Failure Recovery of HLR Mobility Databases and Parameter Optimization for PCS Networks

Yugang Fang

Department of Electrical and Computer Engineering, New Jersey Institute of Technology, University Heights, Newark, New Jersey 07102-1982
E-mail: fang@njit.edu

and

Imrich Chlamtac and Hong-Bing Fei

Erik Jonsson School of Engineering and Computer Science, University of Texas at Dallas, Mail Station EC33, P.O. Box 830688, Richardson, Texas 75083
E-mail: chlamtac@utdallas.edu and hbfei@utdallas.edu

Received January 15, 1999; revised August 15, 1999; accepted December 15, 1999

In PCS networks, new calls to a portable may be lost due to incorrect location information (due to either data corruption or outdated records) in the home location register mobility database. Such database failures can be recovered through the portable’s registration when it initiates a call, through required registration (or deregistration for some systems) as it crosses the location area boundary, or through active location updates. In this paper we analyze the active location update scheme and failure restoration under several different mobile traffic distributions, propose a framework for the performance evaluation of this scheme, and present general analytical results for cost analysis. Through a systematic study, we show that there always exists an optimal choice of location update period to minimize the total cost, one that represents the tradeoffs between location update frequency and call losses. Our results can be used to provide design guidelines for active failure restoration of mobility databases in PCS networks.

Key Words: location update; failure recovery; HLR; PCS; mobility database; mobility management.

1 This work was supported in part by the U.S. Army Research Office under Contract DAAG55-97-1-0312 and DAAG55-97-1-0382. The work of Yugang Fang was also supported, in part, by the New Jersey Institute of Technology under Grant SBR421980 and New Jersey Center for Wireless Telecommunications.

2 To whom correspondence should be addressed.
1. INTRODUCTION

In the EIA/TIA IS-41 standard [7], user location strategies use two-level hierarchical registration schemes. In a PCS network, the home location register (HLR) is the location register that maintains the mobile's identity information containing mobile user 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 the mobile. The visitor location register (VLR) is the location register other than the HLR used to retrieve information for the handling of calls from or to the mobile when it visits another area or system (i.e., when it is roaming). The VLR is the database associated with the PCS network that the mobile user is currently visiting.

When the mobile is in its home system (i.e., the system in which the mobile subscribes to its service), the mobile's location information can be directly accessed from the HLR in the home system. When a call to the mobile is requested, the mobile switching center (MSC) can retrieve the location information from the HLR and direct the call to the mobile. If the mobile moves to 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 identity information is created from the HLR and is stored in the VLR in the visited system and the home MSC updates its current location in its HLR.

To originate a call, the mobile contacts the visited MSC in the network. The call request is processed using the information in VLR and call connection can eventually be established. If a call is to the mobile user, the call is directed to the originating MSC, and then a contact with the home MSC of the mobile is made with the help of its HLR. The location information in the HLR database is used to locate the callee mobile user; if found, the call will be forwarded to the visited MSC where the callee mobile is currently visiting. Thus, PCS mobility databases, especially HLRs and VLRs, are modified and queried frequently for location tracking and call delivery.

Due to such constant changes of location information for a mobile in the HLR, the location information may be corrupted or obsolete, if the mobile does not register often. These conditions lead to HLR database failure, and calls or call connections arriving to the mobile before recovery will be lost, resulting in a blocked call. Service providers have to minimize this blockage to ensure normal system operation. The HLR failure can be recovered either when the mobile initiates a call in a cell, thus revealing the current location of the mobile in the cellular network, or when the mobile crosses a location area (LA) boundary, where a registration/deregistration process is required. However, there is a serious problem with these recovery schemes: if the mobile rarely initiates calls and/or moves very slowly, the location information will become obsolete after some time; although a paging scheme can be used in this case, the time for locating a mobile [2] may be so long that the call to the mobile is lost before the HLR failure is recovered. In order to overcome this problem, an autonomous registration mechanism was proposed [8], in which a mobile periodically reregisters with the system, when used for failure recovery, this scheme is also termed active failure restoration. In this approach a mobile periodically establishes radio contact with the network to confirm its location. These location updates are registrations that update the mobile's current
location information in its HLR. It has been shown that in this way the HLR restoration delay is reduced. However, the method for choosing the location update interval is open to the system designers.

There are some differences in mobility database recovery for the GSM system \[8\] and the EIA/TIA IS-41 system \[7\]. In GSM, the HLR or VLRs are backed up periodically. After a failure, the HLR or VLRs are restored from the nonvolatile storage. In IS-41, the HLR and VLR are not backed up in nonvolatile storage, so after a failure the HLR must incrementally reconstruct the mobile's location records. Several database recovery schemes were recently proposed and analyzed. Lin \[19, 20\] modeled the HLR and VLR restoration with and without checkpointing in IS-41 and GSM. 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 and analyzed its performance.

The failure restoration is directly related to system costs. More frequent location updates will increase the signaling traffic, while less frequent updates will result in more lost calls and lower customer satisfaction. There is a trade-off between these two; consequently an overall cost analysis is needed. The total cost for failure restoration consists of two components: the cost of lost calls and the location update cost. 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. Their work assumes that the LA residence time and the interarrival time of the initiating calls are all exponentially distributed and cost analysis is done mostly via simulations. For system control and evaluation purposes it is highly advantageous to obtain analytical models. For applicability to actual systems it is furthermore necessary to analyze these networks under general, nonexponential, assumptions on the LA residence time and the interarrival time of the new initiating calls, and relate the location update interval to the traffic parameters.

In this paper, we present analytical results for the performance evaluation of the above active location updating scheme for the mobility database failure recovery of PCS networks under more general realistic assumptions. For some specific cases, we give analytic formulas relating the location update interval to the system parameters. The results show that there exist optimal values of location update period, which can minimize the total cost. These results can be used to adaptively change the location update interval according to the traffic and mobility conditions. Although our problem is motivated from second generation systems (IS-41 or GSM), our solution does not depend on specific systems and will be applicable for future wireless cellular networks. We expect that the results will play a significant role in the active failure restoration of mobility databases in PCS networks.

