Analytical Results for Optimal Choice of Location Update Interval for Mobility Database Failure Restoration in PCS Networks

Yuguang Fang, *Senior Member*, *IEEE*, Imrich Chlamtac, *Fellow*, *IEEE*, and Hong-Bing Fei

Abstract—When the location information for a mobile user in the Home Location Register (HLR) is corrupted or obsolete, the new arriving calls to the user may be lost. In order to minimize the effect of such HLR mobility database failure, a location update scheme called *period location updating* is proposed, and the cost analysis is performed. Analytical results for the failure recovery time distribution and the average number of call losses, which are crucial for cost analysis, are presented. The optimal location update period is given analytically as a function of other traffic parameters. This optimal choice of location update period lays the foundation for the adaptive adjustment of the location update period in failure restoration for PCS networks.

Index Terms—Location update, failure recovery, HLR, PCS, mobility database, mobility management.

✦

1 INTRODUCTION

PERSONAL Communications Services (PCS) networks enable mobile users to communicate with each other at any time from any location using any form of services. In order to achieve this goal, it is necessary to locate mobile users within the time limit (the maximum setup time) before the call connections can be established for them to receive calls. This then requires that the location information in the PCS networks be correct and up-to-date ([5], [4]).

In the EIA/TIA IS-41 standard ([6]), the user location strategies use the two-level hierarchical registration schemes. In a PCS network, the Home Location Register (HLR) is the location register which maintains mobile users' identity information containing mobile users' information such as directory number, profile information, current location, authentication information, and billing information. The HLR is a database residing in the home system of a mobile. The Visitor Location Register (VLR) is the location register besides the HLR, which is used to retrieve information necessary for handling calls from or to a mobile which visits another area or system different from its home system, a VLR is a database associated with a PCS network that the mobile user is currently visiting. When the mobile is in its home system (where the mobile subscribes to its service), the location information of the mobile can be directly accessed from the HLR in the home system—when a call is to the mobile, the Mobile Switching Center (MSC)

can directly get the location information from HLR and direct the arriving call to the mobile. This can be better illustrated in Fig. 1, where the major signaling steps are shown, and the detailed procedure can be found in [12]. If the mobile travels in a visiting system (the PCS network the mobile is currently visiting), the mobile initiates a registration process with the new visited MSC. During this process, the mobile's identity information is created from the HLR and is stored in the VLR in the visited system, then the visited MSC sends the mobile's location to the home MSC and location database in its HLR is updated. When a call terminating to the mobile arrives, the call is directed to the originating MSC, then a contact with the home MSC of the mobile is made with the help of its HLR. The location information stored in the HLR database is used, a contact with the MSC in charge of the mobile is initiated, the location of the mobile will be found via its VLR, and the call connection can be established in this way. The main signaling steps for this terminating call process are shown in Fig. 2, detailed procedure can be found in [12].

It is obvious that the location information in HLR is very important for call delivery. When the location information at HLR is incorrect, either system wide paging should be done, which is expensive in terms of traffic generation, or the arriving call is dropped, which is expensive in terms of customer care. The location information will be changed when a call is initiated from the mobile, or when it crosses the location area boundary, or when a location update is performed. Due to such constant changes of location information for a mobile in the Home Location Register (HLR), the location information may be corrupted. When the mobile does not register often, the location information in the HLR may be obsolete. These lead to the HLR mobility database failure, which then results in call blocking. As a

- Y. Fang is with the Department of Electrical and Computer Engineering, University of Florida, 435 Engineering Building, P.O. Box. 116130, Gainesville, FL 32611. E-mail: fang@ece.ufl.edu.
- I. Chlamtac and H.-B. Fei are with the Erik Jonsson School of Engineering and Computer Science, University of Texas, Mail Station EC33, P.O. Box 830688, Richardson, TX 75083. E-mail: {chlamtac, hbfei}@utdallas.edu.
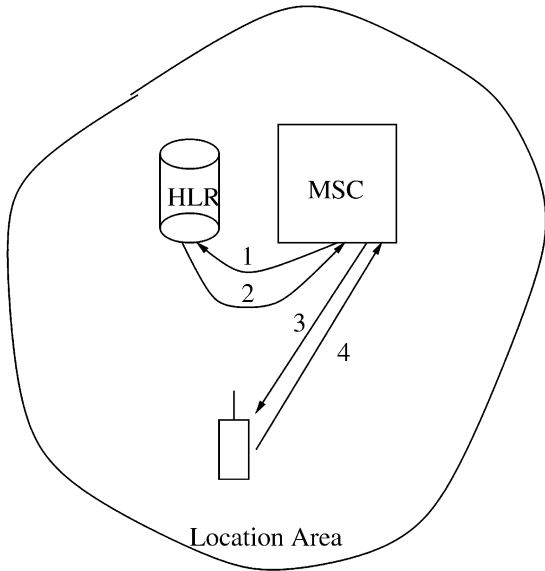
Fig. 1. Terminating call process when the called mobile is located at the home system.

service provider, it has to minimize this blockage. In this study, we neglect the paging possibility. This is because if we do not have the location information of a mobile user, we may need to do system wide paging, which is obviously too expensive. If partial location information is known, we can use the similar procedure together with the technique developed in [3] to carry out the cost analysis. The HLR failure can be recovered either when the mobile initiates a call in a cell, which reveals the current location of the mobile in the PCS network, or when the mobile crosses a Location Area (LA) boundary, where a registration is needed or an active deregistration is carried out. However, there is a significant problem with these two recovery schemes: If the mobile rarely initiates any calls and it also moves very slowly, the location information will be obsolete after some time and the delay via the paging scheme ([1])

may be so long that the calls to the mobile before a HLR failure recovery will be lost. In order to overcome this problem, the *autonomous registration* mechanism was proposed ([7]), in which a mobile periodically reregisters with the system. In this approach, a mobile periodically establishes radio contact with the PCS network to update its location. These location updates are registrations to update the mobile's current location information in its HLR. It has been shown that the HLR restoration delay is reduced.

Failure recovery of PCS mobility databases have been studied intensively in the last few years. In EIA/TIA ([6]) IS-41, the HLR and VLR are not backed up in nonvolatile storage. After a failure, the HLR incrementally reconstructs the mobile's location records when the mobile sends a registration message. In GSM ([7]), the HLR or VLRs are backed up periodically. After a failure, the HLR or VLRs are restored from the nonvolatile storage. Lin ([19], [20]) modeled the HLR and VLR restoration with and without check-pointing in IS-41, and GSM conducted the performance analysis. It is also possible for HLR to aggressively restore its location records by requesting the known VLRs to provide the exact location information. Wang et al. ([23]) proposed a novel aggressive approach for failure recovery of PCS mobility databases and analyzing its performance.

