On optimizing the location update costs in the presence of database failures

Zygmunt J. Haas [a] and Yi-Bing Lin [b]

[a] *School of Electrical Engineering, Cornell University, Ithaca, NY 14853, USA*

[b] *Department of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu, Taiwan, ROC*

This paper studies the database failure recovery procedure for cellular phone networks as part of the Electronic Industries Association/Telecommunications Industry Association Interim Standard 41 (EIA/TIA IS-41). Before the location information of the database is recovered, phone calls may be lost. The restoration process can be sped up by having the mobile phones to periodically confirm their existence by radio contact with the cellular network. We show that, under some cost assumptions, periodic update interval should be chosen to be approximately equal to the call interarrival time, with more frequent updates for more unreliable system. We also show that the cost of an optimized system is relatively small and stable, if the system is even moderately reliable. Finally, if the system is at least moderately reliable, the effects of call origination rate and the rate at which Location Areas are crossed are rather small, assuming that the periodic update interval was chosen as stated above. Thus, in such cases, optimization of the size of the Location Area can be made independent of the optimization of the periodic update process.

1. Introduction

In a cellular mobile phone network, a mobile phone may move from one location to another. To deliver a phone call to a mobile subscriber, the location of the mobile phone must be identified before the connection can be established. The current location of a mobile phone is usually maintained by a two-level hierarchical strategy with two types of databases. We assume that a global cellular phone network consists of several cellular phone service providers (or cellular systems). The *Home Location Register* (HLR) is the location register to which a mobile phone identity is assigned for record purposes such as mobile user information (e.g., directory number, profile information, current location, authentication information, billing, validation period). The HLR is a database residing in the home system of a mobile phone. (A home system is the cellular phone system where the mobile user subscribes to his/her service.) The *Visitor Location Register* (VLR) is the location register, other than the HLR, used to retrieve information for handling of calls to or from a visiting mobile user. A VLR is the database associated with a cellular system (other than the home system) that the mobile phone is currently visiting.

When a mobile phone is in its home system, it can be accessed directly through its home Mobile Switching Center (MSC). If the mobile phone leaves its home system and moves from one visited system to another, it must register with the new visited MSC. We define the coverage of an MSC as a *Location Area* (LA). Thus, a mobile that moves from one LA to another (i.e., crosses a LA boundary), initiates a registration with the new visited MSC. Such a registration is performed by entering the mobile's identity in the VLR of the new visited system (see step 1 in figure 1). The new visited MSC then informs the mobile's home MSC of the mobile's current location (i.e., the address of the new MSC). The home MSC updates the current mobile's location in its HLR and sends an acknowledgement to the new visited MSC (see step 2 in figure 1). The new visited MSC then informs the mobile phone of the successful registration (see step 3 in figure 1). The home MSC also sends a deregistration message to cancel the obsolete location record of the mobile phone in the old visited MSC's VLR. The old visited MSC acknowledges the deregistration (see step 4 in figure 1).

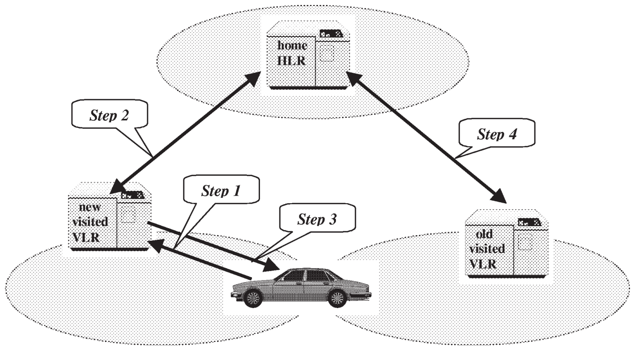

Figure 1. The mobile phone registration process.

To originate a call, a roaming mobile phone first contacts the visited MSC in the cellular network. The call request is processed with the assistance of the local VLR and is eventually connected to the called party.

If a fixed user calls a mobile subscriber, the call is forwarded to a switch (called the *originating switch*) in the public switched telephone network (PSTN). The call is routed to the mobile's home MSC (see step 1 in figure 2).[1] The home MSC, after verifying that the mobile is "away from home" by consulting its HLR, retrieves the

[1] MSC serves the function of a Central Office (CO).

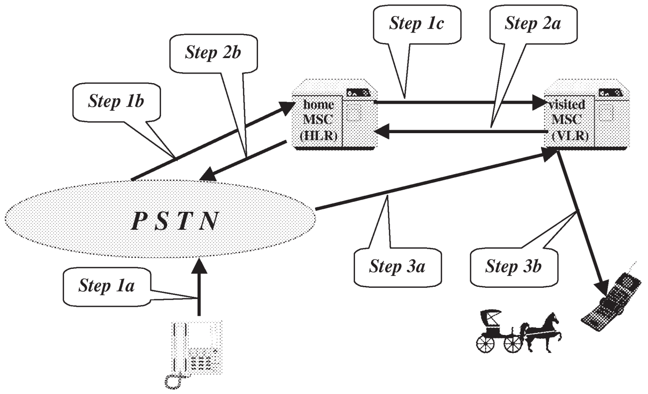

Figure 2. The call setup procedure.

current mobile's location by checking the mobile's VLR entry in the HLR. The home MSC then queries the current mobile's VLR, which returns a routable address (the *Temporary Directory Number* (TDN)) to the originating switch in the PSTN through the home MSC (see step 2 in figure 2). Based on the routable address, a trunk (voice circuit) is then set up from the originating switch to the mobile phone through the visited MSC (see step 3 in figure 2).

If the HLR fails, one may not be able to access a mobile phone. In particular, after a failure, the location information stored in the HLR may either be corrupted or obsolete. Consequently, we assume in this paper that after an HLR failure, the location information is lost. Thus, a database recovery procedure is required after a failure to guarantee the service availability to a mobile phone. Restoration of a mobile's entry within an HLR is done by one of the following three processes:

- by the *autonomous registration* mechanism, in which a mobile periodically reregisters with the system,
- by the mobile's originating a cell, which reveals the current location of the mobile in the network, or
- by the mobile's crossing a LA boundary, which prompts reregistration.

