Reducing Registration Traffic for Multitier Personal Communications Services

Kun Il Park, *Senior Member, IEEE,* and Yi-Bing Lin, *Senior Member, IEEE*

***Abstract*—This paper presents an architecture for the multitier personal communications system (PCS) and intelligent algorithms for mobility management (specifically, the registration procedure). The multitier PCS system architecture considered in this study integrates three individual tiers: a high-tier system, a licensed low tier, and a unlicensed low tier. These three tiers are integrated into a single system by using a single home-location register (HLR) or the multitier HLR (MHLR). Under this architecture, we describe a registration protocol, where the mobile station (MS) is allowed to register to the MHLR on only one tier at any given time. We propose several intelligent algorithms for the MS to determine whether to perform registration or not in various situations to reduce the registration traffic.**

***Index Terms*—Mobility management, multitier, personal communications, registration.**

I. Introduction

WIRELESS technologies are often grouped into "high-tier" and "low-tier" technologies. Although there is no formal definition of these terms, high- and low-tier technologies are distinguished by several characteristics. High-tier technologies use a high-radio-transmitting power with antennas mounted on tall towers covering a large area and are designed to operate at high vehicular speeds. On the other hand, low-tier technologies use a low-radio-transmitting power covering a small area (i.e., a microcell) and are intended to operate at low (e.g., pedestrian) speeds. A low-tier system would be cost effective for high-density environments, such as metropolitan areas and urban residential areas, whereas a high-tier system would be cost effective for covering large areas. The motivation for the *multitier system* [6], [8], [11] is to integrate the high- and low-tier systems into a single system to provide the advantages of both tiers in an integrated manner. We describe the multitier architecture and its registration protocol for roaming management. Then, we propose and analyze several intelligent algorithms to reduce the amount of registration traffic.

II. The Multitier PCS System Architecture

Fig. 1 illustrates a multitier PCS system architecture. This architecture integrates three individual tiers: a high-tier system (e.g., AMPS [1], IS-95 [4], or IS-54 [3]) using the IS-41 network-signaling protocol, a low-tier system operating at the licensed frequency band based on the licensed PACS (the licensed low tier) [2], and a low-tier system operating in the unlicensed frequency band based on the unlicensed PACS system, i.e., PACS-UB (the unlicensed low tier or unlicensed tier for short) [9], [10]. The three tiers are tied together into a single system by connecting the individual tiers' visitor location registers (VLR's) with a common home-location register (HLR) called multitier HLR (MHLR). Both the high- and low-tier VLR's communicate with the MHLR using IS-41 over signaling system 7 (SS7). In concept, the unlicensed tier system can have the same architecture as the high-tier and licensed low-tier system. In reality, however, the two most likely applications of the unlicensed tier system are the wireless private branch exchange (PBX) and the home-base unit, both of which need to be treated differently from the other two tiers. In Fig. 1, the wireless PBX and the home-base unit represent the unlicensed tier systems. In the case of the wireless PBX, the VLR may be part of the PBX or may be a separate entity, such as the "wireline VLR." In both cases, the VLR would communicate with the MHLR using IS-41 over SS7. The home-base unit application may work as follows. As the user comes home with the mobile station (MS), the home-base unit would sense its presence and automatically dial a preassigned number at the switching service point (SSP). Alternatively, the user can dial the same number manually. In either case, this call to the SSP would trigger a sequence of message exchanges between the SSP and service control point (SCP) and between the SSP and the home-base unit to register the MS on the SCP. The SCP would then inform the MHLR that the MS is registered at the SCP. The SCP in this case functions like a VLR. The automatic operation of the home-base unit would require a slight modification to this protocol to replace the announcement used for the human operator with dual-tone multifrequency (DTMF) tones for the home-base unit.