One of the important problems among these few schemes is how to choose the location update interval. Recently, Haas and Lin ([14]) studied the effect of the HLR failures on various system parameters and came up with a set of recommendations for setting up the value of the periodic interval for the autonomous registration mechanism. However, in all aforementioned schemes and analysis, it is commonly assumed that the LA residence time and the interarrival interval of the initiating calls are all exponentially distributed. The cost analysis is done mostly via simulations. It will be more desirable to give some analytical results under general assumptions on the LA residence time and the interarrival time of the new initiating
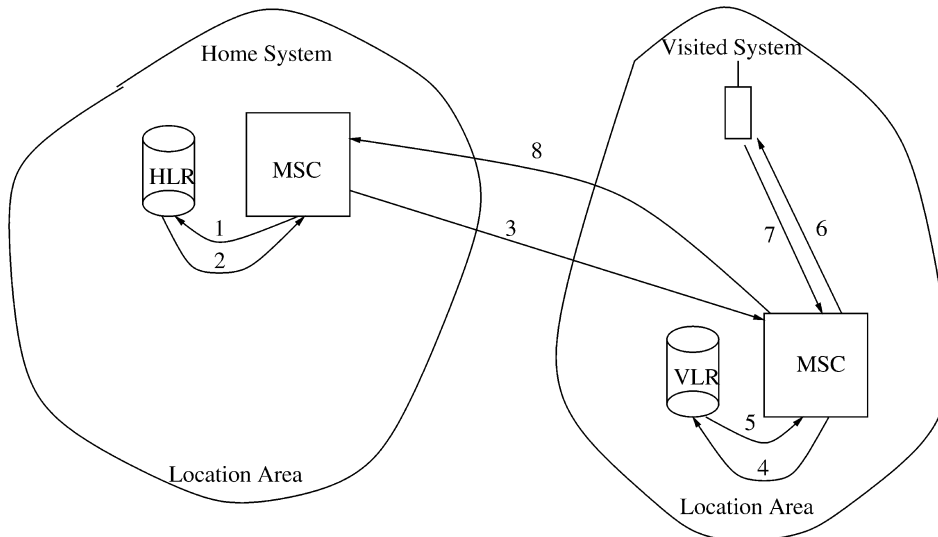

Fig. 2. Terminating call process when the called mobile is located at the visited system.
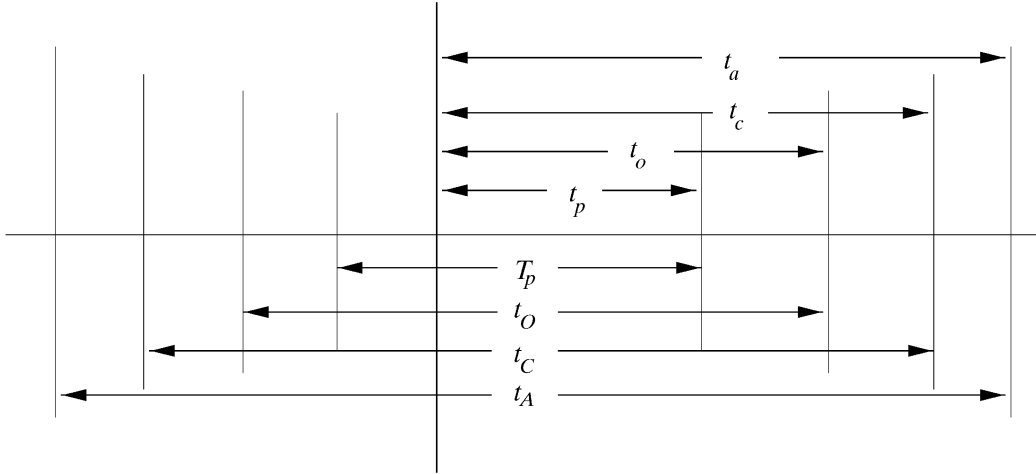
Fig. 3. The time diagram for failure restoration.

calls, and relate the location update interval to the traffic parameters.

In this paper, we present analytical results for the performance analysis for the autonomous registration for the mobility database failure recovery for the PCS networks under more general realistic assumption. For some specific cases, we give analytic formula relating the location update interval to the system parameters. These results can be used to adaptively change the location update interval according to the traffic and mobility conditions. We expect that the results will play a significant role in the active failure restoration of mobility databases in PCS networks.

2 PROBABILITY DISTRIBUTION OF FAILURE RECOVERY TIME

In this section, we proceed to conduct the performance analysis of the periodic location update scheme (autonomous registration) for the HLR mobility database failure recovery. The performance metrics we are interested in are the probability distribution of the failure recovery time, the call loss given the information on the failure rate, LA residence time (i.e., the time that a mobile spends in a LA area), the call arrival process, and the originating call process. As we mentioned earlier, if the failure recovery time is long, then the signaling traffic is not a problem, however, the number of calls arriving to a mobile whose location information is not available due to the failure will increase. On the other hand, if the failure recovery time is short (hence, the location update interval is short), we do reduce the number of lost calls; however, the signaling traffic due to the update will increase. Therefore, there is obviously trade-off here; in order to carry out such cost tradeoff analysis, we will need the probability distribution of the failure recovery time and the average call loss which are used to determine how often the location updates should be.

Consider the scenario for a mobile experiencing a failure in the PCS networks. Let t_O, t_A and t_C denote, respectively, the interarrival time of calls initiating from the mobile, the interarrival time of calls arriving to the mobile and the location area (LA) residence time. Let t_o, t_a, and t_c denote, respectively, the residual life times of t_O, t_A, and t_C when a database failure occurs. Fig. 3 shows the time diagram for our study. Let T_p denote the location update interval and let t_p denote the residual life time when a database failure occurs. From the Residual Life Theorem ([17]), we obtain that the distribution of t_p is uniform in the interval $[0, T_p]$, which is intuitively reasonable. As we mentioned earlier, the database failure in HLR can be recovered at either the next location update instant, or the next initiating call arrival instant, or the instant when the mobile moves out of the LA. Hence, the database failure recovery time is given by

$$t_r = \min\{t_o, t_c, t_p\} = \min\{\min\{t_o, t_c\}, t_p\}$$
$$= \min\{t_u, t_p\}, t_u = \min\{t_o, t_c\}. \quad (1)$$

Let $f_o(t)$, $f_a(t)$, $f_c(t)$, and $f_u(t)$ be the density functions of t_o, t_a, t_c, and t_u with the corresponding Laplace transforms $f_o^*(s)$, $f_a^*(s)$, $f_c^*(s)$, and $f_u^*(s)$, respectively, (we will use star $*$ to denote the Laplace transform for the corresponding density function, and use $x^{(i)}(t)$ to denote the ith derivative for any function $x(t)$). Let $f_p(t)$ denote the density function of t_p, which is given by

$$f_p(t) = 1/T_p, \ 0 \le t_p \le T_p.$$

Since the call origination and the LA crossing are independent events, hence the random variables t_o and t_c are independent. From (1), we obtain the cumulative distribution function of t_u

$$F_u(t) = \Pr(t_u \le t) = \Pr(t_o \le t \text{ or } t_c \le t)$$
$$= \Pr(t_c \le t) + \Pr(t_o \le t) - \Pr(t_c \le t, t_o \le t) \quad (2)$$
$$= \Pr(t_c \le t) + \Pr(t_o \le t) - \Pr(t_c \le t)\Pr(t_o \le t).$$