In this paper, we study the effect of the HLR failures on various system's parameters. In particular, we come up with a set of recommendation for setting up the value of the periodic interval for the autonomous registration mechanism. Among other results, we claim that a good "rule-of-thumb" is to set the periodic update interval to be approximately equal to the average call interarrival time in a practical macrocellular system.

2. Previous and related work

Optimization of location update mechanism has been a subject of intense recent research (e.g., [1,2,6,9]). Most of these studies concentrates on the tradeoff between the registration and the paging processes. Sizing the LA is one design parameter which allows to implement the above tradeoff [12]; i.e., the larger the LA is, the lower the registration cost is, alas at the expense of higher paging cost. The optimum depends on the Call-to-Mobility Ratio; i.e., how often a user receives a call and how mobile the user is (how often the user crosses LA boundaries). Some of these studies impose additional constraints on the user location process, such as paging delay, for example, [10].

There is little discussion in the technical literature related to the reliability issues of the location management system. In [7], failure restoration of HLR is addressed based on a checkpointing procedure. However, since the IS-41 recovery procedure does not rely on checkpointing to repair a failed HLR, that paper does not address the HLR reliability in IS-41 based network. Our work here presents a complete analysis of the IS-41 failure recovery procedure and proposes a framework within which the values of the system's parameters can be sized. Additionally, we extend the failure recovery analysis by including factors which were unaccounted for in the other work.

3. The IS-41 failure restoration procedure

In EIA/TIA IS-41 [3], the HLR recovery procedure works as follows.

1. After a failure, the HLR initiates the recovery procedure by sending an *Unreliable Roamer Data Directive* message to all of its associated VLRs. The VLRs then remove all records of all the mobile phones associated with that HLR.
2. At some future point in time, a 'lost' mobile phone contacts the visited MSC for call origination or by performing the location update procedure (to be described later). The visited MSC then sends a registration message to the mobile's HLR, allowing the home MSC to reconstruct its HLR's internal data structures in an incremental fashion.

After the HLR failure, if the first event of a mobile phone is either a call origination (a request from the mobile phone) or crossing of an LA boundary, then the visited MSC detects the existence of the mobile phone, and the HLR record of the mobile phone is restored through the registration process. If, before the HLR's information is restored, there are attempts to call the mobile phone (i.e., call arrivals), then these call arrivals are lost, since the HLR cannot identify the location of the mobile phone.

The delay to confirm the location of a mobile phone after an HLR failure depends on the traffic from the mobile phone and on the mobility of the mobile user. If a mobile phone is silent or stationary for a long time, all call arrivals during this period will be lost. To allow restoration of a failed HLR in such cases, a mechanism called *periodic location updating*[2] [4] was introduced, which reduces the HLR restoration delay. In this approach a mobile phone periodically establishes radio contact with the network to confirm its location. These location updates are,

[2] We also refer to a location update as location confirmation.

in fact, registrations that allow to update the current mobile's location in its HLR. This paper analyzes the impact of the periodic location updating on the performance (i.e., the number of lost calls) of the IS-41 failure restoration procedure.

The following nomenclature is used throughout this paper:

- C_{total}: the location management cost function, which includes the effects of HLR failures,
- c_n, $n = 1, 2, \ldots$: the cost of losing n calls destined to a mobile phone,
- c_{u}: the cost of a single periodic location confirmation,
- t_{sojourn}: average time a mobile spends in a LA,
- λ_{o}: the arrival rate of call originations by the mobile phone,
- λ_{c}: the rate of crossing LA boundaries by the mobile phone,
- λ_{u} $(= \lambda_{\text{o}} + \lambda_{\text{c}})$: the combined rate of call originations and LA crossings events by the mobile phone,
- λ_{a}: the arrival rate of call arrivals to the mobile phone,
- $p_{\text{L}}(k)$: the probability that after an HLR failure k calls to a mobile phone are lost before its HLR record is recovered,
- t_{O}: the time interval between two consecutive call originations from the mobile phone,
- t_{A}: the time interval between two consecutive call arrivals to the mobile phone,
- t_{C}: the time interval between two LA crossings by the mobile phone,
- T_{p}: the constant time interval between two periodic location update of a mobile phone,
- t_{o}: the time interval between the HLR failure and the next call origination from the mobile phone,
- t_{a}: the time interval between the HLR failure and the next call arrival to the mobile phone,
- t_{c}: the time interval between the HLR failure and the next LA crossing by the mobile phone,
- t_{p}: the time interval between the HLR failure and the next mobile phone periodic location update,
- T_{f}: Mean Time Between Failures (MTBF), the average time between two HLR failures,
- t_{r} $(= \min(t_{\text{p}}, t_{\text{c}}, t_{\text{o}}))$: the time interval between the HLR failure and one of the following three events, whichever occurs first: next location update, next LA crossing, or next call origination,
- t_{u} $(= \min(t_{\text{c}}, t_{\text{o}}))$: the time interval between the HLR failure and one of the following two events, whichever occurs first: next LA crossing, or next call origination,
- $f_{\text{L}}(k, t)$: the joint density function of k call arrivals to a mobile phone before the mobile phone's HLR record is recovered and of the period between the HLR failure and the record restoration time being t,
- f_{r}: the density function for t_{r},
- f_{r}^*: the Laplace transform for f_{r},
- f_{p}: the density function for t_{p},
- f_{o}: the density function for t_{o},
- f_{u}: the density function for t_{u}.