III. Registration and Call Delivery

Registration is a process by which the current location of the MS is updated in the appropriate data bases, such as the VLR and MHLR. Registration is required as the MS moves from one registration area to another. We consider a registration protocol for the multitier system, where the MHLR keeps the registration record of the MS on only one tier at any given time. If the unlicensed tier is available, the MS registers to this tier and ignores both the low and the high tiers. That is, the MS will receive services at the unlicensed tier. If the

Manuscript received September 27, 1995; revised February 13, 1996. This work was supported in part by CCL/ITRI, R.O.C.

K. I. Park is with Bellcore, Red Bank, NJ 07701 USA.

Y.-B. Lin is with Department and Institute of Computer Science and Information Engineering, National Chiao-Tung University, Hsinchu 300, Taiwan, R.O.C. (e-mail: liny@csie.nctu.edu.tw).

Publisher Item Identifier S 0018-9545(97)04619-7.

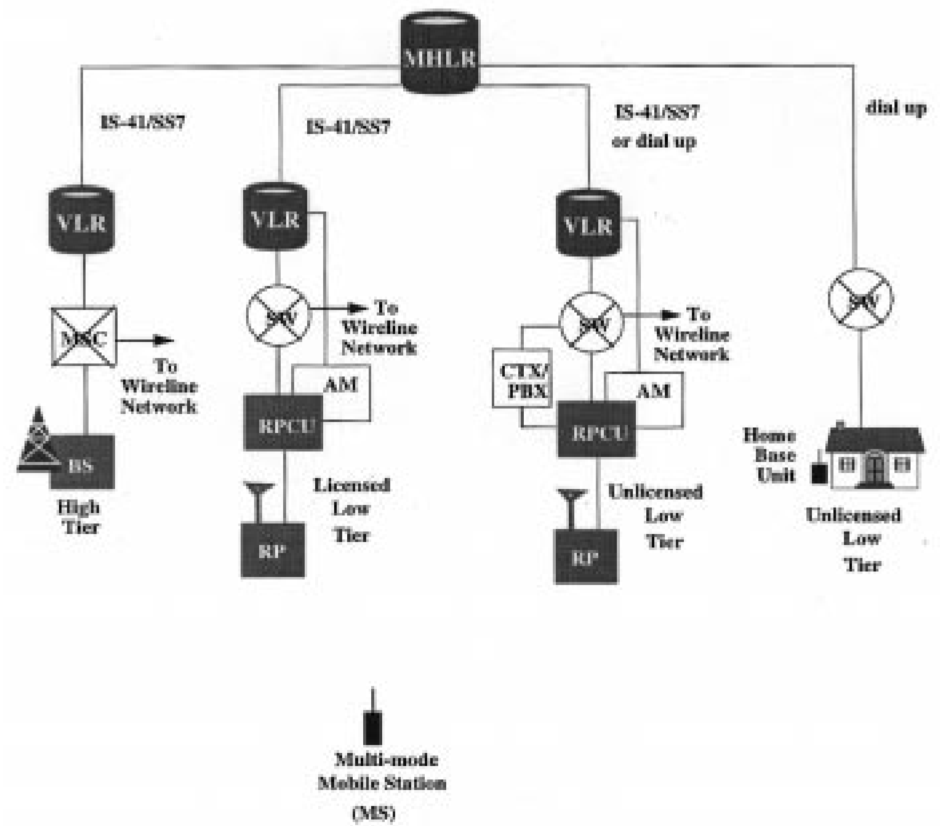

Fig. 1. Multitier PCS system architecture.