Differentiating both sides of (2), we obtain

$$f_u(t) = f_c(t) + f_o(t) - f_c(t)\Pr(t_o \le t) - \Pr(t_c \le t)f_o(t)$$
$$= f_c(t)\int_t^\infty f_o(\tau)d\tau + f_o(t)\int_t^\infty f_c(\tau)d\tau. \tag{3}$$

Notice that the Laplace transform of $\int_t^\infty f(\tau)d\tau$ is $(1 - f^*(s))/s$, where $f(t)$ can be $f_c(t)$ or $f_o(t)$, applying Laplace transform to both sides of (3) and the inverse Laplace transform theorem ([18]), we obtain

$$f_u^*(s) = \int_0^\infty f_c(t)\left[\int_t^\infty f_o(\tau)d\tau\right]e^{-st}dt$$
$$+ \int_0^\infty f_o(t)\left[\int_t^\infty f_c(\tau)d\tau\right]e^{-st}dt$$
$$= \int_0^\infty \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} f_c^*(z)e^{zt}dz\left[\int_t^\infty f_o(\tau)d\tau\right]e^{-st}dt$$
$$+ \int_0^\infty f_o(t)\frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \frac{1 - f_c^*(z)}{z}e^{zt}dze^{-st}dt$$
$$= \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} f_c^*(z)\int_0^\infty \left[\int_t^\infty f_o(\tau)d\tau\right]e^{-(s-z)t}dtdz$$
$$+ \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \frac{1 - f_c^*(z)}{z}\int_0^\infty f_o(t)e^{-(s-z)t}dtdz$$
$$= \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} f_c^*(z)\frac{1 - f_o^*(s-z)}{s-z}dz$$
$$+ \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \frac{1 - f_c^*(z)}{z}f_o^*(s-z)dz$$
$$= \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \left[f_c^*(z)\frac{1 - f_o^*(s-z)}{s-z} + \frac{1 - f_c^*(z)}{z}f_o^*(s-z)\right]dz,$$
$$\tag{4}$$

where σ is the sufficiently small positive number appropriately chosen for the use of the inverse Laplace transforms ([18]). Let σ_o denote the set of poles of $f_o^*(-s)$. Since

$$\lim_{z\to\infty} \frac{1 - f_o^*(s-z)}{s-z} = -f_o^{*(1)}(0), \ \lim_{z\to 0} \frac{1 - f_c^*(z)}{z} = -f_c^{*(1)}(0),$$

we conclude that $z = s$ and $z = 0$ are removable singular points of the integrand of the last equation of (4). Hence, the set of poles of the integrand in the right complex plane will be the same as the one of $f_o^*(s)$. If $f_o^*(s)$ does not have branch points and only has isolated poles, which is the case when it is a rational function, then the Residue Theorem ([18]) can be applied; we can obtain (using a contour in the right half of the complex plan)

$$f_u^*(s) =$$
$$-\sum_{p\in\sigma_o}\text{Res}_{z=s+p}\left[f_c^*(z)\frac{1 - f_o^*(s-z)}{s-z} + \frac{1 - f_c^*(z)}{z}f_o^*(s-z)\right].$$
$$\tag{5}$$

In particular, if the calls originating from the mobile form a Poisson process, i.e., t_o is exponentially distributed with mean $1/\lambda_o$, then $f_o^*(s) = \lambda_o/(s+\lambda_o)$, from (5), we obtain

$$f_u^*(s) = -\sum_{p\in\sigma_o}\text{Res}_{z=s+p}$$
$$\left\{f_c^*(z)[1 - (\lambda_o/(s - z + \lambda_o))] + \frac{1 - f_c^*(z)}{z}[\lambda_o/(s - z + \lambda_o)]\right\}$$
$$= \text{Res}_{z=s+\lambda_o}\text{Res}_{z=s+p}$$
$$\left\{f_c^*(z)[1/(z - (s + \lambda_o))] + \frac{1 - f_c^*(z)}{z}[\lambda_o/(z - (s + \lambda_o))]\right\}$$
$$= \frac{\lambda_o}{s + \lambda_o} + \frac{s}{s + \lambda_o}f_c^*(s + \lambda_o).$$
$$\tag{6}$$

Furthermore, if the LA residence time is also exponentially distributed with parameter λ_c, then we can obtain a simpler form

$$f_u^*(s) = \frac{\lambda_o + \lambda_c}{s + \lambda_o + \lambda_c}, \tag{7}$$

which is the Laplace transform of the exponential distribution with parameter $\lambda_o + \lambda_c$. This result is obtained in [14].

If we exchange the roles of $f_o^*(s)$ and $f_c^*(s)$, we can obtain a similar result. Let σ_c denote the set of poles of the function $f_c^*(-s)$. Summarizing the above, we obtain Theorem 1.

Theorem 1. *The random variable t_u can be characterized as follows:*

$$f_u^*(s) =$$
$$\frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\left[f_c^*(z)\frac{1 - f_o^*(s-z)}{s-z} + \frac{1 - f_c^*(z)}{z}f_o^*(s-z)\right]dz$$
$$= \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\left[f_o^*(z)\frac{1 - f_c^*(s-z)}{s-z} + \frac{1 - f_o^*(z)}{z}f_c^*(s-z)\right]dz.$$
$$\tag{8}$$

If $f_c^*(s)$ and $f_o^*(s)$ are rational functions, then

$$f_u^*(s) = -\sum_{p\in\sigma_o}\text{Res}_{z=s+p}$$
$$\left[f_c^*(z)\frac{1 - f_o^*(s-z)}{s-z} + \frac{1 - f_c^*(z)}{z}f_o^*(s-z)\right]$$
$$= -\sum_{p\in\sigma_c}\text{Res}_{z=s+p}\left[f_o^*(z)\frac{1 - f_c^*(s-z)}{s-z} + \frac{1 - f_o^*(z)}{z}f_c^*(s-z)\right].$$
$$\tag{9}$$

Applying (5) and using the inverse Laplace transform ([18]), we could obtain the density function $f_u(t)$. Using the density function $f_p(t)$, we obtain the density function $f_r(t)$ of t_r (using the similar probability argument as for t_u):

$$f_r(t); = f_u(t)\int_t^\infty f_p(\tau)d\tau + f_p(t)\int_t^\infty f_u(\tau)d\tau$$
$$= \frac{1}{T_p}\int_{T_p}^\infty f_u(\tau)d\tau$$
$$+ \frac{1}{T_p}\int_t^{T_p} f_u(\tau)d\tau + f_u(t)\left(1 - \frac{t}{T_p}\right) (0 \le t \le T_p).$$
$$\tag{10}$$