The following are assumptions used in our analysis:

- The interval between two HLR failures is a random variable with a general distribution. The HLR failure is much less frequent than the call originations and than the periodic location updates.
- The call originations and call arrival processes each constitutes a Poisson process with average rates of λ_{o} and λ_{a}, respectively.
- The sojourn time of a mobile within a LA is exponentially distributed with mean equal to

$$t_{\text{sojourn}} = \frac{9}{3 + 2\sqrt{3}} \frac{R}{V} \,,$$

  where R is the "radius" of a LA and V is the average mobile's velocity. This assumption is based on [5], in which the mobile's sojourn time in a cell is considered. Following the same arguments, the mobile's sojourn time in a LA can be obtained. The rate of LA boundary crossings can then be calculated as $\lambda_{\text{c}} = 1/t_{\text{sojourn}}$.
- The four random processes, call origination, call arrival, mobile's LA crossings, and an HLR failure, are all independent one from another.

4. The model

Figure 3 illustrates the timing diagram before and after an HLR failure. Due to the memoryless property of the exponential distributions, t_{o}, t_{a}, and t_{c} have the same distributions as t_{O}, t_{A}, and t_{C}, respectively. That is, after the HLR failure, the call arrivals are still a Poisson process and the density function of t_{o}, f_{o}, is exponentially distributed with mean $1/\lambda_{\text{o}}$. It can be shown that [8] if the interval between two HLR failures is much longer than T_{p}, which should be the case in any practical system, then t_{p} has a uniform distribution in the interval $[0, T_{\text{p}}]$. That is, the density function f_{p} of t_{p} is

$$f_{\text{p}}(t_{\text{p}}) = \frac{1}{T_{\text{p}}}, \quad 0 \leqslant t_{\text{p}} \leqslant T_{\text{p}}. \tag{1}$$

After an HLR failure, the existence of the mobile phone is known to the HLR at time

$$t_{\text{r}} = \min(t_{\text{o}}, t_{\text{c}}, t_{\text{p}}) = \min\big(\min(t_{\text{o}}, t_{\text{c}}), t_{\text{p}}\big) = \min(t_{\text{u}}, t_{\text{p}}).$$

(Note, in figure 3, $t_{\text{r}} = t_{\text{p}}$.) As both of the random processes, the call origination and the crossing of LA boundaries, are independent and exponentially-distributed,

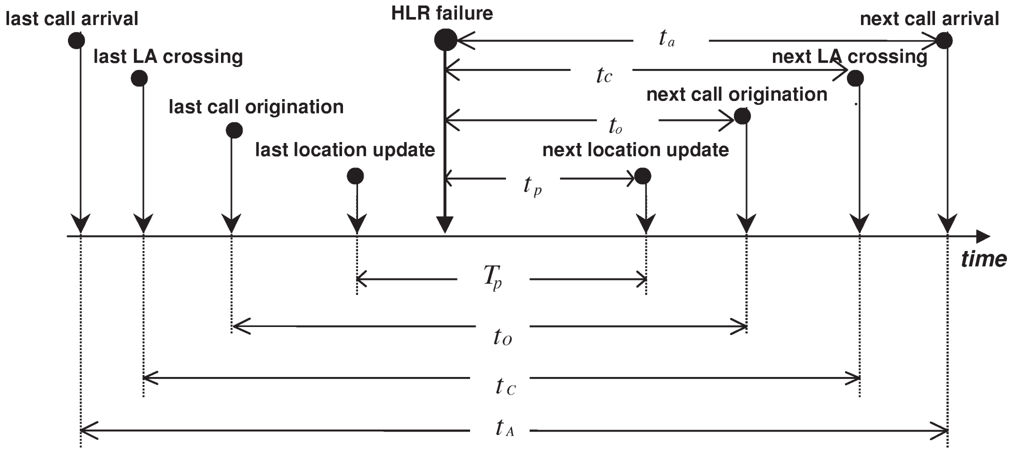

Figure 3. The timing diagram for deriving $p_L(n)$.

the probability density function of the process t_u is as follows:

$$f_u(t_u) = \lambda_o e^{-\lambda_o t_u} e^{-\lambda_c t_u} + \lambda_c e^{-\lambda_c t_u} e^{-\lambda_o t_u} = (\lambda_o + \lambda_c) e^{-(\lambda_o+\lambda_c)t_u} = \lambda_u e^{-\lambda_u t_u}. \quad (2)$$

Thus, the process is also exponential with the rate equal to the sum of the call originations and LA crossing rates: $\lambda_u = \lambda_o + \lambda_c$.

Since $0 \leqslant t_p \leqslant T_p$, it is clear that $0 \leqslant t_r \leqslant T_p$. The density function f_r of t_r is then

$$\begin{aligned} f_r(t_r) &= \int_{t_p=t_r}^{T_p} f_u(t_r) f_p(t_p)\,dt_p + \int_{t_u=t_r}^{\infty} f_u(t_u) f_p(t_r)\,dt_u \\ &= \int_{t_p=t_r}^{T_p} \lambda_u e^{-\lambda_u t_r}\left(\frac{1}{T_p}\right)dt_p + \int_{t_u=t_r}^{\infty} \lambda_u e^{-\lambda_u t_u}\left(\frac{1}{T_p}\right)dt_u \\ &= \begin{cases} \frac{\lambda_u(T_p-t_r)+1}{T_p} e^{-\lambda_u t_r}, & 0 \leqslant t_r \leqslant T_p, \\ 0, & \text{otherwise} \end{cases} \end{aligned} \quad (3)$$

and the Laplace transform of f_r is

$$\begin{aligned} f_r^*(s) &= \int_{t_r=0}^{\infty} f_r(t_r) e^{-st_r}\,dt_r \\ &= \int_{t_r=0}^{T_p} \frac{\lambda_u(T_p - t_r) + 1}{T_p} e^{-(\lambda_u+s)t_r} dt_r = \cdots \\ &= \frac{\lambda_u}{\lambda_u + s} + \frac{s}{T_p(\lambda_u + s)^2}\left[1 - e^{-(\lambda_u+s)T_p}\right]. \end{aligned} \quad (4)$$