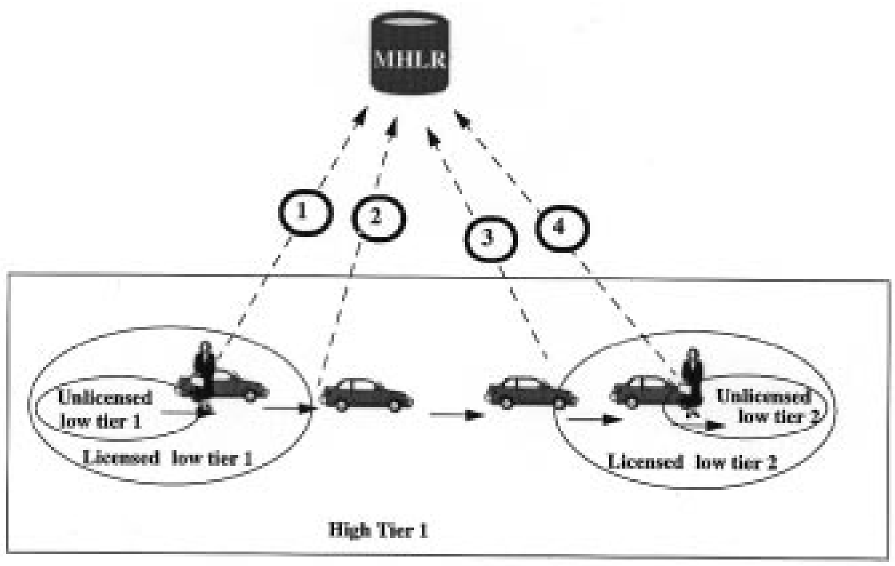

Fig. 2. Multitier registration. (1) Registration on licensed low-tier registration area 1. (2) Registration on high-tier registration area 1. (3) Registration on licensed low-tier registration area 2. (4) Registration on unlicensed low-tier area 2.

unlicensed tier is unavailable, but the low tier is available, then the MS will be in the low tier and ignore the high tier. If both the unlicensed and low tiers are not available, then the MS will receive services at the high tier. The above rules follow the fact that the service cost in the unlicensed tier is the lowest among the three tiers. The cost for the high tier is the highest.

Fig. 2 illustrates the occurrences of registrations as the MS moves from the unlicensed tier into a licensed low tier, from the licensed low tier into a high tier, from the high tier into another licensed low tier, and, finally, from the licensed low tier into another unlicensed tier. Note that when the MS registers at a new tier, the registration to the previous tier is

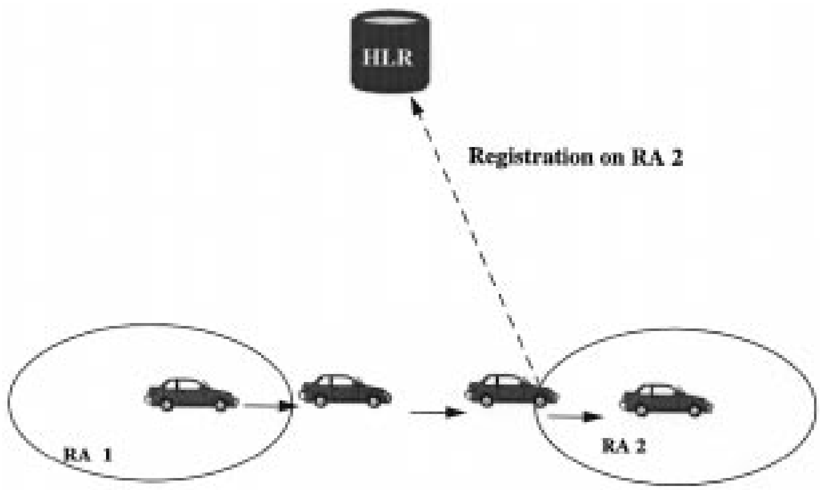

Fig. 3. Registrations in a single-tier system.

cancelled. As a reference for comparison, Fig. 3 illustrates the registrations in a single-tier system.

Call delivery refers to the process of locating the called mobile user and establishing a connection from the calling party to the called party, i.e., the mobile user. Note that in our multitier registration protocol, the MHLR keeps the registration record of the MS on one tier at any given time. When the MHLR receives a location identification request for an incoming call to the MS, it simply queries the VLR of the tier on which the MS is currently registered, and, based on the temporary local directory number (TLDN) supplied by the VLR, the call is routed to the appropriate switch. In other words, with the SR method, the call-delivery algorithm is the same as that in a single-tier system.