It is obvious now that all computation in finding the probability distribution of the failure recovery time can be reduced to the computation of the probability distribution $f_u(t)$ in Theorem 1. If the Laplace transforms of the residual

LA residence time t_c and the residual life of the originating calls (t_o) are rational functions, then $f_u^*(s)$ can be easily computed by (5) via partial fractional expansion. It has been known that probability distribution functions of any nonnegative random variables can be appropriately approximated by either the Cox models ([17]), or the SOHYP models ([21], [22]), or the hyper-Erlang models ([8]), which all have rational Laplace transforms, therefore the random variables t_o and t_c can be directly modeled by one of the three types of aforementioned models based on the field data, then our analytical results can be used. Among the three types of models, the hyper-Erlang model is the simplest, which has also intuitive interpretation of its universal approximation capability ([8]). We will use the hyper-Erlang model to illustrate the computation.

The *hyper-Erlang* distribution has the following density function and Laplace transform:

$$f_{he}(t) = \sum_{i=1}^{M} \alpha_i \frac{(m_i\eta_i)^{m_i} t^{m_i-1}}{(m_i-1)!} e^{-m_i\eta_i t} \ (t \ge 0),$$

$$f_{he}^*(s) = \sum_{i=1}^{M} \alpha_i \left(\frac{m_i\eta_i}{s+m_i\eta_i}\right)^{m_i}, \tag{11}$$

where

$$\alpha_i \ge 0, \ \sum_{i=1}^{M} \alpha_i = 1,$$

and $M, m_1, m_2, \ldots, m_M$ are nonnegative integers, $\eta_1, \eta_2, \ldots, \eta_M$ are positive numbers. Hyper-Erlang distribution models contains the exponential distribution, Erlang distribution, the hyper-exponential distribution.

From Theorem 1, if t_o is hyper-Erlang distributed with distribution in (11), then we have

$$f_u^*(s) = \frac{1}{2\pi j} \int_{\sigma+s-j\infty}^{\sigma+s+j\infty}$$

$$\left[f_c^*(s+w)\frac{f_o^*(-w)-1}{w} + \frac{1-f_c^*(s+w)}{s+w}f_o^*(-w) \right] dw$$

$$= \frac{1}{2\pi j} \int_{\sigma+s-j\infty}^{\sigma+s+j\infty} \left[\frac{f_c^*(s+w)}{w} + \frac{1-f_c^*(s+w)}{s+w} \right] f_o^*(-w) dw$$

$$= \sum_{i=1}^{M} \alpha_i \frac{1}{2\pi j} \int_{\sigma+s-j\infty}^{\sigma+s+j\infty} \left[\frac{f_c^*(s+w)}{w} + \frac{1-f_c^*(s+w)}{s+w} \right]$$

$$\left(\frac{m_i\eta_i}{-w+m_i\eta_i} \right)^{m_i} dw$$

$$= -\sum_{i=1}^{M} \alpha_i(-1)^{m_i} \mathrm{Res}_{w=m_i\eta_i} \left[\frac{f_c^*(s+w)}{w} + \frac{1-f_c^*(s+w)}{s+w} \right]$$

$$\left(\frac{m_i\eta_i}{w-m_i\eta_i} \right)^{m_i}$$

$$= -\sum_{i=1}^{M} \alpha_i \frac{(-1)^{m_i}(m_i\eta_i)^{m_i}}{(m_i-1)!} \frac{\partial^{m_i-1}g(w;s)}{\partial w^{m_i-1}}\Big|_{w=m_i\eta_i}, \tag{12}$$

where

$$g(w;s) = \frac{f_c^*(s+w)}{w} + \frac{1-f_c^*(s+w)}{s+w}.$$

In the above derivation, we have used the fact that

$$\int_{\sigma+s-j\infty}^{\sigma+s+j\infty} h(w)dw = 0$$

for proper rational function $h(w)$ analytic in the right complex plane. Thus, we obtain Corollary 1.

Corollary 1. *If $f_c^*(s)$ is a proper rational function and $f_o^*(s) = f_{he}^*(s)$ as in (11), then we have*

$$f_u^*(s) = -\sum_{i=1}^{M} \alpha_i \frac{(-1)^{m_i}(m_i\eta_i)^{m_i}}{(m_i-1)!} \frac{\partial^{m_i-1}g(w;s)}{\partial w^{m_i-1}}\Big|_{w=m_i\eta_i}, \tag{13}$$

where

$$g(w;s) = \frac{f_c^*(s+w)}{w} + \frac{1-f_c^*(s+w)}{s+w}.$$

In practice, the field data may be collected for the random variables t_O, t_C, and t_A, respectively. Since t_o, t_c and t_a are residual life times of t_O, t_C, and t_A, respectively, we can draw the distributions of t_o, t_c, and t_a by applying the Residual Life Theorem ([17]) as follows:

$$f_o^*(t) = \frac{1-f_O^*(s)}{T_O s}, \ f_c^*(t) = \frac{1-f_C^*(s)}{T_C s}, \ f_a^*(t) = \frac{1-f_A^*(s)}{T_A s},$$

where $f_O^*(s)$, $f_C^*(s)$, and $f_A^*(s)$ are the Laplace transform of the probability density functions of the random variables t_O, t_C and t_A, respectively, with T_O, T_C, and T_A to be their mean values, respectively. If we use the hyper-Erlang models to approximate the distributions of t_O and t_C, then we can also obtain simple analytical result as in Corollary 1.

3 AVERAGE NUMBER OF LOST CALLS

In order to perform the cost analysis, we need to compute the average number of lost calls because of the database failure. We assume that the call arrivals to the mobile form a Poisson process with arrival rate λ_a. The probability that $K = k$ call arrivals occur in a known period X is given by

$$\Pr(K = k \mid X = t) = \frac{(\lambda_a t)^k}{k!} e^{-\lambda_a t},$$

where K is the number of call arrivals. The joint density function of (K, X) is given by

$$f_L(k,t) = \Pr(K = k \mid X = t)f_r(t) = \frac{(\lambda_a t)^k}{k!} e^{-\lambda_a t} f_r(t).$$