2. SYSTEM PARAMETERS RELATED TO THE HLR MOBILITY DATABASE FAILURE RESTORATION IN PCS NETWORKS

In this section, we present the system parameters and notations used in the analysis of the HLR mobility database failure recovery.
Let $t_c$ denote the recovery time interval, that is, the time interval between the instants of the HLR failure and its recovery. Let $t_o$ be the time interval between the instant of HLR failure and the instant that the mobile crosses the LA boundary (which we will term the *residual residence time*), let $t_p$ be the interval between the instant of failure and the location update. The recovery time interval $t_r$ is the minimum among the three values $t_c$, $t_o$, or $t_p$, or $t_r = \min\{t_c, t_o, t_p\}$. Let $f_r(t_r)$, $f_c(t_c)$, $f_o(t_o)$, and $f_p(t_p)$ denote the probability density functions of $t_r$, $t_c$, $t_o$, and $t_p$.

In what follows, we will assume that the time variables $t_c$, $t_o$, and $t_p$ are mutually independent. The following are additional system parameters we will use. $C_{tot}$ is the total cost of the location management including the recovery of HLR failure. $c_n$ is the cost of losing $n$ calls to the mobile. $c_r$ is the cost of a single location update. $\lambda_0$ is the arrival rate of calls originated by the mobile. $\lambda_c$ is the rate of LA boundary crossings by the mobile. $\lambda_o$ is the arrival rate of call arrivals to the mobile. $t_o$ is the time interval between two consecutive call originated from the mobile. $t_d$ is the time interval between two consecutive call arrivals to the mobile. $t_c$ (residence time) is the time interval between two LA crossings by the mobile. $T_p$ is the constant time interval between two periodic location updates of mobile. $T_f$ is the mean time between two HLR mobility database failures. $E_{loss}$ is the average number of lost calls.

Some of the time-related parameters discussed above are shown in Fig. 1.

### 3. GENERAL ANALYTICAL RESULTS

To calculate the number of lost incoming calls, we need to obtain the probability distribution of the recovery time $t_r$. In this section we present several analytical results for the probability density function $f_r(t_r)$ and its Laplace transform.

From the previous section, we know that $t_r = \min\{t_c, t_o, t_p\} = \min\{t_r, \min\{t_c, t_o, t_p\}\}$. To characterize $t_r$, we need to compute the density of $\min\{t_1, t_2\}$ for independent random variables $t_1$ and $t_2$. The following result serves this purpose.
Theorem 3.1 [11]. If \( t_m = \min(t_1, t_2) \), let the probability density functions of \( t_1 \) and \( t_2 \) be \( f_1(t_1) \) and \( f_2(t_2) \), then \( f_m(t_m) \), the probability density function of \( t_m \), is given by

\[
f_m(t_m) = f_1(t) \int_{l_z}^\infty f_2(\tau) \, d\tau + f_2(t) \int_{l_z}^\infty f_1(\tau) \, d\tau.
\]

Using Eq. (1) recursively, we can obtain \( f_1(t) \) from \( f_2(t) \), \( f_3(t) \), and \( f_4(t) \). Unfortunately, in many cases, the explicit form of \( f_1(t) \) is difficult to calculate. Instead, it is easier to compute the Laplace transform of \( f_1(t) \), which provides the same information. We will concentrate on the computation of the Laplace transform next. The following result is fundamental for the computation of the Laplace transform of \( f_1(t) \).

Theorem 3.2. The Laplace transform \( f_m^*(s) \) of \( f_m(t_m) \) is given as

\[
f_m^*(s) = f_1^*(s) + f_2^*(s) - \int_{l_z} \frac{dz}{2\pi j} \frac{f_2^*(s-z) f_1^*(z)}{z} - \int_{l_z} \frac{dz}{2\pi j} \frac{f_1^*(s-z) f_2^*(z)}{z},
\]

where \( f_1^*(s) \) and \( f_2^*(s) \) are the Laplace transforms of \( f_1(t_1) \) and \( f_2(t_2) \), respectively, and \( l_z \) is the integration path, \( l_z = \{ z | z = e + j\xi, e = 0^+, \xi: -\infty \rightarrow \infty \}. \)

Proof. The proof of this theorem can be found in Appendix 1.

Equation (2) can be simplified if both \( f_1^*(z) \) and \( f_2^*(z) \) do not have poles between \((0, \infty)\). Using the following variable substitution \( z \rightarrow s - z \) in the last term in Eq. (2), we obtain

\[
\int_{l_z} \frac{dz}{2\pi j} \frac{f_2^*(s-z) f_1^*(z)}{z} = \int_{l_z'} \frac{dz}{2\pi j} \frac{f_2^*(s-z) f_1^*(z)}{s-z},
\]

where \( l_z' \) is the following curve in the complex plane: \( l_z' = \{ z | z = (s-0^+) + j\xi, \xi: -\infty \rightarrow \infty \}. \)

Due to the facts that \( f_1(z) \rightarrow 0 \) and \( f_2(z) \rightarrow 0 \) when \( |z| \rightarrow \infty \) and that both \( f_1^*(z) \) and \( f_2^*(z) \) do not have poles between \((0, \infty)\), \( f_1^*(z) \) and \( f_2^*(z) \) are analytic within the range \((0, s)\) for any \( s \), Re \( s > 0 \), and we have

\[
\int_{l_z'} \frac{dz}{2\pi j} \frac{f_2^*(s-z) f_1^*(z)}{s-z} = \int_{l_z'} \frac{dz}{2\pi j} \frac{f_1^*(s-z) f_2^*(z)}{s-z}.
\]