IV. Intelligent Algorithms for MS Registration

Since registrations are required when an MS switches tiers in the multitier system, the operation of such a system may significantly increase the network-signaling traffic. This section proposes intelligent algorithms to reduce the amount of registration traffic for the multitier system. These algorithms are exercised at the MS to determine whether the registration operation must be performed without changing the network protocols of registration. We first describe the tier-registration criteria.

Case 1) *When the MS is first turned on.* The MS sequentially checks the unlicensed tier, the licensed low tier, and then the high tier. If all three tiers are not available, the MS keeps monitoring the three tiers.

Case 2) *When the MS is in the unlicensed tier.* The MS then monitors the unlicensed tier only and ignores the other two tiers as long as the unlicensed tier is available. If the unlicensed tier becomes unavailable, Case 1) is exercised.

Case 3) *When the MS is in the licensed low tier.* The MS monitors the licensed low tier and also scans the unlicensed tier. If the unlicensed tier becomes available, the MS switches to the unlicensed tier as in Case 2). If the licensed low tier becomes unavailable, Case 1) is exercised.

Case 4) *When the MS is in the high tier.* The MS also scans both the unlicensed and licensed tiers. If either of the low tiers is available, the MS switches to that low tier (where the unlicensed tier has higher priority over the licensed tier). If the high tier becomes unavailable, then Case 1) is exercised.

To simplify our discussion, the remainder of this paper only considers the two-tier system (the high tier and the licensed low tier). The extension to the three-tier system is trivial. Based on Case 4), an MS registers to the low tier when it moves from the high to the low tier (i.e., when the MS detects that the low tier is available). In some cases, the movement into the low tier is *transient*—the MS will move back to the high tier shortly. An example is that the mobile user is driving in the highway, where the low tier is not available due to the high speed of the vehicle. The vehicle may stop at a toll booth and then speed up again. When the vehicle is “temporarily” idle, the MS may detect the availability of the low tier and register to the low tier. 5 min later, the MS will lose the contact to the low tier and have to reregister to the high tier again. During the 5 min, the MS may not receive any phone call, and the two registrations are not desirable. To avoid the extra registration traffic due to the transient situation, three registration strategies are proposed.

1) *Fixed delay registration (FDR).* When the MS detects that the low tier is available at time t, it does not register to the low tier immediately. Instead, it waits for a fixed period T_d and actually registers at the end of the delay period if the low tier is still available. During the period T_d, a call termination is handled by the more expensive high tier. Two output measures are considered to evaluate the delay registration: the probability $p_{\rm reg}$ that a registration is required at $t+T_d$ (i.e., the low tier is still available at time $t+T_d$) and the probability $p_{\rm call}$ that a call termination arrives during $[t, t+T_d]$ (and is delivered by the high tier). Note that $p_{\rm reg}$ and $p_{\rm call}$ are conflict goals, which complicate the selection of T_d.

The delay T_d should be carefully chosen to satisfy the network design goals. The value of T_d for an MS can be preassigned based on the user profile. However, the user mobility may change dynamically, and the fixed T_d approach may not be effective.

2) *Adaptive delay registration (ADR).* This algorithm is the same as the FDR except that it automatically adjusts the registration delay T_d to accommodate the changing of the user mobility. Suppose that an MS is originally in the high tier. The MS detects that the low tier is available at time t.

 a) *Case 1).* A call arrives during $[t, t+T_d]$. This case indicates that the selection of T_d is too long. The MS decreases T_d.

 b) *Case 2).* No call arrives during $[t, t+T_d]$, and the low tier is still available at $t+T_d$. This case may imply that the selection of T_d is too short and may not be able to capture the transient situation. The MS increases T_d.