Call arrivals are a Poisson counting process with the arrival rate λ_a. The probability that k call arrivals occur in a known period X is

$$\Pr[K = k \mid X = t] = \frac{(\lambda_a t)^k}{k!} e^{-\lambda_a t}. \quad (5)$$

Suppose that during t_r, there are k call arrivals to the mobile phone. The joint density function that $X = t_r$ and $K = k$ is

$$f_L(k, t_r) = \Pr[K = k \mid X = t_r] f_r(t_r) = \frac{(\lambda_a t_r)^k}{k!} e^{-\lambda_a t_r} f_r(t_r). \quad (6)$$

The probability that k call arrivals (to a mobile phone) arrive between the HLR failure and the next call origination, next location confirmation, or next LA crossing, whichever occurs sooner, is

$$p_L(k) = \int_{t_r=0}^{\infty} f_L(k, t_r)\,dt_r = \frac{\lambda_a^k}{k!}\int_{t_r=0}^{\infty} t_r^k f_r(t_r) e^{-\lambda_a t_r}\,dt_r = \frac{\lambda_a^k}{k!}\left[(-1)^k \frac{d^k}{ds^k} f_r^*(s)\right]\Bigg|_{s=\lambda_a}. \quad (7)$$

The last equality in (7) follows from the fact that the Laplace transform of $tf(t)$ is $-df_r^*(s)/ds$ (see [11]). From equations (4) and (7), we can compute the value of $p_L(k)$ for any k. For example,

$$p_L(0) = \frac{\lambda_u}{\lambda_u + \lambda_a} + \frac{\lambda_a}{T_p(\lambda_u + \lambda_a)^2}\left[1 - e^{-(\lambda_u+\lambda_a)T_p}\right], \quad (8)$$

$$p_L(1) = \frac{\lambda_u\lambda_a}{(\lambda_u + \lambda_a)^2} + \frac{\lambda_a(\lambda_a - \lambda_u)}{T_p(\lambda_u + \lambda_a)^3}\left[1 - e^{-(\lambda_u+\lambda_a)T_p}\right] - \frac{\lambda_a^2}{(\lambda_u + \lambda_a)^2} e^{-(\lambda_u+\lambda_a)T_p}. \quad (9)$$

The average number of calls arriving between an HLR failure and its repair, E_{loss}, is thus:

$$E_{loss} = \sum_{i=0}^{\infty} i p_L(i) = \sum_{i=0}^{\infty} i \frac{\lambda_a^i}{i!}\left[(-1)^i \frac{d^i}{ds^i} f_r^*(s)\right]\Bigg|_{s=\lambda_a} = \left\{e^{(-\lambda_a(d/ds))}\left[-\lambda_a \frac{d}{ds} f_r^*(s)\right]\right\}\Bigg|_{s=\lambda_a}. \quad (10)$$

E_{loss} can be calculated for given values of λ_o, λ_c, λ_a, and T_p using equations (4), (7), (10), and the definition of λ_u.

The crossing rate of LA boundaries, λ_c, strongly depends on the actual parameters of the system (i.e., the size of the LAs and the mobility of the mobiles). Since we prefer our conclusions to be general, independent on the particular system implementation, we evaluate the E_{loss} as a function of λ_u, rather than separating the contribution of λ_o and λ_c. To apply our results to an actual system implementation, the values of the LA size (R) and the average mobile's velocity (V) are used in the equation for $t_{sojourn}$ in section 3

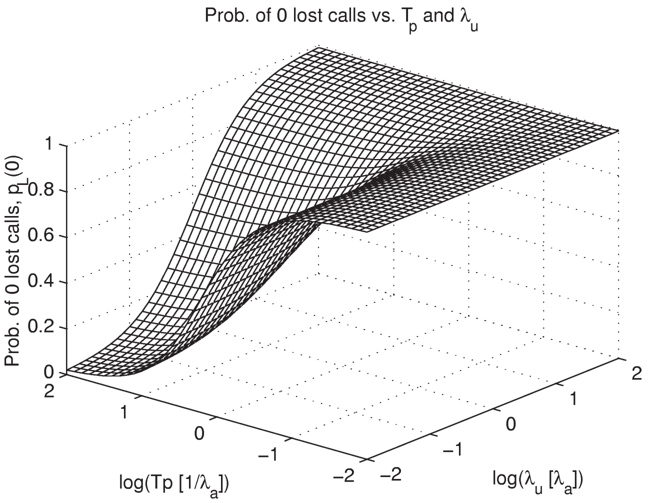

Figure 4. Probability of loosing 0 calls as a function of T_p and λ_u.

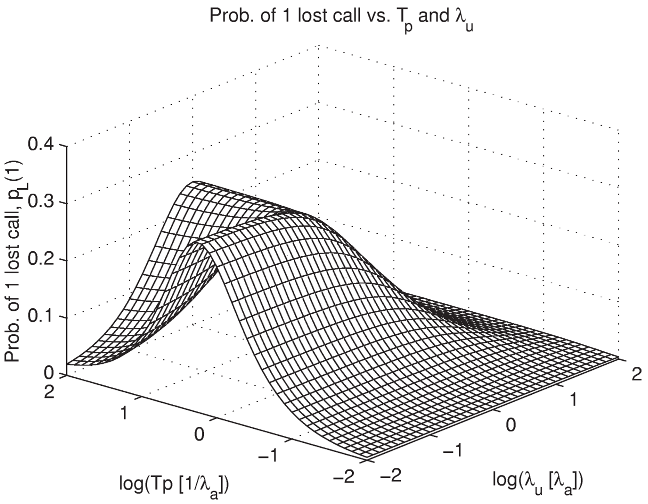

Figure 5. Probability of loosing 1 call as a function of T_p and λ_u.

to evaluate λ_c. This value is then used in substitution of $\lambda_\mathrm{c} + \lambda_\mathrm{o}$ in place of λ_u. We further notice that in a practical macrocellular system, the contribution of λ_c will be small, due to the fact that the LAs tend to be big. This is especially true for a system with relatively slowly moving mobiles. In such cases, the average number of calls that a users receives while in an LA is much larger than 1. Thus, $\lambda_\mathrm{u} \approx \lambda_\mathrm{o}$. We refer here to such a system as a *Low Mobility System* (LMS).