From Eqs. (2) and (4), we obtain

\[
f_m^*(s) = f_1^*(s) + f_2^*(s) - \int_{l_z} \frac{dz}{2\pi j} \frac{f_2^*(s-z) f_1^*(z)}{z} - \int_{l_z} \frac{dz}{2\pi j} \frac{f_1^*(s-z) f_2^*(z)}{z} = f_1^*(s) + f_2^*(s) + \int_{l_z} \frac{dz}{2\pi j} f_1^*(s-z) f_2^*(z) \frac{s}{z(z-s)},
\]

(5)
From Eq. (5) we have the following useful result:

**Corollary 3.1.** If $f_\tau^*(z)$ and $f_\mathbb{M}^*(z)$ do not have any poles between $(0, \infty)$, then Eq. (2) has the form

$$f^*_M(s) = f^*_{\tau}(s) + f^*_{\mathbb{M}}(s) + \int_{1}^{s} \frac{dz}{2\pi i} f^*_{\tau}(s-z) f^*_M(z) \frac{s}{z(z-s)}. \tag{6}$$

Let $t_M = \min(t_1, t_2, t_3) = \min(t_o, t_e)$; using Theorem 3.2 twice, we obtain

**Theorem 3.3.** If $t_M = \min(t_1, t_2, t_3)$, then $f^*_M(s)$, the Laplace transform of $f_M(t_M)$, is given by

$$f^*_M(s) = f^*_{\tau}(s) + f^*_{\mathbb{M}}(s) + f^*_o(s)$$

$$= -\int_{0}^{s} \frac{dz}{2\pi i} \left[ f^*_{\mathbb{M}}(s-z) f^*_{\tau}(z) + f^*_{\tau}(s-z) f^*_{\mathbb{M}}(z) \right]$$

$$+ f^*_{\mathbb{M}}(s-z) f^*_{\tau}(z) + f^*_{\tau}(s-z) f^*_{\mathbb{M}}(z)$$

$$+ f^*_{\mathbb{M}}(z) f^*_{\tau}(s-z) f^*_{\tau}(z') + f^*_{\tau}(z) f^*_{\mathbb{M}}(s-z) f^*_{\tau}(z'). \tag{7}$$

### 4. SPECIFIC DISTRIBUTION MODELS OF PRACTICAL INTEREST

To calculate the average number of lost calls, we need to find $f(t_o)$, or its Laplace transform. The distribution of $t_o$ depends on the distribution of other time parameters: $t_e$, $t_c$, and $t_p$. In this section, we discuss some probability distribution models for these parameters. Let $f^*_{\tau}(s)$, $f^*_{\mathbb{M}}(s)$, and $f^*_{\tau}(s)$ denote the Laplace transforms of $f_\tau(t_o)$, $f_\mathbb{M}(t_o)$, and $f_\tau(t_o)$, respectively.

In an IS-41 system, HLR mobility database failures occur infrequently. Therefore the time interval between two successive HLR failures is longer than location update interval $T_p$; i.e., a failure may occur at any point in the location update interval, and thus the density distribution function $f_p(t_p)$ of $t_p$ may be approximated well by the uniform distribution:

$$f(t_p) = \frac{1}{T_p}, \quad 0 \leq t_p < T_p. \tag{8}$$

It is a reasonable assumption that the call originated from a mobile forms a Poisson process, i.e., $t_o$ has exponential distribution. The distribution of $t_o$, called the *residual residence time*, is relatively complex. The distribution of $t_e$ is related to $t_c$, the *residence time of a mobile in a LA*, through the residual life theorem [17]. The distribution $t_c$ depends on the geometry and area of LA, as well as on the traffic distribution of mobiles inside the LA. In [14], the authors assume that $t_c$ (or $t_e$) simply has an exponential distribution. A recent study [11] indicates that this assumption may not be appropriate for some practical situations. A more


general distribution is needed for the LA residence time. In [10] we proposed the so-called hyper-Erlang distribution for cell residence time. It has been shown that hyper-Erlang distributions can provide a powerful approximation to any general distribution. Thus, in this paper the hyper-Erlang distribution can also be used to characterize some time variables such as residual residence time.

We start with a simpler case first. Assume that \( t_c \) is Erlang distributed with the probability density function

\[
f_c(t_c) = \frac{m^* c (m^* c t_c)^{m-1} e^{-m^* c t_c}}{(m-1)!},
\]

where \( t \geq 0 \), and the following Laplace transform of \( f_c(t_c) \):

\[
f_c^*(s) = \left( \frac{m^* c}{m^* c + s} \right)^m.
\]

We also assume that \( t_0 \) is exponentially distributed with the explicit forms

\[
f_0(t_0) = \lambda_0 e^{-\lambda_0 t_0}.
\]

Let \( t_f = \min\{t_0, t_c\} \) and let \( f_f(t_f) \) denote the probability distribution of \( t_f \). By noticing that both \( f_0^*(z) \) and \( f_c^*(z) \) do not have poles between \((0, \infty)\), we can use Eq. (6) to compute the Laplace transform of probability density function \( f_f(t_f) \).

\[
f_f^*(s) = \mathcal{L}\{f_f(t_f)\} = f_0^*(s) + f_c^*(s) + \int_{L_1} \frac{dz}{2\pi i} \frac{f_0^*(s-z) f_c^*(z)}{z(z-s)}.
\]

Because \( f_0^*(s-z) = o(\frac{1}{z}) \) and \( f_c^*(z) = o(\frac{1}{z}) \), we can change the integration along \( L_1 \) into a contour integration \( \oint_{C_R} \), where the contour consists of the \( L_1 \) and \( C_R = \{z : \arg(z) : \pi/2 \to -\pi/2\} \). The final result can be obtained from residue theorem [18]

\[
f_f^*(s) = f_0^*(s) + f_c^*(s) + \oint_{C_R} \frac{dz}{2\pi i} \frac{f_0^*(s-z) f_c^*(z)}{z(z-s)}
\]

\[
= \left( \frac{m^* c}{m^* c + s} \right)^m + \frac{\lambda_0}{s + \lambda_0} + \oint_{C_R} \frac{dz}{2\pi i} \frac{\lambda_0}{s + \lambda_0 + z} \left( \frac{m^* c}{m^* c + z} \right)^m \frac{s}{z(z-s)}
\]

\[
= \frac{\lambda_0}{s + \lambda_0} \left[ 1 - \left( \frac{m^* c}{s + \lambda_0 + m^* c} \right)^m \right] + \left( \frac{m^* c}{s + \lambda_0 + m^* c} \right)^m.
\]