 If the MS originates a call during $[t, t+T_d]$, it selects the appropriate tier for the service and the delay-registration timer is disabled. The initial T_d value for the MS is arbitrarily chosen (our performance study indicates that the initial T_d value does not affect the results). When the MS decreases T_d, the reduction may be based on several criteria. For example, suppose that the goal is to make sure that $p_{\text{call}} < 0.05$. In Case 1), the MS may decrease T_d based on the p_{call} statistic collected so far. If $p_{\text{call}} > 0.05$, the MS may significantly reduce T_d (e.g., the new value will be 0.1 times of the old value), and if $p_{\text{call}} < 0.05$, the MS may moderately increase T_d (e.g., the new value is 1.05 times of the old value). Another possibility is to consider the past "history" in Case 1). If calls arrived during the last two registration delays, then T_d is significantly reduced (e.g., by 24% in our experiments). Otherwise, T_d is moderately reduced (e.g., by 60% in our experiments). Similarly heuristics can be used in Case 2).

3) *Event-driven registration (EDR).* In this algorithm, the MS does not periodically monitor the low tier. Instead, the MS looks for a low tier only if a certain event occurs: when it originates a call or just after receiving an incoming call or when it loses the high tier.

V. Performance of the Intelligent MS Algorithms

This Section analyzes the performance of the intelligent MS algorithms. We first compare the performance of the adaptive algorithm with the fixed approach. The performance study is conducted by a discrete-event simulation approach [7]. We assume that the interval t_l, during which the low tier is available, has a Gamma distribution with mean $1/\lambda_l$ and the variance V [5]. (The Gamma distribution is selected because it can be conveniently used to approximate many other distributions.) We assume that the arrivals of the call terminations are a Poisson process with the arrival rate λ. Assume that the mean low-tier available period $E[t_l] = 500$ min with the variance $V = 1000$. Figs. 4 and 5 indicate that

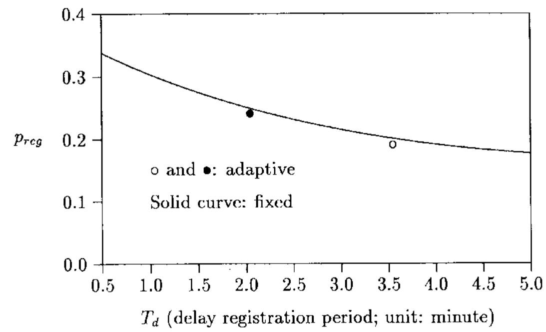

Fig. 4. Comparing the fixed and the adaptive registration delays: p_{reg}. o: one call arrival per 50 min; •: one call arrival per 25 min. p_{reg}: the probability that a registration is required after the low tier is available.

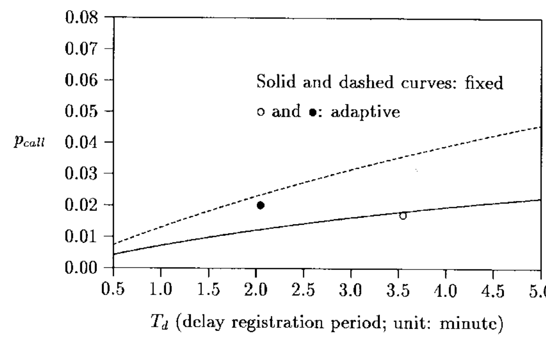

Fig. 5. The comparison of the fixed and adaptive registration delays: p_{call}. Solid curve: fixed with one call arrival per 50 min; Dashed curve: fixed with one call arrival per 25 min; o: adaptive with one call arrival per 50 min; •: adaptive with one call arrival per 25 min. p_{call}: the probability that a call is delivered in the high tier when the low tier is available.

for $\lambda = 10\lambda_l$ (i.e., one call arrival per 50 min) and $20\lambda_l$ (i.e., one call arrival per 25 min), the expected T_d in the adaptive algorithm is $0.007/\lambda_l$ (= 3.5 min) and $0.004/\lambda_l$ (= 2 min), respectively. Both p_{call} and p_{reg} for the adaptive algorithm are lower than the fixed T_d approach at the same T_d values. The improvement is not significant and probably not very important. The major advantage of the adaptive algorithm (with our selected parameters) guarantees that $p_{\text{call}} < 0.1$ for all λ values and different γ values for the lower tier residence times (see Fig. 6). This goal cannot be achieved by the FDR.