Figures 4–8 show the $p_\mathrm{L}(n)$ for $n = 0, \ldots, 4$, as a function of log(T_p) and log(λ_u). In these figures and in the rest of this paper, we express both T_p and λ_u in units of $1/\lambda_\mathrm{a}$ and λ_a, respectively. As can be observed from figure 4, for large values of λ_u and/or small values of T_p, there will be, essentially, no lost calls due to HLR failure. (Of course, this is intuitively so, since at these values the HLR restoration is fast.) As n increases, the maximum at which $p_\mathrm{L}(n)$ occurs is shifted towards larger values of T_p and smaller values of λ_u. For large T_p and small λ_u, the probability of loosing a small number of calls is very small, while the

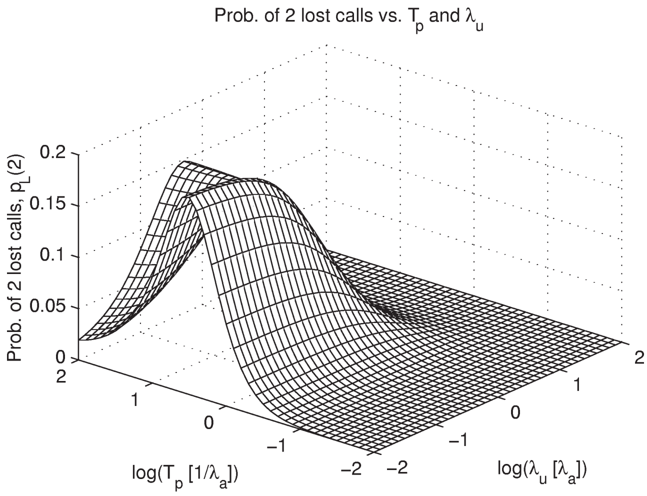

Figure 6. Probability of loosing 2 calls as a function of T_p and λ_u.

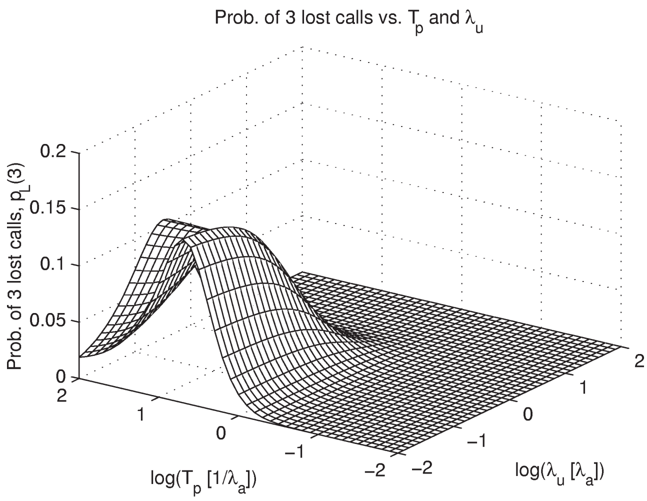

Figure 7. Probability of loosing 3 calls as a function of T_p and λ_u.

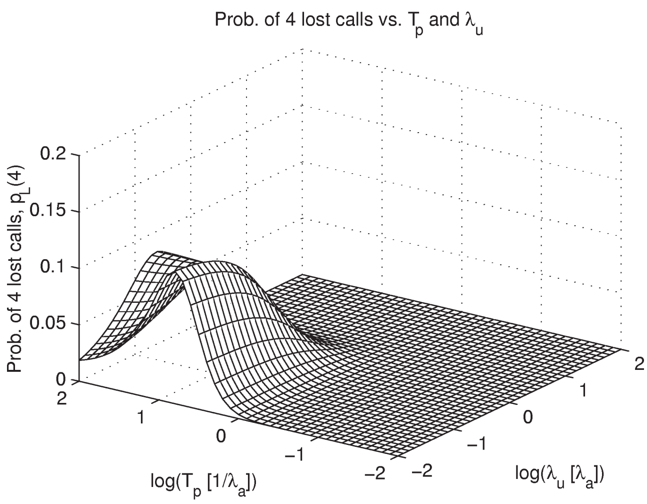

Figure 8. Probability of loosing 4 calls as a function of T_p and λ_u.

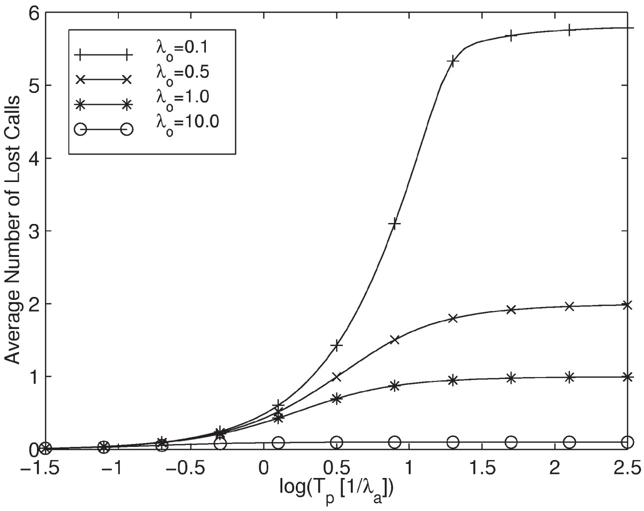

Figure 9. Average number of lost calls as a function of T_p.