Let $p_L(k)$ denote the probability that k calls to the mobile arrive between an HLR database failure and the failure recovery instant, then it is given by

$$p_L(k) = \int_0^\infty f_L(k,t)dt = \frac{\lambda_a^k}{k!} \int_0^\infty t^k f_r(t)e^{-\lambda_a t}dt$$

$$= \frac{(-\lambda_a)^k}{k!} f_r^{*(k)}(\lambda_a) = \sum_{k=0}^{\infty} a_k(-\lambda_a)^k, \tag{14}$$

where $a_k = f^{*(k)}(\lambda_a)/k!$ is the coefficient of the Taylor expansion of the Laplace transform $f_r^*(s)$ at $s = \lambda_a$:
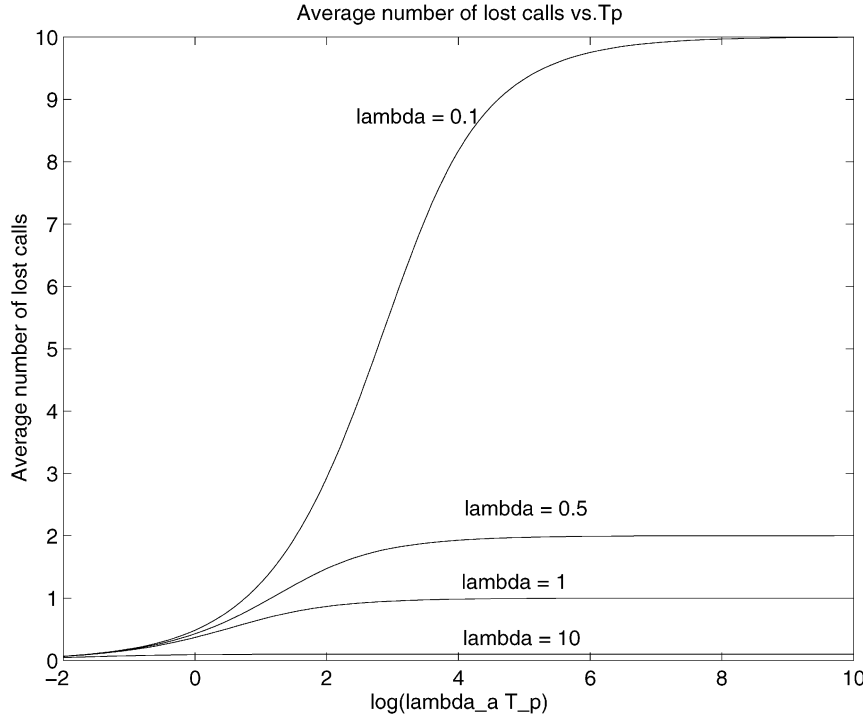
Fig. 4. Average number of lost calls vs. location update interval ($\lambda_a T_p$).

$$f_r^*(s) = \sum_{i=0}^{\infty} \frac{f^{*(i)}(\lambda_a)}{i!}(s - \lambda_a)^i = \sum_{i=0}^{\infty} a_i(s - \lambda_a)^i. \qquad (15)$$

Let E_{loss} denote the average number of calls arriving during between the HLR database failure and the failure recovery instant, i.e., the average number of arriving calls lost. Then we have

$$
\begin{aligned}
E_{loss} &= \sum_{i=0}^{\infty} i \cdot p_L(i) = \sum_{i=1}^{\infty} i \cdot p_L(i) \\
&= (-\lambda_a) \sum_{i=1}^{\infty} \frac{(-\lambda_a)^{i-1}}{(i-1)!} \frac{d^{i-1}}{ds^{i-1}} f_r^{*(1)}(s)\Big|_{s=\lambda_a} \\
&= (-\lambda_a) \sum_{i=0}^{\infty} \frac{(-\lambda_a)^i}{i!} g^{(i)}(\lambda_a) \; (g(s) = f_r^{*(1)}(s)) \\
&= (-\lambda_a) \sum_{i=0}^{\infty} \frac{g^{(i)}(\lambda_a)}{i!}(-\lambda_a)^i \\
&= (-\lambda_a)g(0) = -\lambda_a f_r^{*(1)}(0) = \lambda_a E[t_r].
\end{aligned}
\qquad (16)
$$

From (10) and (16), we obtain Theorem 2.

Theorem 2. *The average number of lost calls is given by*

$$E_{loss} =$$

$$\frac{\lambda_a T_p}{2} \int_{T_p}^{\infty} f_u(t)dt + \lambda_a \int_0^{T_p} t f_u(t)dt - \frac{\lambda_a}{2T_p} \int_0^{T_p} t^2 f_u(t)dt. \qquad (17)$$

In the last section, we have given the computation of $f_u(t)$ (i.e., $f_u^*(s)$), thus we can easily compute the average number of lost calls.

If both t_o and t_c are exponentially distributed, from the last section, we know that t_u is exponentially distributed

with parameter $\lambda_u = \lambda_o + \lambda_c$. Hence, we obtain the following close-form formula:

$$E_{loss} = \frac{\lambda_a}{\lambda_u}\left[1 - \frac{1 - e^{-\lambda_u T_p}}{\lambda_u T_p}\right]. \qquad (18)$$

In [14], Haas and Lin only provided a procedure to compute E_{loss} numerically. Here, we provide an analytical formula clearly showing the relationship between the average number of lost calls and different parameters (λ_u, λ_a and T_p). Let $\lambda = \lambda_u/\lambda_a$, $x = \lambda_a T_p$ (we use λ_a to normalize some quantities). We obtain

$$E_{loss} = \frac{1}{\lambda}\left[1 - \frac{1 - e^{-\lambda x}}{\lambda x}\right].$$

Figs. 4 and 5 show the plots of the average numbers of lost calls with respect to the location update interval and the "natural" registration rate λ ($= \lambda_u/\lambda_a$). Obviously, as λ increases, i.e., the registration rate (either due to the call origination or due to the LA crossing) is higher, the average number of lost calls decreases, which is intuitively true; as the location update interval decreases, the average number of lost calls decreases. These observations are consistent with the results obtained in [14], in which a different computational method is used.

4 COST ANALYSIS

As we mentioned earlier, there is a trade-off between the cost of signaling and the cost of lost call connections. In this section, we conduct the cost analysis to determine the optimal location update interval T_p.