Fig. 7 indicates that under the constraint that $p_{\text{call}} < 0.1$, the adaptive algorithm still reduces the number of low-tier registrations when the variance of t_l is large (i.e., $V = 1000$). When the variance of the t_l distribution is small (i.e., $V = 0.001$), the transient situations seldom occur, and we expect that p_{reg} is large. Note that in the real world, the variance of the low-tier residence times is large, and the ADR is very effective. The low-variance case is meant to be included for the completeness of our study.

The ADR and EDR are compared as follows. Let the intercall-arrival time be t_c and the low-tier available time (i.e.,

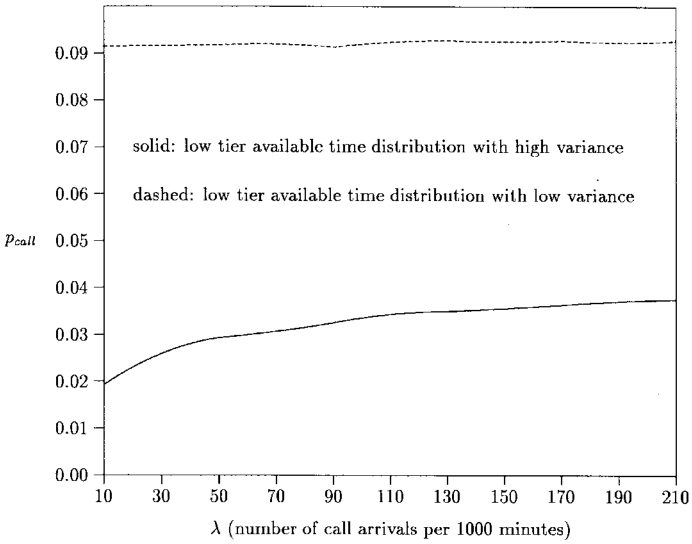

Fig. 6. The probability p_{call} of the adaptive algorithm as a function of call-arrival rate λ ($E[t_l] = 500$ min). p_{call}: the probability that a call is delivered in the high tier when the low tier is available.

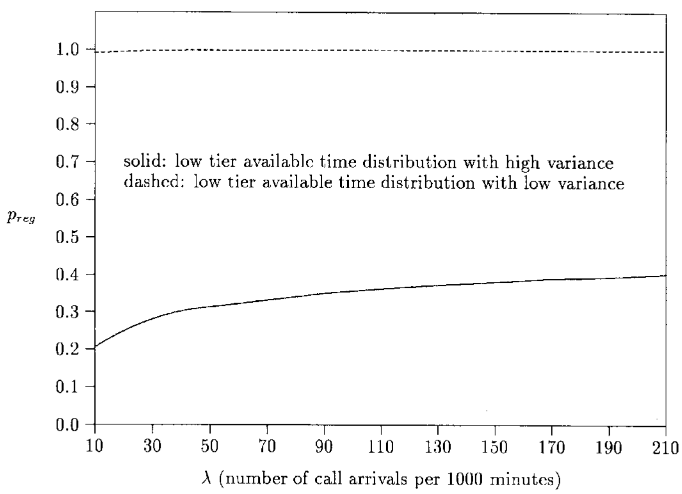

Fig. 7. The probability p_{reg} of the adaptive algorithm as a function of call-arrival rate λ ($E[t_l] = 500$ min). p_{reg}: the probability that a registration is required after the low tier is available.

t_l) distribution be f_l. For the EDR

$$p_{\text{call}} = \int_{t_l=0}^{\infty} \int_{t_c=0}^{t_l} \lambda e^{-\lambda t_c} f_l(t_l) dt_c dt_l = 1 - f_l^*(\lambda)$$

where $f_l^*(s)$ is the Laplace transform of f_l [12]. If f_l is a Gamma distribution with mean $1/\lambda_l$ and the variance $1/(\gamma\lambda_l^2)$, then from [12]

$$f_l^*(\lambda) = \left(\frac{\gamma\lambda_l}{\gamma\lambda_l + \lambda}\right)^{\gamma} \quad \text{and} \quad p_{\text{call}} = 1 - \left(\frac{\gamma\lambda_l}{\gamma\lambda_l + \lambda}\right)^{\gamma}.$$