The recovery time is given by \( t_r = \min\{t_p, t_a, t_c\} = \min\{t_p, t_f\} \). Since \( f_p^*(z) \) and \( f_a^*(z) \) do not have any poles between \((0, \infty)\), we can use Eq. (6) to compute the Laplace transform of probability density function \( f_f(t_f) \).
\[ f^*_p(s) = f^*_p(s) + \frac{1}{2\pi \lambda_0} \int_0^\infty \frac{dz}{z} \frac{f^*_p(s-z) f^*_p(z) s}{z(z-s)} \]
\[ = \frac{1}{T_p s} (1 - e^{-T_p s}) + \frac{\lambda_o}{\lambda_o + s} \left[ 1 - \left( \frac{m \lambda_c}{s + \lambda_o + m \lambda_c} \right)^m \right] + \left( \frac{m \lambda_c}{\lambda_o + m \lambda_c + s} \right)^m \]
\[ + \int_0^\infty \frac{dz}{z} \left( \frac{\lambda_o}{\lambda_o + s - z} \left[ 1 - \left( \frac{m \lambda_c}{s + \lambda_o + m \lambda_c - z} \right)^m \right] + \left( \frac{m \lambda_c}{\lambda_o + m \lambda_c + s - z} \right)^m \right) \times \frac{1}{T_p s} (1 - e^{-T_p s}) \frac{s}{z(z-s)} \]  
\[ = \frac{1}{T_p s} (1 - e^{-T_p s}) \frac{s}{z(z-s)}. \]  
\[ (14) \]

Detailed calculations can be found in Appendix 2. The final result is
\[ f^*_p(s) = \frac{\lambda_o}{\lambda_o + s} \frac{1}{T_p (\lambda_o + s)^2} \]
\[ + \left( \frac{m \lambda_c}{\lambda_o + m \lambda_c + s} \right)^m \left[ 1 - \frac{\lambda_o}{\lambda_o + s} \frac{s}{T_p (\lambda_o + s)^2} - \frac{ms}{T_p (\lambda_o + m \lambda_c + s)} \right] \]
\[ - \frac{(-m \lambda_c T_p)^m}{(m-1)! (m \lambda_0 + s)} \mathcal{F}(m-1, 1, T_p(s + \lambda_0 + m \lambda_c)) \]
\[ - \frac{(-m \lambda_c T_p)^m}{(m-1)! \lambda_o + s} \mathcal{F}(m-1, 2, T_p(s + \lambda_0 + m \lambda_c)) \]
\[ + \frac{(-m \lambda_c T_p)^m}{(m-1)!} \frac{s e^{-(\lambda_0 + m \lambda_c) T_p}}{T_p (\lambda_o + s)^2} \mathcal{F}(m-1, 1, m \lambda_c T_p). \]
\[ (15) \]

where
\[ \mathcal{F}(n, k, z) = (-1)^n \frac{e^{-z}}{z^n} + \sum_{i=1}^n (-1)^i \binom{n}{i} k(k+1) \cdots (k+i-1) \frac{e^{-z}}{z^i}. \]
\[ (16) \]

We notice that \( f^*_p(s) \) in Eq. (15) is a function of the parameters \( s, \lambda_o, \lambda_c, m, T_p \). For convenience, we may express \( f^*_p(s) \) explicitly as a function of these parameters,
\[ f^*_p(s) = \mathcal{E}(s, \lambda_o, \lambda_c, m, T_p). \]
\[ (17) \]

The explicit form of \( \mathcal{E}(s, \lambda_o, \lambda_c, m, T_p) \) can be found on the right-hand side of Eq. (15). If we set \( m = 1 \), Eq. (15) is simply reduced to the result obtained in [14].

Finally, we study the general case where the hyper-Erlang distribution is used. Hyper-Erlang distribution has the following probability density function and Laplace transform
\[ f_{he}(t) = \sum_{i=1}^M x_i (m_i \lambda_{c_i} t)^{m_i-1} e^{-m_i \lambda_c t}, \]
\[ f^*_{he}(s) = \sum_{i=1}^M x_i \left( \frac{m_i \lambda_{c_i}}{s + m_i \lambda_{c_i}} \right)^{m_i}. \]
\[ (18) \]
where $M > 0$ is a positive integer, $m_i$ $(i \geq 0)$ are nonnegative integers, $\lambda_i$ $(i \geq 0)$ are positive numbers, and
\[
\sum_{i=1}^{M} \alpha_i = 1, \quad 0 \leq \alpha_i \leq 1. \tag{19}
\]

Comparing Eq. (18) with Eq. (9), we find that hyper-Erlang distribution is just a linear combination of weighted Erlang distributions.

If $t_c$ is hyper-Erlang distributed with the above distribution density, then the Laplace transform of $f_r(t_c)$ is
\[
f^*_r(s) = \sum_{i=1}^{M} \alpha_i \tilde{\epsilon}(s, \lambda, \lambda_i, m_i, \lambda_c, \sum_{j=1}^{i} \lambda_j)^{-1}. \tag{20}
\]

We observe that the parameters $M, \alpha_i, \lambda_i$ in the hyper-Erlang distribution can be used to fit the field data.

