, *Queuing Systems, Vol. 1: Theory* John Wiley & Sons, 1975.

[8] X. Wu, B. Mukherjee, and B. Bhargava, "A low-cost, low-delay location update/paging scheme in hierarchical cellular networks," in *Proc. ACM MobiDE 2003*.

[9] R. Gau and Z. J. Haas, "Concurrent search of mobile users in cellular networks," *IEEE Trans. Networking*, vol. 12, no. 1, pp. 117-130, Feb. 2004.

[10] X. Wu, B. Mukherjee, and B. K. Bhargava, "A crossing-tier location update/paging scheme in hierarchical cellular networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 4, pp. 839-848, 2006.