On Optimizing the Location Update Costs in the Presence of Database Failures

Zygmunt J. Haas
School of Electrical Engineering, Cornell University, Ithaca, NY, 14853
haas@anise.ee.cornell.edu, http://www.ee.cornell.edu/∼haas/wnl.html

Yi-Bing Lin
Dept. Comp. Sci. and Info. Engr., National Chiao Tung University, Hsinchu, Taiwan, R.O.C.
liny@csie.nctu.edu.tw

Abstract

This paper studies the database failure recovery procedure for cellular phone networks as part of the EIA/TIA IS-41 standard. 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

The current location of a mobile phone is maintained in today's mobile phone systems by a two-level hierarchical strategy with two types of databases. 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 *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.

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. 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. The new visited MSC then informs the mobile phone of the successful registration. 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.

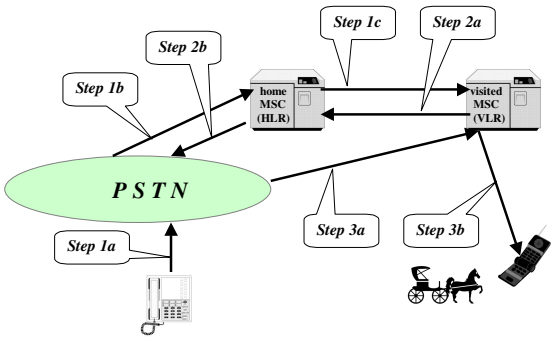

Figure 1: The call setup procedure

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 in the public switched telephone network (PSTN). Referring to Fig. 1, the call is routed to the mobile's home MSC (Step 1). The home MSC, after verifying that the mobile is "away from home" by consulting its HLR, retrieves the current mobile's location from its HLR. The home MSC then queries the current mobile's VLR, which returns a routable address (the *Temporary Directory Number*) to the originating switch in the PSTN through the home MSC (Step 2). Based on the routable address, a voice circuit is then set up from the originating switch to the mobile phone through the visited MSC (Step 3).

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 and we assume in this paper that after an HLR failure, the location information is lost. 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 call, which reveals the current location of the mobile in the network, or
- by the mobile's crossing a LA boundary, which prompts reregistration.

In particular, In EIA/TIA IS-41 [1], 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 VLR-s. The VLR-s 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 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. 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 failed HLR in such cases, a mechanism called *periodic location updating*, also referred to here as a as location confirmation, ([2]) was introduced, which reduces the HLR restoration delay. A mobile phone periodically establishes radio contact with the network to confirm its location.

In this paper, we study the effect of the HLR failures on various system's parameters in an IS-41 system. 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.

The following are assumption 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 $t_{sojourn} = \frac{9}{3+2\sqrt{3}}\frac{R}{V}$, where R is the "radius" of a LA and V is the average mobile's velocity. This assumption is based on [3], 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 $\lambda_c = 1/t_{sojourn}$.

2 HLR Failure Evaluation

Fig. 2 illustrates the timing diagram before and after an HLR failure. Due to the memoryless property of the exponential

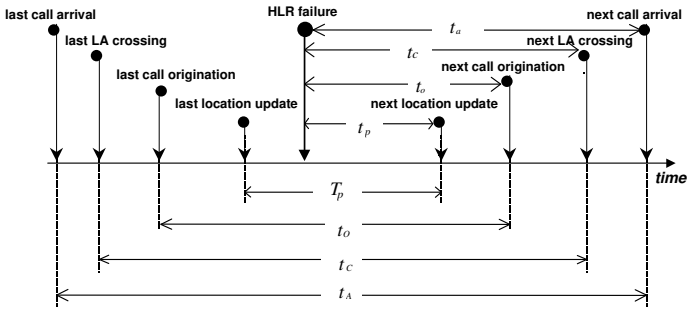

Figure 2: The timing diagram for deriving $p_L(n)$

distributions, t_o, t_a, and t_c have the same distributions as t_O, t_A, and t_C, respectively, whereas:

- t_O is the time interval between two consecutive call originations from the mobile phone,
- t_A is the time interval between two consecutive call arrivals to the mobile phone,
- t_C is the time interval between two LA crossings by the mobile phone,
- t_o is the time interval between the HLR failure and the next call origination from the mobile phone,
- t_a is the time interval between the HLR failure and the next call arrival to the mobile phone, and
- t_c: the time interval between the HLR failure and the next LA crossing by the mobile phone.