In the EDR, if the MS detects that the low tier is available after a high-tier call delivery, then it registers to the low tier. Thus

$$p_{\text{reg}} = p_{\text{call}}.$$

Fig. 8 compares the probabilities p_{reg} and p_{call} for both the ADR and EDR (in this figure, $V = 1000$ for the high-

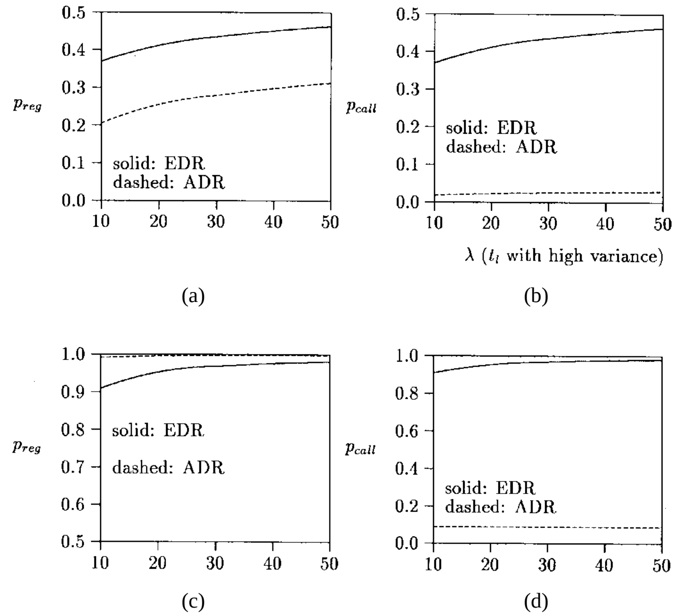

Fig. 8. The comparison of the ADR and EDR. p_{reg}: the probability that a registration is required afer the low tier is available. p_{call}: the probability that a call is delivered in the high tier when the low tier is available. t_l: the lower available period ($E[t_l] = 500$ min). λ: the number of call arrivals per 500 min.

variance case and $V = 0.001$ for the low-variance case). The figure indicates that the ADR is better than the EDR in terms of p_{call}. For p_{reg}, the EDR is better than the ADR when the variance for the t_l distribution is small.

In summary, our study indicates that the ADR is always better than the FDR and that the ADR is better than the EDR in terms of p_{call}. In terms of p_{reg}, the ADR is better than the EDR under certain cases and vice versa. Note that p_{call} and p_{reg} are conflict goals in view of the delay period. In a PCS system, several other factors contribute to p_{call}. For example, no radio channels are available, no network resources (e.g., the circuits between the base stations and mobile switching centers) are available, or the network response times are too long (e.g., call setup timer expires). In the intelligent MS algorithms, it is useless to select a delay period that results in a p_{call} (caused by delay registration) smaller than the probability of the lost calls due to other factors. Thus, a typical guideline to engineering the delay period is to minimize p_{reg} based on a p_{call} value determined by the capabilities of other network resources.

VI. Conclusions

We described an architecture and a registration protocol for the multitier PCS system. In the proposed registration protocol, the multitier HLR keeps the registration of the MS on only one tier at any given time. Since tier switching may significantly increase the network signaling traffic, it is desirable to reduce the registration operations in a multitier system. For this purpose, we proposed three delay-registration algorithms: the FDR, ADR, and EDR algorithms. These algorithms merely determine when to register and are independent of the network protocols, that is, the algorithms do not change the registration protocol *per se*. The decision on which algorithms to use in the

MS will depend on the tradeoff between the MS complexity and the benefits to the network, such as a reduced registration traffic. It should be noted that use of the ADR or the EDR is not a prerequisite, that is, the multitier registration protocol works independently of these intelligent MS algorithms.