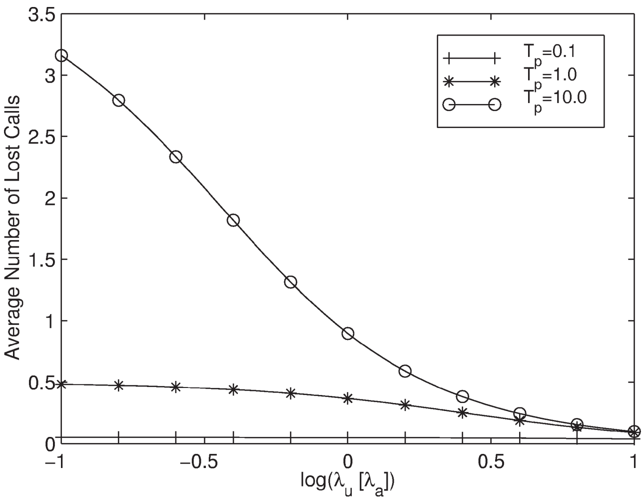

Figure 10. Average number of lost calls as a function of λ_u.

probability of loosing many calls is large. In other words, $p_\mathrm{L}(n)$ at a particular point ($T_\mathrm{p}, \lambda_\mathrm{u}$) is rather a sharp function. Finally, the maximum value of $p_\mathrm{L}(n)$ is rather fast decreasing with n. These observations suggest that only a limited number of terms is necessary in equation (10) for good accuracy. The actual number of terms required for a given accuracy can be estimated from the values of T_p and $\lambda_\mathrm{u}[\lambda_\mathrm{a}]$. Using this strategy, E_loss has been numerically evaluated and the results are depicted in figures 9 and 10, as a function of T_p and λ_u.

From what we have concluded before, it is not surprising that the average number of lost calls increase with an increase in T_p and with a decrease in λ_u. As a point of reference we note the condition in which the average number of lost calls is at most 0.5, which occurs for $T_\mathrm{p} \leqslant 1[1/\lambda_\mathrm{a}]$, for the range of λ_u studied. Another point of interest is the case when $\lambda_\mathrm{u} = 1[\lambda_\mathrm{a}]$, which corresponds to an LMS with balanced in and out traffic (i.e., when the incoming traffic density equals outgoing traffic density). In this condition, the average number of lost calls is below 1, for $T_\mathrm{p} \leqslant 10$, as can be seen from figure 10.

Of course, one would like to reduce the probability that a call is lost due to HLR failure. This could be done by the mobile's updating its location often (i.e., small T_p). However, location updates consume wireless resources and, thus, should not be too frequent. On the other hand, infrequent updates results in more lost calls. Thus, there is a tradeoff between the penalty associated with failing to deliver a call and the cost of user tracking. This tradeoff will depend on the ratio between the call arrival and the call origination rate.

We label by c_u the cost of a single location update and by $c_n,\ n = 1, 2, \ldots$, the cost of loosing n calls. Assuming that the average time between HLR failure is T_f, the cost function, C_total, which expresses the average cost per unit time of location updates and the average penalty per unit time of lost calls, is

$$C_\mathrm{total} = \frac{1}{T_\mathrm{f}} \sum_{n=1}^{\infty} c_n p_\mathrm{L}(n) + \frac{c_\mathrm{u}}{T_\mathrm{p}}. \tag{11}$$

For the sake of simplicity, we will reasonably assume that the cost of loosing the nth call after HLR failure is independent of n. Thus, $\forall n,\ c_n = nc_1$. Consequently,

$$C_\mathrm{total} = \frac{c_1}{T_\mathrm{f}} \sum_{n=1}^{\infty} n p_\mathrm{L}(n) + \frac{c_\mathrm{u}}{T_\mathrm{p}} = \frac{c_1}{T_\mathrm{f}} E_\mathrm{loss} + \frac{c_\mathrm{u}}{T_\mathrm{p}}. \tag{12}$$

We further normalize the C_total, so that it is expressed in units of c_u. Thus,

$$C_\mathrm{total}[c_\mathrm{u}] = \frac{c_1[c_\mathrm{u}]}{T_\mathrm{f}} E_\mathrm{loss} + \frac{1}{T_\mathrm{p}}. \tag{13}$$

5. Discussion

$C_\mathrm{total}[c_\mathrm{u}]$ is shown in figures 11–14, as a function of $c_1[c_\mathrm{u}]$, $T_\mathrm{f}[1/\lambda_\mathrm{a}]$, $T_\mathrm{p}[1/\lambda_\mathrm{a}]$, and $\lambda_\mathrm{u}[\lambda_\mathrm{a}]$. The default values of parameters in these figures[3] are $\lambda_\mathrm{u} = 1$, $c_1 = 1000$, $T_\mathrm{p} = 1$ and $T_\mathrm{f} = 1{,}000$. In our evaluation, we used the following ranges for the independent parameters: $\lambda_\mathrm{u} \in [0.01, 100]$, $c_1 \in [100, 10^4]$, $T_\mathrm{p} \in [0.1, 10^3]$ and $T_\mathrm{f} \in [10, 10^5]$. These ranges of parameters' values were chosen based on the assumption that in most practical situations 10[min] $\leqslant \lambda_\mathrm{a} \leqslant$ 1[day]. Thus, for example, it will be reasonable to assume that the database failure will not be more frequent than once an hour or that the periodic update interval will not be more frequent than every minute and will occur at least once a week. Similarly, if one postulates that the cost of an update is on the order of 1 cent, than the "cost" of loosing a call can be reasonably assumed to be in the range of 1 $ to 100 $.

Our first observation from graph in figure 11 is that, unless the HLR is extremely unreliable (i.e., as long as $T_\mathrm{f} > 100$), the total cost is not very strongly affected by

[3] I.e., the values of undisplayed independent parameters.