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 $1/\lambda_o$. We label by T_p the constant time interval between two periodic location updates. It can be shown that [4] 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 (the time interval between the HLR failure and the next mobile phone periodic location update) has a uniform distribution in the interval $[0, T_p]$. That is, the density function f_p of t_p is: $f_p(t_p) = \frac{1}{T_p}, \quad 0 \le t_p \le T_p$. After an HLR failure, the existence of the mobile phone is known to the HLR at time $t_r = \min(t_o, t_c, t_p) = \min(min(t_o, t_c), t_p) = \min(t_u, t_p)$, whereas $t_u = \min(t_c, t_o)$ is the time interval between the HLR failure and one of the following two events, whichever occurs first: next LA crossing, or next call origination.

As both of the random processes, the call origination and the crossing of LA boundaries, are independent and exponentially-distributed, the probability density function of the process t_u, $f_u(t_u)$, is:

$$f_u(t_u) \quad = \quad ... = (\lambda_o + \lambda_c) e^{-(\lambda_o + \lambda_c) t_u} = \lambda_u e^{-\lambda_u t_u} \ , \quad (1)$$

whereas $\lambda_u = \lambda_o + \lambda_c$. Thus, the process is also exponential with the rate equal to the sum of the origination and LA crossing rates: $\lambda_u = \lambda_o + \lambda_c$. Since $0 \le t_p \le T_p$, it is clear that $0 \le t_r \le T_p$. The density function, f_r, of t_r is then:

$$\begin{aligned} f_r(t_r) \quad &= \quad \int_{t_p=t_r}^{T_p} f_u(t_r) f_p(t_p) dt_p + \int_{t_u=t_r}^{\infty} f_u(t_u) f_p(t_r) dt_u \\ &= \quad ... = \begin{cases} [\frac{\lambda_u (T_p - t_r) + 1}{T_p}] e^{-\lambda_u t_r} & 0 \le t_r \le T_p \\ 0 & \text{otherwise} \end{cases} \quad (2) \end{aligned}$$

and the Laplace Transform of f_r, f_r^*, is:

$$f_r^*(s) \quad = \quad \frac{\lambda_u}{\lambda_u + s} + \left[\frac{s}{T_p (\lambda_u + s)^2} \right] \left[1 - e^{-(\lambda_u + s) T_p} \right] \quad (3)$$

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:

$$\Pr[K = k | X = t] = \frac{(\lambda_a t)^k}{k!} e^{-\lambda_a t} \quad . \quad (4)$$

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

$$f_L(k, t_r) = \frac{(\lambda_a t_r)^k}{k!} e^{-\lambda_a t_r} f_r(t_r) \quad . \quad (5)$$

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

$$\begin{aligned} p_L(k) \quad &= \quad \int_{t_r=0}^{\infty} f_L(k, t_r) dt_r = ... \\ &= \quad \frac{\lambda_a^k}{k!} \left[(-1)^k \frac{d^k}{ds^k} f_r^*(s) \right] \Big|_{s=\lambda_a} \quad . \quad (6) \end{aligned}$$