5. COST ANALYSIS

The total cost of location management comes from two sources: the cost of lost incoming calls and the cost of location updates. To evaluate the cost due to lost calls, we need to find the average number of lost incoming calls during the procedure of HLR recovery. In this section, we compute the average number of lost incoming calls.

Assuming that the call arrivals to the mobile form a Poisson process with arrival rate $\lambda_a$, the probability that $K = k$ call arrivals occur in a known period $X$ is given by
\[
\Pr(K = k \mid X = t) = \frac{\lambda_a^k t^k}{k!} e^{-\lambda_a t}, \tag{21}
\]
where the $K$ is the number of arriving calls.

Let $p_L(k)$ denote the probability that $k$ calls arrive at the mobile during the interval between an HLR mobility database failure and the failure recovery; then $p_L(k)$ is given by [14]
\[
p_L(k) = \int_0^\infty dt \Pr(K = k \mid X = t) f_r(t) = \frac{\lambda_a^k}{k!} \int_0^\infty dt^k f_r(t) e^{-\lambda_a t}
\begin{align*}
&= \frac{\lambda_a^k}{k!} \frac{d^k}{ds^k} f^*_r(s)|_{s = \lambda_a}.
\end{align*} \tag{22}
\]

The average number of lost calls during this interval is [13]
\[
E_{loss} = \sum_{i=0}^{\infty} i p(L) = \sum_{i=1}^{\infty} i \left( -\lambda_a \right)^{i-1} \frac{d^{i-1}}{dt^{i-1}} f_r(s)|_{s = \lambda_a}
\begin{align*}
&= \left( -\lambda_a \right) \sum_{i=1}^{\infty} \frac{(i-1)!}{(i-1)!} \left( -\lambda_a \right)^{i-1} \frac{d^{i-1}}{ds^{i-1}} f^*_r(s)|_{s = \lambda_a}
&= \left( -\lambda_a \right) \left[ \frac{d}{ds} f^*_r(s)|_{s = 0} \right]. \tag{23}
\end{align*}
\]
In Eq. (23), the average number of lost calls is related to the Laplace transform of $f_r^*(s)$. By noticing that $\frac{d}{ds} f_r^*(s)|_{s = 0} = E[t_r]$, we obtain

$$E_{\text{loss}} = \lambda_a E[t_r].$$  \hspace{1cm} (24)

This last result can also be obtained from Little’s law.

Next we compute the average number of lost calls in the case that residence time is Erlang distributed. Because the hyper-Erlang distribution is the linear combination of Erlang distributions, once the number of lost calls is obtained for Erlang distribution, the number of lost calls for the hyper-Erlang distributions can easily be computed.

From Eqs. (15) and (24), the average number of lost calls for Erlang distribution is

$$E_{\text{loss}} = \lambda_a + \left( \frac{m \lambda_c}{\lambda_a + m \lambda_c} \right)^m \left[ \frac{\lambda_a}{\lambda_a + m \lambda_c} + \frac{m \lambda_c}{T_p \lambda_a (\lambda_a + m \lambda_c)} \right] + \lambda_a \left( \frac{m \lambda_c}{\lambda_a + m \lambda_c} \right)^m \left[ \frac{1}{T_p \lambda_a (m - 1)!!} \mathcal{F}(m - 1, 1, T_p(\lambda_a + m \lambda_c)) \right] + \mathcal{F}(m - 1, 1, T_p(\lambda_a + m \lambda_c)) - \frac{e^{-\lambda_a T_p}}{T_p \lambda_a (m - 1)!!} \mathcal{F}(m - 1, 1, m \lambda_c T_p) \right]$$

$$\equiv \mathcal{L}(\lambda_a, \lambda_c, \lambda_s, m, T_p).$$ \hspace{1cm} (25)

When $t_r$ is hyper-Erlang distributed, we can easily apply the above result to obtain the following formula for computing $E_{\text{loss}}$:

$$E_{\text{loss}} = \sum_{i=1}^{M} \lambda_s \mathcal{L}(\lambda_a, \lambda_c, \lambda_s, m_i, T_p), \quad \lambda_s = \left( \sum_{i=1}^{M} \lambda_s \lambda_s \right)^{-1}.$$

Using the average number of lost calls, we can evaluate the cost for HLR database recovery and mobile location update of the system in the presence of HLR database failure. Assume that $c_n$ is the cost per lost call when there are $n$ calls lost during the failure recovery period and $c_r$ is the cost for sending out one location update signal. $C_{\text{tot}}$ denotes the total cost per unit time, which is given by

$$C_{\text{tot}} = \frac{1}{T_f} \sum_{n=1}^{\infty} c_n n \rho^*(n) + \frac{1}{T_p} c_r.$$ \hspace{1cm} (26)

For simplicity, we assume that $c_n$ is a constant, i.e., $c_n = c_1$; then we have

$$C_{\text{tot}} = c_1 \frac{1}{T_f} \sum_{n=1}^{\infty} n \rho^*(n) + \frac{1}{T_p} c_r$$

$$C_{\text{tot}} = c_1 \frac{1}{T_f} \sum_{n=1}^{\infty} n \rho^*(n) + \frac{1}{T_p} c_r.$$ \hspace{1cm} (27)

Equation (27) shows the relationship between $E_{\text{loss}}$ and $C_{\text{tot}}$. $E_{\text{loss}}$ can be computed from $f_r^*(s)$, the Laplace transform of $f_r(t_r)$. Once $f_r^*(s)$ is obtained, Eq. (26) can be used to evaluate the total cost $C_{\text{tot}}$. 

440 FANG, CHLAMTAC, AND FEI
6. DISCUSSIONS

In this section we present numerical results for the average number of lost calls and the cost of location updates. The analytical expressions for $E_{\text{loss}}$ and $C_{\text{tot}}$ presented in the previous section are used to evaluate $E_{\text{loss}}$ and $C_{\text{tot}}$. The final results are shown in Fig. 2-9.