Let c_u denote the cost of a single location update and let c_n denote the cost of losing n calls ($n = 1, 2, \ldots$). Assume that the average interfailure time (the average time between
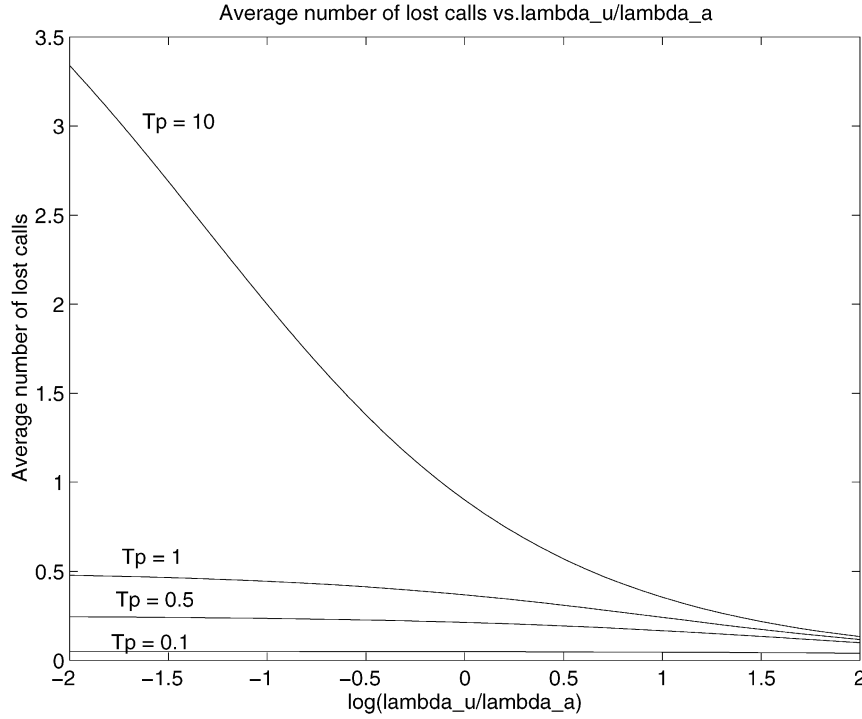
Fig. 5. Average number of lost calls vs. mobile registration rate (λ_u/λ_a).

HLR failures) is T_f. Let C_{total} denote the total cost per unit time of location updates and lost calls. The total cost can be expressed as

$$C_{total} = \frac{1}{T_f}\sum_{n=1}^{\infty} nc_n p_L(n) + \frac{1}{T_p} c_u, \qquad (19)$$

where $p_L(n)$ can be found using the Taylor expansion at $s = \lambda_a$ as in (14) and (15). The above cost function is reasonable. If the signaling traffic is more expensive than the call losses, then we can choose larger unit cost c_u. Hence, in order to minimize the overall cost, T_p should be greater, i.e., the less location update should be applied. If call losses are more expensive (i.e., in the business area), we can increase the unit cost c_n to lower the call loss probability $p_L(n)$, or reduce T_p does not change the overall cost too much.

If c_n is a constant, let $c_n = c_1$, then the total cost becomes

$$C_{total} = \frac{c_1}{T_f}\sum_{n=1}^{\infty} np_L(n) + \frac{1}{T_p} c_u = \frac{c_1}{T_f}E_{loss} + \frac{c_u}{T_p}. \qquad (20)$$

In particular, when t_o and t_c are both exponentially distributed, we have the following analytical formula

$$C_{total} = \frac{\lambda_a c_1}{\lambda_u T_f}\left[1 - \frac{1 - e^{-\lambda_u T_p}}{\lambda_u T_p}\right] + \frac{c_u}{T_p}. \qquad (21)$$

If we can normalize the cost by c_u, let $c = c_1/c_u$ and the total cost be denoted as $C(T_p)$, then we obtain Theorem 3.

Theorem 3. *The total cost is given by*

$$C(T_p) = \frac{\lambda_a c}{\lambda_u T_f}\left[1 - \frac{1 - e^{-\lambda_u T_p}}{\lambda_u T_p}\right] + \frac{1}{T_p}. \qquad (22)$$

Our objective now is to minimize this cost with respect to the location update interval T_p. Obviously, $C(0) = +\infty$, which means that it is not worthy to update too often; $C(+\infty) = \lambda_a c/\lambda_u T_f$, which implies that when there is no location update for the scheme, the total cost is fixed. There must be an optimal value T_p which minimizes the total cost. Differentiating the total cost $C(T_p)$ with respect to T_p, and set it to zero, we obtain the stationary equation

$$1 - e^{-\lambda_u T_p} - \lambda_u T_p e^{-\lambda_u T_p} = \frac{\lambda_u^2 T_f}{c\lambda_a}. \qquad (23)$$

Let $a = \lambda_u^2 T_f/(c\lambda_a)$ and $y = \lambda_u T_p$. We can use any numerical algorithm to find the solution of the following equation

$$1 - e^{-y} - ye^{-y} = a. \qquad (24)$$

Let y^* denote the solution of equation (24), then the optimal location update interval is given by

$$T_p^* = \frac{y^*}{\lambda_u}.$$

Therefore, we obtain the analytical result for the location update interval.

Remark. Equation (24) may not have a solution for some choice of parameters. This implies that there is no stationary point for the cost function, so period location update is not necessary for the failure recovery. This may be true when there are many calls originating calls (λ_a is large) or the mobile user travels fast (λ_c is large).

One important observation is that the optimal value T_p^* is related to the traffic parameters λ_o and λ_a, the failure rate T_f and the mobility parameter λ_c. From this scheme, we can adaptively adjust the location update interval T_p according to the real situation in the
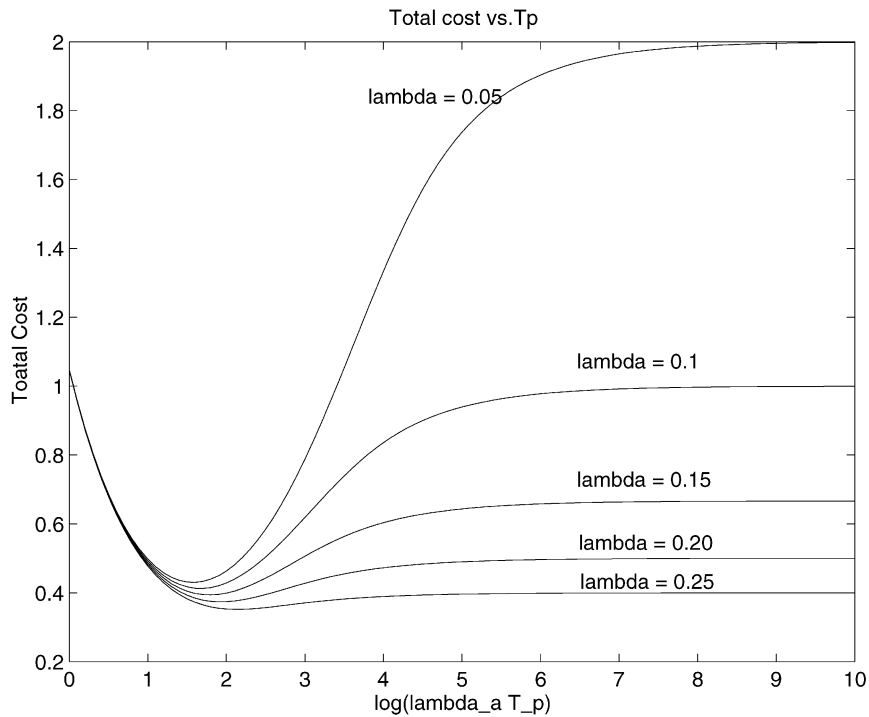
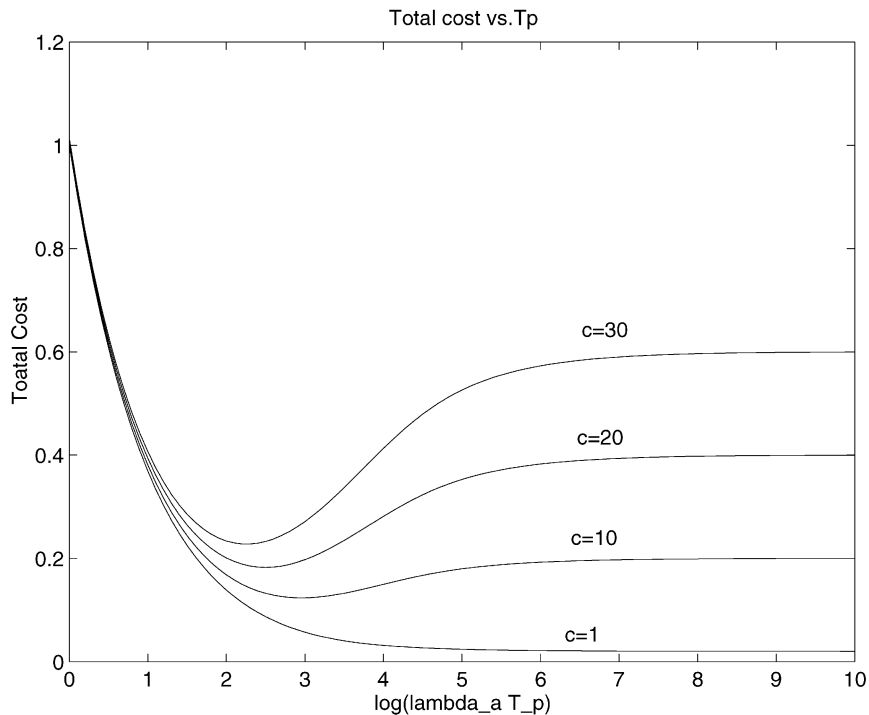Fig. 6. Total cost vs. location update interval: varying traffic factor ($\lambda_u \lambda_a$).