From equations (3) and (6), we can compute the value of $p_L(k)$ for any k. For example,

$$\begin{aligned} p_L(1) \quad &= \quad \frac{\lambda_u \lambda_a}{(\lambda_u + \lambda_a)^2} + \left[\frac{\lambda_a (\lambda_a - \lambda_u)}{T_p (\lambda_u + \lambda_a)^3} \right] \cdot \quad (7) \\ &\quad \left[1 - e^{-(\lambda_u + \lambda_a) T_p} \right] \left[\frac{\lambda_a^2}{(\lambda_u + \lambda_a)^2} \right] e^{-(\lambda_u + \lambda_a) T_p} \ . \end{aligned}$$

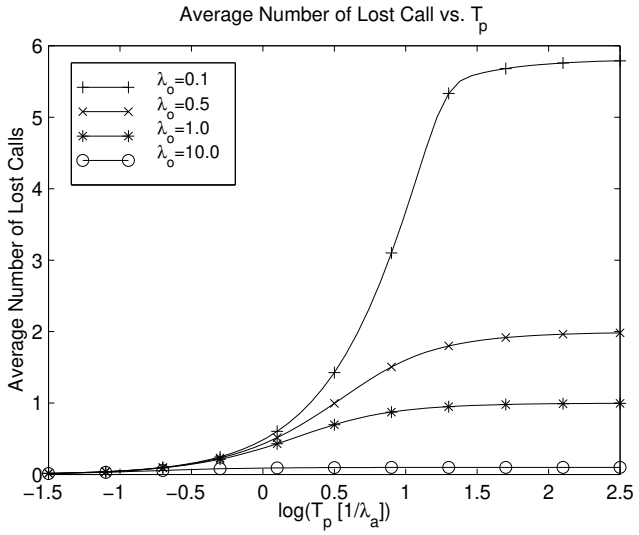

Figure 3: Average number of lost calls as a function of T_p

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

$$\begin{aligned} E_{loss} \quad &= \quad \sum_{i=0}^{\infty} i \cdot p_L(i) = \sum_{i=0}^{\infty} i \frac{\lambda_a^i}{i!} [(-1)^i \frac{d^i}{ds^i} f_r^*(s)] |_{s=\lambda_a} \\ &= \quad \{ e^{(-\lambda_a \frac{d}{ds})} \cdot [-\lambda_a \frac{d}{ds} f_r^*(s)] \} |_{s=\lambda_a} \quad . \quad (8) \end{aligned}$$

E_{loss} can be calculated for given values of λ_o, λ_c, λ_a, and T_p using equations (3), (6), (8), and the definition of λ_u.

The crossing rate of LA boundaries, λ_c, strongly depends on the the size of the LA-s 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 to evaluate λ_c. This value is then used in substitution of $\lambda_c + \lambda_o$ 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 LA-s 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 a LA is much larger than 1. Thus, $\lambda_u \approx \lambda_o$. We refer here to such a system as a *Low Mobility System (LMS)*.

E_{loss} has been numerically evaluated and the results are depicted in Figs. 3 and 4, as a function of T_p and λ_u.

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 $\lambda_u = 1[\lambda_a]$, 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 $T_p \le 10$, as can be seen from Fig. 4.