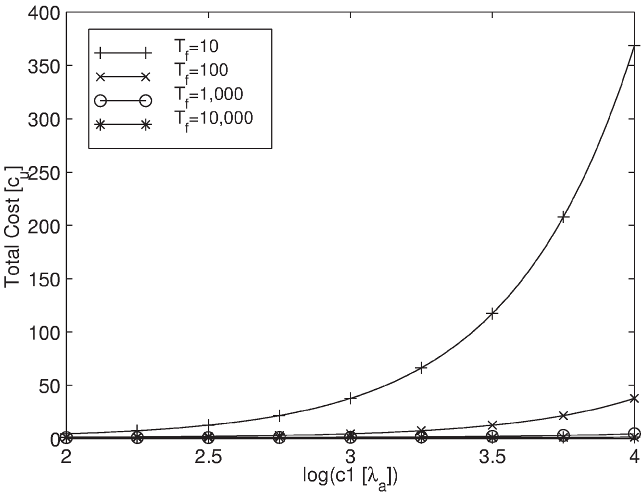

Figure 11. Average cost of a failure as a function of c_1.

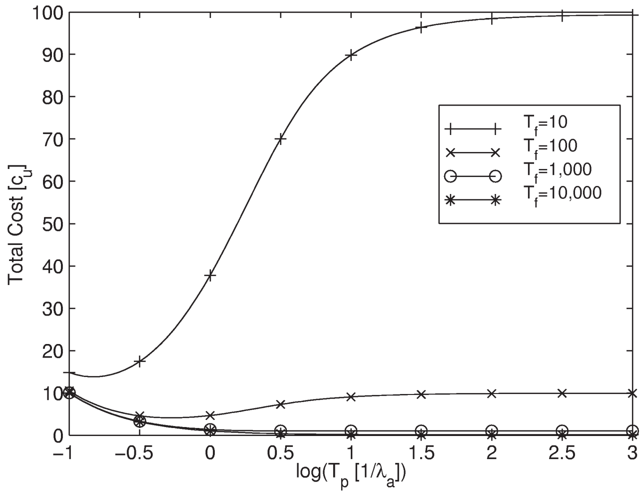

Figure 13. Average cost of a failure as a function of T_p.

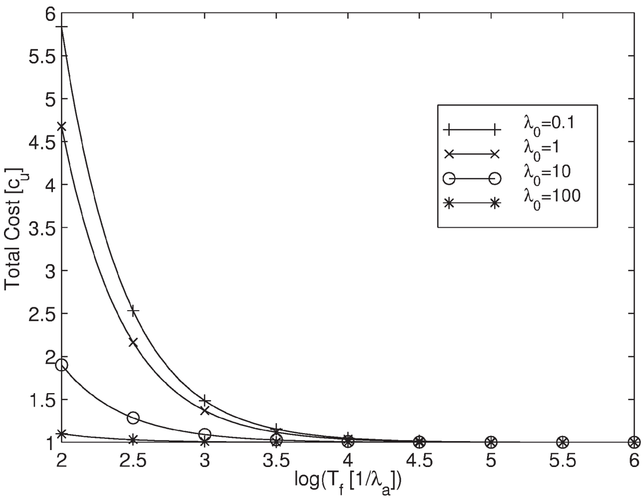

Figure 12. Average cost of a failure as a function of T_f.

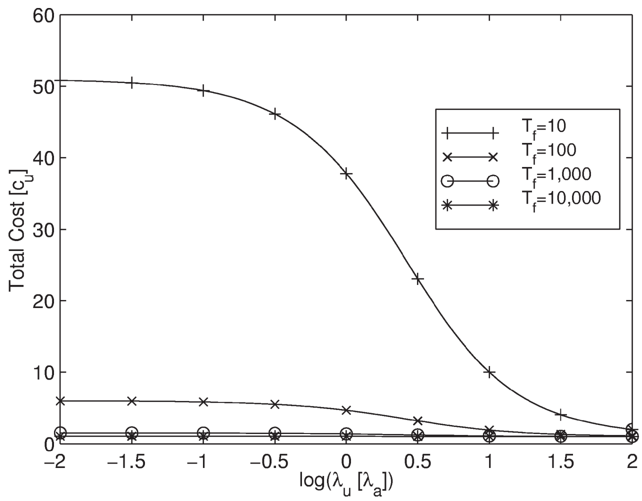

Figure 14. Average cost of a failure as a function of λ_u.

the actual value of c_1. For very unreliable systems, the total cost is highly dependent on the value of c_1, resulting in a considerable increase in the cost for larger c_1.

As indicated in figure 12, as long as the T_f is larger than $10{,}000/\lambda_\mathrm{a}$, the total cost remains independent of the actual value of λ_u (and, thus, of λ_o as well) and of T_p (for the latter, see figure 13). Thus, if the system parameters are optimized, the cost remains relatively stable and small for moderate to highly reliable systems.

The behavior of the cost as a function of T_p depends on the value of T_f. For a reliable HLR, larger values of T_p are preferred, as the frequency of loosing calls due to HLR failure is small and large T_p reduces the cost of periodic update. However, as the system gets more unreliable, smaller values of T_p correspond to smaller total cost (figure 13). We claim that a good setting for T_p is to be approximately equal to the value of call interarrival time, with larger values for a more reliable system.

Finally, as a function of λ_u (figure 14), the cost is always smaller with an increase in λ_u. The cost is much more sensitive to λ_u when the system is fairly unreliable. We conclude that, as the system is at least moderately reliable, the call origination rate and the rate of LA crossing has only minor effect on the system cost, *provided that the update period was set up as recommended above*. Thus, for at least moderately reliable system and properly set up update period, the determination of the size of the LA can be made independent from the HLR failure consideration. This conclusion is in particular true for LMS systems with balanced in and out traffic.

6. Summary

We have presented here a framework for evaluation of the cost of mobility management in the presence of unreliable HLR. Upon failure of an HLR, the HLR database needs resynchronization, which is achieved either by periodic location updates performed by the mobile, by the mobile's originating a call, or by the mobile's registration due to its crossing of an LA boundary. The longer the HLR restoration lasts, there is larger probability that calls will be lost.