In Fig. 2, the average number of lost calls is drawn as a function of the shape parameter $m$ in the Erlang distribution. In this figure, we choose $\lambda_a T_p = 20$ for illustration purposes. Because the variance of the Erlang distribution is $1/(m\mu^2)$ ($1/\mu$ is the mean of the Erlang distribution), the shape parameter $m$ represents the variance of the Erlang distribution. We find that as $m$ increases, the average number of lost calls will increase. When $m$ increases from $m = 1$ to $m = 10$, there is a significant change of $E_{\text{loss}}$. Further increase of $m$ does not have a significant effect on $E_{\text{loss}}$. This phenomenon can be explained by the characteristic of Erlang distribution. With the increase of $m$, the variation of $t_r$ decreases, so the failure recovery time $t_r = \min\{t_a, t_o, t_c\} \approx \min\{t_a, t_o, 1/\lambda_a\}$ is independent of sufficient large $m$.

In Figs. 3-6, we show the average number of lost calls as a function of $\lambda_a$, $\lambda_o$, $\lambda_s$, and $T_p$. In all these figures, $m$ assumes values 1, 2, 4, 8. Figures 3 and 6 show the average number of lost calls as a function of $\lambda_a/\lambda_o$ and $\log(\lambda_a T_p)$, in which normalization by $\lambda_a$ is applied. The variable $\lambda_o/\lambda_a$ chosen to be 1 in Fig. 3 and to be 0.5 in Fig. 4.

Figures 5 and 6, show the average number of lost calls as a function of $\lambda_s/\lambda_a$ and $\log(\lambda_a T_p)$. The variable $\lambda_s/\lambda_o$ takes 1 and 0.5 in Figs. 5 and 6, respectively.

![FIG. 2. Average number of lost calls as a function of $m$.](image-url)
FIG. 3. Average number of lost calls as a function of $\lambda_c/\lambda_a$ and $\log(\lambda_aT_p)$; $\lambda_c/\lambda_a$ is taken to be 1.

FIG. 4. Average number of lost calls as a function of $\lambda_c/\lambda_a$ and $\log(\lambda_aT_p)$; $\lambda_c/\lambda_a$ is taken to be 0.5.
FIG. 5. Average number of lost calls as a function of $\lambda_c/\lambda_a$ and log($\lambda_a T_p$); $\lambda_o/\lambda_a$ is taken to be 1.

FIG. 6. Average number of lost calls as a function of $\lambda_c/\lambda_a$ and log($\lambda_a T_p$); $\lambda_o/\lambda_a$ is taken to be 0.5.
FIG. 7. Total cost as a function of log(\(\lambda / T_p\)).

FIG. 8. Total cost as a function of log(\(\lambda / T_p\)).
From Figs. 5 and 6, we find that when $T_p$ goes to zero, the average number of lost calls also goes to zero. This conclusion is consistent with the intuition: more frequent location updates always reduces the number of lost calls. As matter of fact, by using the relationship

$$
\mathcal{F}(n, k, z) \approx \frac{(k-1+n)!}{(k-1)!} (-1)^{k-z} \frac{e^{-z}}{z^{k+n}} + n(-1)^k \frac{(k+n-z)!}{(k-1)!} \frac{e^{-z}}{z^{k+n-1}}, \quad (28)
$$

we can show analytically from Eq. (25) that when $T_p \rightarrow 0$, $E_{\text{loss}} \rightarrow 0$.

When either $\lambda_o/\lambda_0$ or $\lambda_c/\lambda_0$ increases, the average number of lost calls decreases. This is reasonable, because the increase of $\lambda_o/\lambda_0$ or $\lambda_c/\lambda_0$ implies more frequent call generations from the mobile or more frequent boundary crossings of the mobile; either case will reduce the time needed to recover HLR database failure and thus cause the $E_{\text{loss}}$ to decrease.

Comparing Figs. 3 and 4, we find that when $\lambda_c/\lambda_c$ decreases, the change rate of $E_{\text{loss}}$ with respect to $\lambda_o/\lambda_o$ increases. From Figs. 4 and 5, we can find that when $\lambda_o/\lambda_0$ decreases, the change rate of $E_{\text{loss}}$ with respect to $\lambda_c/\lambda_c$ increases.

Next we discuss the total cost of database failure recovery and location update in the presence of HLR database failure. Equation (26) can be written as

$$
C_{\text{tot}} = \frac{c_1}{T_f} E_{\text{loss}} + \frac{c_r}{T_p} = \frac{c_1}{T_f} \left[ E_{\text{loss}} + \frac{T_f}{c_1} \frac{c_r}{T_p} \right], \quad (29)
$$
where $c_1/T_f$ is the cost of one lost call per unit time. For convenience, we can set $c_1/T_f$ to 1. $C_{\text{tot}}$ is regarded as a function of the dimensionless normalized location update interval $\lambda_u T_p$ (or location update period $T_p$ in units of $1/\lambda_u$). To simplify our discussion, we also define a new parameter, $c \equiv \lambda_u c_v$, which is the cost of $\lambda_u$ location updates.

Figures 7–9 show the total cost of failure recovery and location update. The cost is in units of $c_1/T_f$, which is taken to be 1. Curves (a), (b), (c), and (d) correspond to different values of $c$, which are 0.01, 0.1, 1, and 10, respectively.

Figure 7 shows the total cost as $T_p$ changes (remember $T_p$ is in units of $1/\lambda_u$ and we take the log$_{10}$ scale for $T_p$). We choose $m = 1, 2, 4, 8, \lambda_o/\lambda_u = 0.1, \lambda_c/\lambda_u = 0.1$.

Figures 7–9 show that when $m > 1$ there always exist optimal values of $T_p$ for the above chosen parameters. When $T_p \to \infty$, $C_{\text{tot}}$ approaches a constant. In fact for $m = 1$, for some system parameters the optimal $T_p$ can be analytically obtained [13]. For the system parameters we used above, $E_{\text{loss}}$ may be too small to guarantee the existence of optimal $T_p$ when $m = 1$, and cost from location update dominates the total cost.