Acknowledgment

The authors would like to thank N. Crystal, D. Lukacs, H. Sherry, and M. Ulema for the discussions and comments they provided on the initial draft of this document and the reviewers who provided valuable comments to improve the quality of this paper. The assistance of Prof. J. R. Cruz is highly appreciated.

References

[1] "Mobile station-land station compatibility specification," ANSI/EIA/TIA, Tech. Rep. 553, 1989.
[2] "Generic criteria for version 0.1 wireless access communications systems (WACS) and supplement," Bellcore, Tech. Rep. TR-INS-001313, 1994.
[3] "Cellular system dual-mode mobile station—Base station compatibility standard," EIA/TIA, Tech. Rep. IS-54, 1992.
[4] "Mobile station-base station compatibility standard for dual-mode wideband spread spectrum cellular system," EIA/TIA, Tech. Rep. IS-95, 1993.
[5] N. L. Johnson, *Continuous Univariate Distributions-1*. New York: Wiley, 1970.
[6] Y.-B. Lin, "A comparison study of the two-tier and the single-tier personal communications services systems," *ACM/Baltzer Mobile Networks Applicat.*, vol. 1, no. 1, pp. 29–38, 1996.
[7] Y.-B. Lin and V. K. Mak, "Eliminating the boundary effect of a large-scale personal communication service network simulation," *ACM Trans. Model. Comput. Simulation*, vol. 4, no. 2, pp. 165–190, 1994.
[8] Y.-B. Lin, L. F. Chang, A. R. Noerpel, and K. Park, "Performance modeling of multi-tier PCS system," *Int. J. Wireless Inform. Networks,* vol. 3, no. 2, pp. 67–78, 1996.
[9] A. R. Noerpel, L. F. Chang, and Y.-B. Lin, "Performance modeling of polling de-registration for unlicensed PCS," *IEEE J. Select. Areas Commun.*, vol. 14, no. 4, pp. 728–734, 1996.
[10] A. R. Noerpel, L. F. Chang, and R. A. Ziegler, "PACS-UB, A protocol for the unlicensed spectrum," to be published.
[11] J. F. Rizzo and N. Sollenberger, "Multitier wireless access," *IEEE Personal Commun. Mag.*, vol. 2, no. 3, pp. 18-31, 1995.
[12] E. J. Watson, *Laplace Transforms and Applications*. Birkhauserk, 1981.

Kun Il. Park (S'69–M'72–SM'85) received the B.S. degree from Seoul National University, Seoul, S. Korea, in 1966 and the M.S. and Ph.D. degrees from the University of Pennsylvania, Philadelphia, in 1968 and 1972, respectively, all in electrical engineering.

From 1973 to 1983, he was with Bell Laboratories, Holmdel, NJ, where he was the Supervisor of a Data Communications Network Planning Group. Since 1984, he has been with Bell Communications Research, Inc. as a District Manager of an ISDN Planning Group responsible for ISDN/broadband ISDN network planning, ATM congestion control, and network performance standards. He was also the Director of a Network Operations Analysis Group responsible for network operations systems software requirements and a Professional Services Consultant responsible for systems engineering for PCS and wireless communications. He is also an Adjunct Professor of Electrical Engineering at the Stevens Institute of Technology, Hoboken, NJ. His current areas of interest include PCS, trunked radio systems (TRS's), wireless signaling protocols such as IS-41 and GSM MAP, wireless data, short-message services, wireless-access standards, and wireless access to the information highway, such as the Internet. He has published papers for conferences and journals, including the *Bell System Technical Journal*, and the IEEE Transactions on Communications. He is the author of *Personal and Wireless Communications: Digital Technology and Standards* (Boston: Kluwer, 1996). He is a Coinventor of a patent on a multitier PCS system currently awaiting approval.

Dr. Park participated in the standards activities of Committee T1, the National Bureau of Standards, the Corporation for Open Systems, and, most recently, the Joint Technical Committee for PCS standards.

Yi-Bing Lin (SM'95), for a photograph and biography, see this issue, p. 596.