The tradeoff is between the cost of location updates and the penalty due to lost calls (reduction of which requires more frequent periodic location updates).

We have showed how the average number of lost calls can be computed and we have set up a cost function that encompasses the tradeoff between location update cost and the penalty of lost calls. The evaluation of the cost function for different practical parameter ranges was performed. Our results indicate that, for a reliable system, the net cost is relatively insensitive to the value of the penalty of a single lost call. The cost is also small and relatively constant for a reliable system (MTBF larger than 10,000 times average call interarrival time). A good rule-of-tumb would be that, if the reliability of the system is unknown, the location update should be on the order of the average interarrival time. For a more unreliable system, the values of the location update interval should be smaller than the average interarrival time. Finally, we have demonstrated that unless the system is highly unreliable, the cost is relatively insensitive to the call origination and to the LA-crossing rates. Thus, in such cases, the design of the LA size can be made independent of the HLR reliability considerations.

References

[1] I.F. Akyildiz and J.S.M. Ho, Dynamic mobile user location update for wireless PCS networks, Wireless Networks 1 (1995) 187–196.

[2] A. Bar-Noy, I. Kessler and M. Sidi, Mobile users: To update or not to update?, Wireless Networks 1 (1995) 175–186.

[3] EIA/TIA, Cellular radio-telecommunications intersystem operation: Automatic roaming, Technical Report IS-41.3-B, EIA/TIA (1991).

[4] ETSI/TC, Restoration Procedures, Version 4.2.0, Technical Report Recommendation GSM 03.07, ETSI (1993).

[5] R.A. Guerin, Channel occupancy time distribution in a cellular radio system, IEEE Transactions on Vehicular Technology 35(3) (August 1987).

[6] R. Jain and Y.-B. Lin, An auxiliary user location strategy employing forwarding pointers to reduce network impacts of PCS, Wireless Networks 1 (1995) 197–210.

[7] Y.-B. Lin, Failure restoration of mobility databases for personal communication networks, Wireless Networks 1 (1995) 365–372.

[8] A.R. Noerpel, L.F. Chang and Y.-B. Lin, Polling deregistration for unlicensed PCS, IEEE Journal on Selected Areas in Communications 14(4) (May 1996) 728–734.

[9] C. Rose, Minimizing the average cost of paging and registration: A timer-based method, IEEE ICC (1995).

[10] C. Rose and R. Yates, Minimizing the average cost of paging under delay constraints, Wireless Networks 1 (1995) 211–219.

[11] E.J. Watson, *Laplace Transforms and Applications* (Birkhäuser, Basel, 1981).

[12] H. Xie, S. Tabbane and D.J. Goodman, Dynamic location area management and performance analysis, in: *IEEE Vehicular Technology Conference* (1993).

Zygmunt J. Haas received his B.Sc. in EE in 1979 and M.Sc. in EE in 1985. In 1988, he earned his Ph.D. from Stanford University and subsequently joined AT&T Bell Laboratories in the Network Research Department. There he pursued research on wireless communications, mobility management, fast protocols, optical networks, and optical switching. From September 1994 till July 1995, Dr. Haas worked for the AT&T Wireless Center of Excellence, where he investigated various aspects of wireless and mobile networking, concentrating on TCP/IP networks. As of August 1995, he joined the faculty of the School of Electrical Engineering at Cornell University as Associate Professor. Dr. Haas is an author of numerous technical papers and holds twelve patents in the fields of high-speed networking, wireless networks, and optical switching. He has organized several Workshops, delivered tutorials at major IEEE and ACM conferences, and serves as editor of several journals. He has been a guest editor of three IEEE JSAC issues ("Gigabit Networks", "Mobile Computing Networks" and "Ad-Hoc Networks"). He also served as a co-chair of ACM/IEEE Mobicom '96. Dr. Haas is a Senior Member of IEEE, a voting member of ACM, and a secretery of the IEEE Technical Committee on Personal Communications. His interests include: mobile and wireless communication and networks, personal communication service, and high-speed communication and protocols.
E-mail: haas@ee.cornell.edu
http://www.ee.cornell.edu/~haas/wn1.html

Yi-Bing Lin received his BSEE degree from National Cheng Kung University in 1983, and his Ph.D. degree in computer science from the University of Washington in 1990. From 1990 to 1995, he was with the Applied Research Area at Bell Communications Research (Bellcore), Morristown, NJ. In 1995, he was appointed as a Professor in the Department of Computer Science and Information Engineering (CSIE), National Chiao Tung University (NCTU). In 1996, he was appointed as Deputy Director of Microelectronics and Information Systems Research Center, NCTU. Since 1997, he has been Chairman of CSIE, NCTU. His current research interests include design and analysis of personal communications services network, mobile computing, distributed simulation, and performance modeling. Dr. Lin is a subject area editor of the Journal of Parallel and Distributed Computing, an associate editor of the International Journal on Computer Simulation, an associate editor of IEEE Network, an associate editor of the Simulation magazine, an area editor of ACM Mobile Computing and Communication Review, a columnist of ACM Simulation Digest, a member of the editorial board of International Journal of Communications Systems, an editor of Journal of Information Science and Engineering, a member of the editorial board of ACM/URSI/Baltzer Wireless Networks, a member of the editorial board of Computer Simulation Modeling and Analysis, Program Chair for the 8th Workshop on Distributed and Parallel Simulation, General Chair for the 9th Workshop on Distributed and Parallel Simulation. Program Chair for the 2nd International Mobile Computing Conference, the publicity chair of ACM Sigmobile, Guest Editor for the ACM/Balzer MONET special issue on Personal Communications, and Guest Editor for IEEE Transactions on Computers special issue on Mobile Computing.
E-mail: liny@csie.nctu.edu.tw