As $T_p$ increases, the total cost mainly results from the cost of lost calls. From Fig. 7 we find that when $m$ increases, the asymptotic values with $T_p \to \infty$ increase, especially for $m = 1$ and $m = 2$. This result is consistent with the observation that when $m$ increases from $m = 1$ to $m = 2$, $E_{\text{loss}}$ has a significant change. However, further increase of $m$ does not bring the same effect on $E_{\text{loss}}$. This observation can greatly simplify the discussion of hyper-Erlang distribution.

Figure 8 shows the total cost as $T_p$ changes. We take $m = 4, \lambda_o/\lambda_u = 0.1, \lambda_c/\lambda_u = 0.01, 0.1, 0.5, 1$.

Figure 8 shows that when $\lambda_c/\lambda_u$ increases, the existence of optimal $T_p$ becomes less obvious. Especially when $\lambda_c/\lambda_u = 0.1$, 1 and $c = 1$, there is no optimal $T_p$ at all. The reason that uniqueness of an optimal value disappears is as follows: as $\lambda_c/\lambda_u$ increases, there are fewer incoming calls lost; when $c$ is larger (here $c = 1$) the cost due to location updates plays a more significant role. The figure also show that when $\lambda_c/\lambda_u$ increases, the asymptotic value of $C_{\text{tot}}$ as $T_p \to \infty$ decreases due to the loss of fewer incoming calls.

Figure 9 shows the total cost as the $T_p$ changes. Here, we take $m = 4, \lambda_o/\lambda_u = 0.1, \lambda_c/\lambda_u = 0.01, 0.1, 0.5, 1$. We observe that there exists a unique optimal value of $T_p$ for some cases. When both $\lambda_o/\lambda_u$ and $c$ increase, the optimal value of $T_p$ is not unique.

7. CONCLUSIONS

In this paper we analyzed the active recovery scheme of HLR mobility database failure in wireless networks and mobile computing systems. We presented general analytical results for computing the probability density distribution of the failure recovery time and the average number of lost calls when HLR mobility database failure occurs. The cost analysis for the failure recovery strategy was conducted analytically and an optimal determination of location update interval was also given. Our approach, which can be applied to any traffic distribution, provides an optimal scheme for the design of location update interval to actively combat the
mobility database failures in wireless networks and mobile computing systems, in particular the PCS networks.

APPENDIX 1: PROOF OF THEOREM 3.2

The Laplace transform of \( f_m(t) \) can be calculated from Eq. (1),

\[
\mathcal{L}[f_m(t)] = \mathcal{L} \left[ f_1(t) \int_{t}^{\infty} f_2(\tau) \, d\tau + f_3(t) \int_{t}^{\infty} f_4(\tau) \, d\tau \right],
\]

where \( \mathcal{L} \) denotes the Laplace transformation operator.

The first term in the right-hand side is

\[
\mathcal{L} \left[ f_1(t) \int_{t}^{\infty} f_2(\tau) \, d\tau \right] = f_1(s) - \int_{0}^{\infty} d\tau f_2(\tau) \int_{0}^{\infty} dt f_4(\tau) = f_1(s) - \int_{0}^{\infty} d\tau f_2(\tau) \int_{0}^{\infty} dt f_4(\tau). \tag{31}
\]

Using \( \delta(t' - t) = \int_{-\infty}^{\infty} \frac{d\epsilon}{2\pi} e^{i\epsilon(t' - t)} \), we obtain

\[
\mathcal{L} \left[ f_3(t) \int_{t}^{\infty} f_2(\tau) \, d\tau \right] = f_3(s) - \int_{-\infty}^{\infty} \frac{d\epsilon}{2\pi} \int_{0}^{\infty} d\tau f_2(\tau) \int_{0}^{\infty} dt e^{-s+\epsilon t} f_4(\tau) \int_{0}^{t} dt' \delta(t' - t) \int_{0}^{t} d\tau f_3(\tau), \tag{32}
\]

where we introduce \( \epsilon = 0^+ \) to ensure the convergence of the integral.

Using the partial integration formula, we have the relationship

\[
\int_{0}^{\infty} dt e^{-s+\epsilon t} f_2(\tau) = \frac{1}{\epsilon - j\omega} \int_{0}^{\infty} dt' f_2(t') e^{-(s-j\omega)t} = f_2^{*}(\epsilon - j\omega). \tag{33}
\]

From (31) and (32), we have

\[
\mathcal{L} \left[ f_1(t) \int_{t}^{\infty} f_2(\tau) \, d\tau \right] = f_1^{*}(s) - \int_{-\infty}^{\infty} \frac{d\epsilon}{2\pi} f_2^{*}(s - j\omega) f_4^{*}(\epsilon + j\omega) \frac{\epsilon}{\epsilon + j\omega}. \tag{34}
\]

If we introduce a complex variable \( z = \epsilon + j\omega \), then Eq. (34) can be written in the form

\[
\mathcal{L} \left[ f_1(t) \int_{t}^{\infty} f_2(\tau) \, d\tau \right] = f_1^{*}(s) - \int_{-\infty}^{\infty} \frac{dz}{2\pi j} \frac{f_2^{*}(s - z) f_4^{*}(z)}{z}. \tag{35}
\]
Similarly, we have
\[
\mathcal{L} \left[ f_2(t) \right] \left[ e^{-t \lambda} \right] = f_2^*(s) - \int \frac{dz}{2\pi j} \frac{f_2^*(s-z) f_2^*(z)}{z}.
\]  
(36)

Combing Eqs. (35) and (36) we can complete the proof of Theorem 3.1.