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). How-

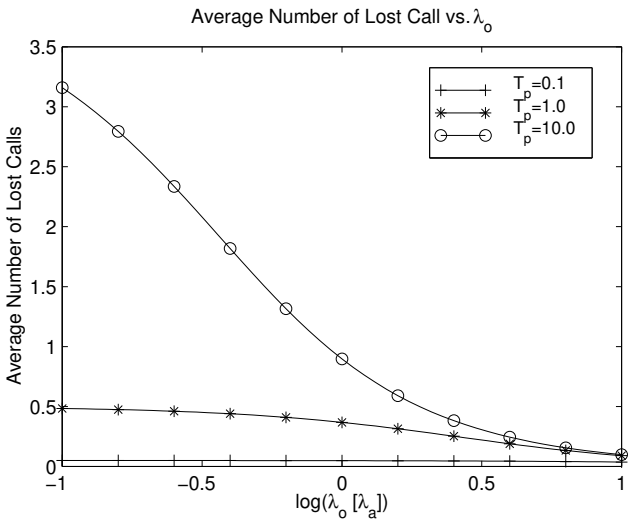

Figure 4: Average number of lost calls as a function of λ_u

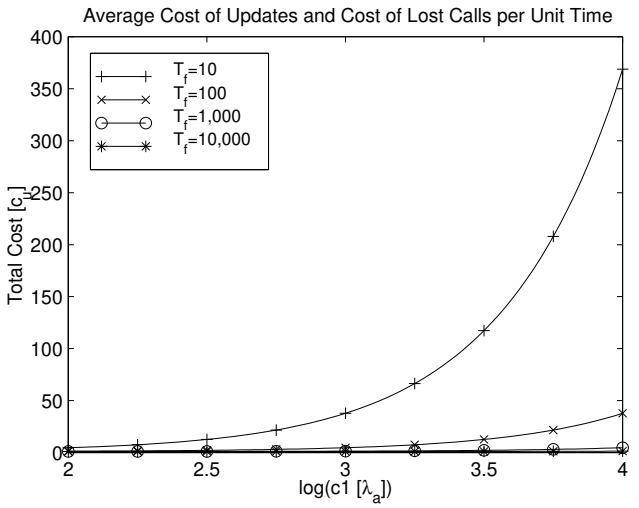

Figure 5: Average cost of a failure as a function of c_1

ever, 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 (i.e., Mean Time Between Failures or MTBF), 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:

$$C_{total} = \frac{1}{T_f} \sum_{n=1}^{\infty} n \cdot c_n \cdot p_L(n) + \frac{c_u}{T_p} \quad . \tag{9}$$

For the sake of simplicity, we will reasonably assume that the cost of loosing the n^{th} call after HLR failure is independent of n. Thus, $\forall n$, $c_n = n \cdot c_1$. Consequently,

$$C_{total} = \frac{c_1}{T_f} \sum_{n=1}^{\infty} n p_L(n) + \frac{c_u}{T_p} = \frac{c_1}{T_f} E_{loss} + \frac{c_u}{T_p} \quad . \tag{10}$$

We further normalize the C_{total}, so that it is expressed in units of c_u. Thus,

$$C_{total}[c_u] = \frac{c_1[c_u]}{T_f} E_{loss} + \frac{1}{T_p} \quad . \tag{11}$$

3 Discussion

$C_{total}[c_u]$ is shown in Figs. 5-8, as a function of $c_1[cu]$, $T_f[1/\lambda_a]$, $T_p[1/\lambda_a]$, and $\lambda_u[\lambda_a]$. The default values of parameters in these figures[1] are: $\lambda_u = 1$, $c_1 = 1000$, $T_p = 1$, and $T_f = 1,000$. In our evaluation, we used the following ranges for the independent parameters: $\lambda_u \in [0.01, 100]$, $c_1 \in [100, 10^4]$, $T_p \in [0.1, 10^3]$, and $T_f \in [10, 10^5]$. These ranges of parameters' values were chosen based on the assumption that in most practical situations $10[\text{min}] \leq \lambda_a \leq 1[\text{day}]$. 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 Fig. 5 is that, unless the HLR is extremely unreliable (i.e., as long as $T_f > 100$), the total cost is not very strongly affected by 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 Fig. 6, as long as the T_f is larger than $10,000/\lambda_a$, 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 Fig. 7). 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 (Fig. 7). 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 (Fig. 8), 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