Fig. 7. Total cost vs. location update interval: varying the unit cost (c).

PCS networks to improve the performance of the failure recovery scheme via the periodic location update.

5   DISCUSSIONS

In this section, we conduct the cost analysis using our analytical results for the system with a specific set of parameters.

We first consider a special case to show how to find the optimal location update interval. Let $\lambda_a = 1$, $\lambda_u = 1$, $c = 1,000$, $T_f = 1,000$, then $a = \lambda_u^2 T_f/(c\lambda_a) = 1$. There is no solution for (24), which implies that it is not necessary to do the location update for the failure recovery. In fact, it is true because the failure rate is very low ($1/T_f = 0.001$). If the $T_f = 100$, then $a = 0.1$, solving (24), we obtain $y^* = 0.535$,

hence, the optimal location update interval is $T_p^* = y^*/\lambda_u = 0.535$.

Next, we want to show how the cost is affected by the λ_u/λ_a and $\lambda_a T_p$.

Applying Theorem 3 with the aforementioned parameters, we can obtain the cost curves. Figs. 6 and 7 show the total cost vs. the location update interval varying traffic factor and the unit cost, respectively. In Fig. 6, $c = 10$ and $T_f = 100$; it shows that there exists an optimal location update interval for each traffic situation. As the natural registration rate (i.e., the registrations resulting from the originating call or the LA boundary crossing) increases, the optimal location update interval increases. This is as expected. Fig. 7 shows that when the natural registration cost and the location update cost is close enough (i.e., $c = 1$), it is not necessary to do the location update. If $c > 1$, optimal location update interval exists; If $c < 1$, it is not necessary to do location update. Thus, the set of curves like in the figures provide all the information for optimal choice of location update for failure recovery in PCS networks.

As a final remark, we can apply our result to design a dynamic location update scheme, or the "run-time" scheme. The basic idea is to collect all statistics to find the arrival rates λ_a, λ_c and λ_o, hence, λ_u, then find the optimal T_p^* which minimizes the cost in Theorem 3 and, subsequently, carry out the location update according to this location update interval. We can implement this scheme epoch by epoch, or we can change the location update interval only when there are significant changes in some of the above parameters.

6 CONCLUSION

In this paper, we discuss the active recovery procedure of HLR failure in the EIA/TIA IS-41 system. We presented a general formula to calculate probability density distribution of the recovery time of the system and the average number of lost calls due to a failure of mobility database. Assuming that the interarrival time of the incoming calls has an exponential distribution, we obtain an analytical result of the average number of lost calls during the failure recovery time of the system, from which the cost analysis can be conducted analytically. An optimization problem is analytically solved to give the optimal location update interval for failure recovery of mobility databases in PCS networks.

REFERENCES

[1] A. Abutaleb and V.O.K. Li, "Paging Strategy Optimization in Personal Communication Systems," *Wireless Networks,* vol. 3, pp. 195–204, 1997.

[2] I.F. Akyildiz and J.S.M. Ho, "Dynamic Mobile User Location Update for Wireless PCS Networks," *Wireless Networks,* vol. 1, pp. 187–196, 1995.

[3] I.F. Akyildiz, J.S.M. Ho, and Y-B. Lin, "Movement-Based Location Update and Selective Paging for the PCS Networks," *IEEE/ACM Trans. Networking,* vol. 4, pp. 629–638, 1996

[4] A. Bar-Noy, I. Kessler, and M. Sidi, "Mobile Users: To Update or Not to Update?" *Wireless Networks,* vol. 1, pp. 175–186, 1995.

[5] D.C. Cox, "Wireless Personal Communications: What Is It?" *IEEE Personal Comm. Magazine,* pp. 20–35, Apr. 1995.

[6] EIA/TIA, *Cellular Radio-Telecommunication Intersystem Operation: Automatic Roaming,* Technical Report IS-41.3., BEIA/TIA, 1991.

[7] ETSI/TC, *Restoration Procedures,* version 4.2.0, Technical Report Recommendation GSM 03.07, ETSI, 1993.

[8] Y. Fang and I. Chlamtac, "A New Mobility Model and Its Application in the Channel Holding Time Characterization in PCS Networks," *Proc. INFOCOM' 99,* Mar. 1999.

[9] Y. Fang, I. Chlamtac, and Y. B. Lin, "Channel Occupancy Times and Handoff Rate for Mobile Computing and PCS Networks," *IEEE Trans. Computers,* June 1998.

[10] Y. Fang, I. Chlamtac, and Y.B. Lin, "Modeling PCS Networks Under General Call Holding Times and Cell Residence Time Distributions," *IEEE Trans. Networking,* vol. 5, no. 6, pp. 893–906, Dec. 1997.

[11] Y. Fang, I. Chlamtac, and Y.B. Lin, "Call Performance for a PCS Network," *IEEE J. Selected Area in Comm.,* vol. 15, no. 8, pp. 1,568–1,581, Oct. 1997.