**APPENDIX 2: COMPUTATION OF** \( f_2^*(s) \)

To begin with,
\[
f_2^*(s) = \frac{1}{T_p s} (1 - e^{-T_p s}) + \frac{1}{\lambda_0 + s} \left[ \frac{m \lambda_c}{s + \lambda_0 + m \lambda_c} \right]^m
\]
\[+ \int \frac{dz}{2\pi j} \frac{i_{\lambda_0} \lambda_0 + s - z}{\lambda_0 + s - z} \left( \frac{m \lambda_c}{\lambda_0 + m \lambda_c + s - z} \right)^m \frac{s e^{-T_p s} - 1}{T_p z^2(s-z)}.
\]  
(37)

The third term in the above expression has two parts, the part with exponential factor \( e^{-T_p s} \) and the part without an exponential factor. The latter is
\[
= -\int \frac{dz}{2\pi j} \frac{i_{\lambda_0} \lambda_0 + s - z}{\lambda_0 + z - \lambda_0} \frac{s}{T_p z^2(z-s)}
\]
\[\left[ \frac{(-m \lambda_c)^m s}{T_p} \int \frac{dz}{2\pi j} \frac{1}{[z-(\lambda_0 + s)][z-(\lambda_0 + m \lambda_c + s)]} \right]^m.
\]  
(38)

The two integrals in Eq. (38) can be calculated by combining \( l_z \) with \( C_R \) and changing the above integrals into contour integrals, where \( C_R = \{ z \mid z > \infty, \arg(z) \}$\( z \rightarrow \frac{2\pi}{2} \}$ \) is the half circle in the left-hand side of vertical axis in the complex plane. The integrals along \( C_R \) with the same integrand as in Eq. (38) will disappear. From the residue theorem, we can easily obtain
\[
\int \frac{dz}{2\pi j} \frac{i_{\lambda_0} \lambda_0 + s - z}{\lambda_0 + s - z} \left( \frac{m \lambda_c}{\lambda_0 + m \lambda_c + s - z} \right)^m \frac{s}{T_p z^2(s-z)}
\]
\[= \frac{1}{T_p} \left\{ \frac{s}{(\lambda_0 + s)^2} - \frac{s}{T_p (\lambda_0 + s)} \frac{1}{(\lambda_0 + s) T_p} \left( \frac{m \lambda_c}{\lambda_0 + m \lambda_c + s} \right)^m \left[ \frac{1}{\lambda_0 + s} + \frac{m}{\lambda_0 + m \lambda_c + s} \right] \right\}.
\]  
(39)

The part with an exponential factor in the third term on the right-hand side of Eq. (37) is
Combining Eqs. (37), (39), and (41), we can obtain the expression for $f$ in Eq. (14).

$$
\int \frac{dz}{2\pi j} \frac{\lambda_0}{\lambda_0 + s - z} + \frac{s - z}{\lambda_0 + s - z} \left( \frac{m\lambda_c}{\lambda_0 + m\lambda_c + s - z} \right)^m \frac{se^{-T_p z}}{T_p z^2(s - z)}
$$

$$
= \int \frac{dz}{2\pi j} \frac{\lambda_0}{\lambda_0 + s - z} \frac{se^{-T_p z}}{T_p} \frac{-(-m\lambda_c)^m s}{T_p} \frac{dz}{2\pi j} \frac{1}{z - (\lambda_0 + s)} \frac{se^{-T_p z}}{T_p z^2(s - z)}
$$

from which we can similarly obtain

$$
\int \frac{dz}{2\pi j} \frac{\lambda_0}{\lambda_0 + s - z} + \frac{s - z}{\lambda_0 + s - z} \left( \frac{m\lambda_c}{\lambda_0 + m\lambda_c + s - z} \right)^m \frac{se^{-T_p z}}{T_p z^2(s - z)}
$$

$$
= \frac{1}{T_p} \left[ \frac{s}{(\lambda_0 + s) - T_p(\lambda_0 + s)^2} \right] \frac{e^{-T_p z}}{T_p} \frac{s}{(m - 1)!} \frac{d^{m-1}}{dz^{m-1}} \left[ \frac{1}{(\lambda_0 + s) - T_p(\lambda_0 + s)^2} \right] \left[ \frac{e^{-T_p z}}{T_p} \right] \left[ \frac{1}{(\lambda_0 + s) - T_p(\lambda_0 + s)^2} \right]
$$

$$
+ \frac{-s}{T_p} \left[ \frac{1}{(\lambda_0 + s) - T_p(\lambda_0 + s)^2} \right] \frac{e^{-T_p z}}{T_p} \frac{s}{(m - 1)!} \frac{d^{m-1}}{dz^{m-1}} \left[ \frac{1}{(\lambda_0 + s) - T_p(\lambda_0 + s)^2} \right] \left[ \frac{e^{-T_p z}}{T_p} \right] \left[ \frac{1}{(\lambda_0 + s) - T_p(\lambda_0 + s)^2} \right]
$$

$$
+ \frac{e^{-T_p z}}{T_p} \frac{s}{T_p} \frac{1}{(\lambda_0 + s) - T_p(\lambda_0 + s)^2} \left[ \frac{-m\lambda_c}{} \right] \left[ \frac{1}{(\lambda_0 + s) - T_p(\lambda_0 + s)^2} \right] \left[ \frac{e^{-T_p z}}{T_p} \right] \left[ \frac{1}{(\lambda_0 + s) - T_p(\lambda_0 + s)^2} \right]
$$

$$
+ \frac{s}{T_p} \frac{1}{(\lambda_0 + s) - T_p(\lambda_0 + s)^2} \left[ \frac{-m\lambda_c}{} \right] \left[ \frac{1}{(\lambda_0 + s) - T_p(\lambda_0 + s)^2} \right] \left[ \frac{e^{-T_p z}}{T_p} \right] \left[ \frac{1}{(\lambda_0 + s) - T_p(\lambda_0 + s)^2} \right]
$$

Combining Eqs. (37), (39), and (41), we can obtain the expression for $f^*(s)$ given in Eq. (14).

REFERENCES