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

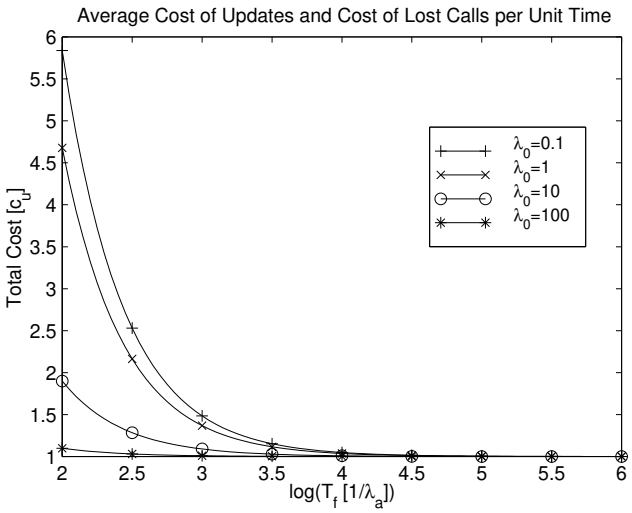

Figure 6: Average cost of a failure as a function of T_f

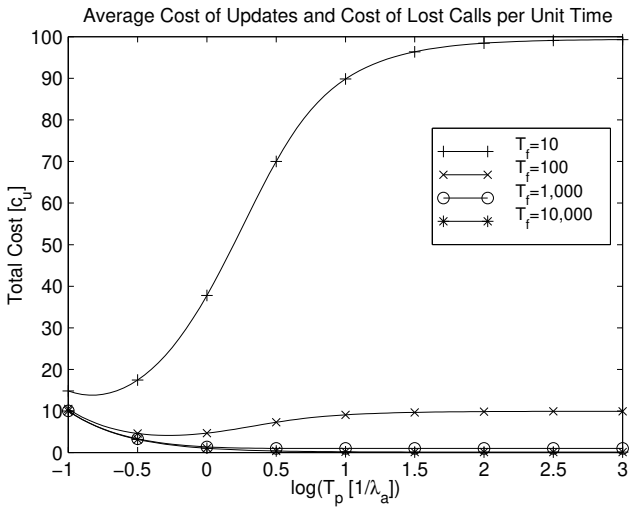

Figure 7: Average cost of a failure as a function of T_p

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.

4 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 update performed by the mobile, by the mobile's originating a call, or by the mobile's registration due to its crossing of a 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 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

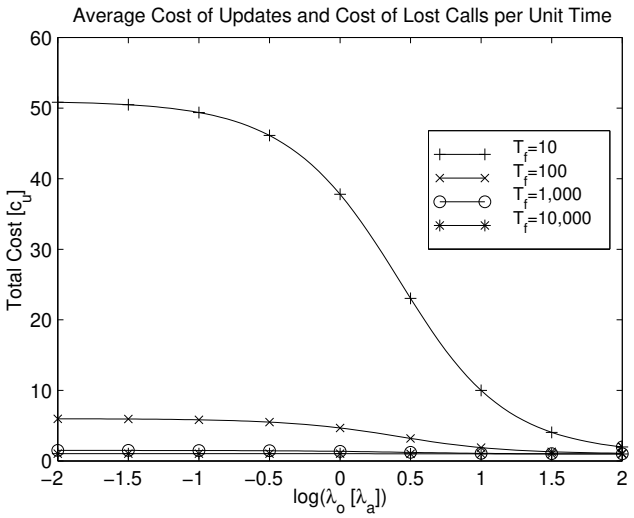

Figure 8: Average cost of a failure as a function of λ_o

the HLR reliability considerations.

References

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

[2] ETSI/TC, "Restoration Procedures," Version 4.2.0, Technical Report Recommendation GSM 03.07, ETSI, 1993.

[3] R. A. Guerin, "Channel Occupancy Time Distribution in a Cellular Radio System," *IEEE Trans. on Vehicular Tech.,* vol. VT-35, no. 3, August 1987.

[4] A. R. Noerpel, L. F. Chang, and Y.-B. Lin, "Polling deregistration for unlicensed PCS," *IEEE Journal on Selected Areas in Communications,* vol.14, no.4, pp. 728-34, May 1996.