[12] V.K. Garg and J.E. Wilkes, *Wireless and Personal Comm. Systems.* Upper Saddle River, New Jersey: Prentice-Hall PTR, 1996.

[13] R.A. Guerin, "Channel Occupancy Time Distribution in a Cellular Radio System," *IEEE Trans. Vehicle Technology,* vol. 35, no. 3, pp. 89–99, 1987.

[14] Z.J. Haas and Y.B. Lin, "On the Optimizing the Location Update Costs in the Presence of Database Failures," *Wireless Networks,* 1998.

[15] D. Hong and S.S. Rappaport, "Traffic Model and Performance Analysis for Cellular Mobile Radio Telephone Systems with Prioritized and Nonprioritized Handoff Procedures," *IEEE Trans. Vehicle Technology,* vol. 35, no. 3, pp. 77–92, 1986.

[16] C. Jedrzycki and V.C.M. Leung, "Probability Distributions of Channel Holding Time in Cellular Telephony Systems," *Proc. IEEE Vehicle Technology Conf.,* pp. 247–251, May 1996.

[17] L. Kleinrock, *Queueing Systems: Theory, Volume I.* New York: John Wiley & Sons, 1975.

[18] W.R. LePage, *Complex Variables and the Laplace Transform for Engineers.* New York: Dover Publications, Inc., 1980.

[19] Y.B. Lin, "Database Failure Recovery for Cellular Phone Networks," Submitted for publication.

[20] Y.B. Lin, "Failure Restoration of Mobility Databases for Personal Comm. Networks," *Wireless Networks,* vol. 1, no. 3, pp. 365–372, 1995.

[21] P. Orlik and S.S. Rappaport, "A Model for Teletraffic Performance and Channel Holding Time Characterization in Wireless Cellular Communication," *Proc. Int'l Conf. Universal Personal Comm. (ICUPC' 97),* pp. 671–675, Oct. 1997.

[22] P. Orlik and S.S. Rappaport, "Traffic Performance and Mobility Modeling of Cellular Communications with Mixed Platforms and Highly Variable Mobilities," Technical Report no. 727, College of Engineering and Applied Sciences, State Univ. of New York at Stony Brook. Also to appear in *Proc. IEEE.*

[23] T.P. Wang, C.C. Tseng, and W.K. Chou, "An Aggressive Approach to Failure Restoration of PCS Mobility Databases," *ACM Mobile Computing and Comm. Review,* vol. 1, no. 3 pp. 21–28, 1997.
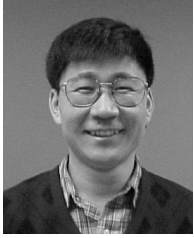
Yuguang Fang (S'96-M'98-SM'99) received the BS and MS degrees in mathematics from Qufu Normal University, Qufu, Shandong, China, in 1984 and 1987, respectively, the PhD degree in systems and control engineering from Department of Systems, Control and Industrial Engineering at Case Western Reserve University, Cleveland, OH, in January 1994, and the PhD degree in electrical engineering from Department of Electrical and Computer Engineering at Boston University, MA, in May 1997.

From 1987 to 1988, he held research and teaching positions in both the Department of Mathematics and the Institute of Automation at Qufu Normal University. From September 1989 to December 1993, he was a teaching/research assistant in the Department of Systems, Control and Industrial Engineering at Case Western Reserve University, where he also held a research associate position from January 1994 to May 1994. He held a postdoctoral position in the Department of Electrical and Computer Engineering at Boston University from June 1994 to August 1995. From September 1995 to May 1997, he was a research assistant in the Department of Electrical and Computer Engineering at Boston University. From June 1997 to July 1998, he was a visiting assistant professor in the Department of Electrical Engineering at the University of Texas at Dallas. From July 1998 to May 2000, he was an assistant professor in the Department of Electrical and Computer Engineering at New Jersey Institute of Technology, Newark, NJ. Since June 2000, he has been an assistant professor in the Department of Electrical and Computer Engineering at the University of Florida, Gainesville, FL. He has published more than 40 papers in professional journals and refereed conferences. He is an editor for *IEEE Transactions on Communication*, an editor for *IEEE Journal on Selected Areas in Communications: Wireless Communications Series*, and an area editor for *ACM Mobile Computing and Communications Review*. His recent research interests include wireless networks and mobile communications, personal communication services (PCS), call admission control, and resource allocations. He is a senior member of the IEEE.

Dr. Fang is the program vice-chair for the 2000 IEEE Wireless Communications and Networking Conference (WCNC '2000), the program vice-chair for the Seventh International Conference on Computer Communications and Networking (ICN '98), a member of the technical program committee for the International Conference on Computer Communications (INFOCOM '98 and INFOCOM '00), a member of the technical program committee for the First International Workshop for Wireless and Mobile Multimedia (WOW-MoM '98), and a member of the technical program committee for the 1999 IEEE Wireless Communication and Networking Conference (WCNC '99).

Imrich Chlamtac (M'86-SM'86-F'93) received the PhD degree in computer science from the University of Minnesota in 1979. Since 1997, he holds the Distinguished Chair in Telecommunications at the University of Texas at Dallas and is the director of CATSS, the Center for Advanced Telecommunications Systems and Services. Prior to joining UTD, Dr. Chlamtac was a professor of electrical engineering, a member of the Photonic Center at Boston University, and President of BCN Inc., a company dealing with network design. Dr. Chlamtac holds the honorary titles of senior professor at Tel Aviv University, and of university professor at the Technical University of Budapest. Dr. Chlamtac is a fellow of the IEEE and ACM societies. He has published more than two hundred and fifty papers in refereed journals and conferences, chapters in books and encyclopedias, as well as the first textbook on LANs in 1981, and delivered dozens of distinguished lectures, plenary, and keynote addresses. Dr. Chlamtac serves as the founding editor-in-chief of the *ACM/URSI/Baltzer Wireless Networks (WINET)*, *Mobile Networks and Applications (MONET)* journals, and the new *SPIE/Baltzer Optical Networks Magazine*. In the past, he was on the editorial boards of most major publications in telecommunications. Dr. Chlamtac was the general chair of leading the ACM and IEEE conferences and workshops, including ACM Sigcomm, ACM/IEEE MobiCom and IEEE CCW. Dr. Chlamtac is the founder and steering committee chair of the ACM/IEEE MobiCom and the founding chairman of ACM Sigmobile.

Hong-Bing Fei received the BS degree in physics from Nanjing University in 1989, the MS degree in physics from the Institute of Theoretical Physics, Chinese Academy of Science in 1994, and the MS degree in electrical engineering from Boston University in 1997. He is currently pursuing his PhD degree in electrical engineering at the University of Texas at Dallas. He has published several papers in professional journals and refereed conferences, which cover the areas of both physics and engineering.