

Resolving Mobile Database Overflow With Most Idle Replacement

Hui-Nien Hung, Yi-Bing Lin, *Senior Member, IEEE*, Nan-Fu Peng, and Shun-Ren Yang

Abstract—In a personal communications service (PCS) network, mobility databases called visitor location registers (VLRs) are utilized to temporarily store the subscription data and the location information for the roaming users. Because of user mobility, it is possible that the VLR is full when a mobile user arrives. Under such a circumstance, the incoming user has no VLR record and thus cannot receive PCS services. This issue is called VLR overflow. To resolve the VLR overflow problem, a VLR record can be selected for replacement when the VLR is full and then the reclaimed storage is used to hold the record of the requesting user. This paper considers the most idle replacement policy to provide services to mobile users without VLR records. In this policy, the record with the longest idle time is selected for replacement. We propose an analytic model to investigate the performance of this replacement policy. The analytic results are validated against simulation experiments. The results indicate that our approach effectively resolves the VLR overflow problem.

Index Terms—Database overflow, mobility management, personal communications service, visitor location register.

I. INTRODUCTION

ADVANCES in wireless communication technologies allow a personal communications service (PCS) network to provide flexible broadband telecommunication services to users with mobility. Through roaming agreements, mobile users can visit various PCS networks and receive services in the visited network. To support user mobility, PCS networks adopt a *two-level* database architecture that consists of two types of databases: *home location register* (HLR) and *visitor location register* (VLR). This architecture was proposed in 2G mobile networks [10] and has been followed by the 3G UMTS specifications [1]. In this architecture, when a user subscribes to the services of a PCS operator (called the *home system* of the user), a record is created in the operator's HLR. The record stores the user's subscription data as well as location information. A VLR of a PCS network temporarily stores the subscription data and

the location information for the users currently residing in the visited PCS network. The VLR is associated with one or more mobile switches. When a call arrives to a visiting user, the VLR is queried to retrieve necessary information to set up the call.

A PCS network is partitioned into several location areas. Every VLR controls some of the location areas. When a mobile station (MS) moves to a new location area, it informs the visited VLR and the HLR of the new location. This procedure is referred to as *registration* and is described in the following steps.

- Step 1.1 When an MS moves from a VLR to another VLR, the MS sends a registration message to the new VLR. The new VLR then contacts with the HLR to update the location information of the MS.
- Step 1.2 If the location update operation is successful, the HLR sends an acknowledgment to the new VLR. This message includes the subscription information of the MS. The VLR creates a record for the MS to store the subscription information received from the HLR and sends an acknowledgment to inform the MS of the successful registration.

To originate a call from a mobile user, the following steps are performed.

- Step 2.1 The dialed telephone number of the called party is sent to a mobile switch through radio interface.
- Step 2.2 When the mobile switch receives the call setup message, it analyzes the request and consults the VLR to check whether the call can be accepted.
- Step 2.3 If the call request is accepted, the mobile switch sets up the voice connection to the called party.

To deliver a call to a mobile user in a PCS network, the following steps are performed.

- Step 3.1 Someone dials the telephone number of a mobile user. The call request is forwarded to a gateway mobile switch.
- Step 3.2 The gateway mobile switch asks the HLR for routing information. The HLR then queries the corresponding VLR to obtain a routable address.
- Step 3.3 The VLR searches the mobile user's record. Based on the location information, the VLR generates the routable address and sends it back to the gateway mobile switch.
- Step 3.4 The gateway mobile switch then sets up the voice connection to the serving mobile switch. The serving mobile switch pages the MS in the location area indicated by the VLR. Finally, the call path to the MS is established.

Manuscript received December 1, 2000; revised June 1, 2001. H.-N. Hung was supported by National Science Council under Contract NSC 89-2118-M009007. Y.-B. Lin was supported in part by the MOE Program of Excellence Research under Contract 89-E-FA04-4, in part by TAHOE Network, in part by Ericsson, in part by InterVideo, in part by FarEastone, in part by National Science Council under Contract NSC 89-2213-E-009-203, and in part by the Lee and MTI Center for Networking Research, NCTU.

H.-N. Hung and N.-F. Peng are with the Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan, R.O.C.

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

S.-R. Yang is with the Department of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu, Taiwan, R.O.C. (e-mail: sjyoun@csie.nctu.edu.tw).

Publisher Item Identifier S 0733-8716(01)08466-9.

Details of mobility management and call setup procedures can be found in [2], [10], and [12].

Since a VLR provides services to visiting users from various PCS networks, the number of records in a VLR changes dynamically. Furthermore, a VLR may overflow if the number of visiting users is larger than the capacity of the VLR. If the VLR is full when a mobile user arrives, the user cannot register using the standard registration procedure and thus cannot receive PCS services. In this case, the VLR is *overflow*, and the incoming users (who cannot receive services) are referred to as the *overflow users*.

To resolve the VLR overflow issue, we proposed a scheme in [8], which modifies registration, mobile call origination, and mobile call delivery procedures as follows.

Modified Registration: If the VLR is full at Step 1.2 for an incoming user u , a record is selected for replacement. The selected record is deleted, and its storage is then used to store u 's information. In this case, the user of the replaced record becomes an overflow user. Alternatively, user u may be considered as the overflow user, and no record replacement is required.

Modified Mobile Call Origination: When an overflow user u attempts to make a call, the VLR cannot find the corresponding record at Step 2.2 and declines the call request. The MS performs an overflow registration operation to create a record for u . (Note that, in this registration, the user u cannot be selected as the overflow user.) The MS then resends the call request, and the normal call origination procedure is executed to set up the call.

Modified Mobile Call Delivery: When an incoming call arrives to an overflow user u , the HLR identifies that u is an overflow user and queries the VLR to obtain the routing information at Step 3.2. The query message includes the user profile of u . Through a replacement at Step 3.3, the VLR creates a record for u to store the received user profile from the HLR. The VLR also generates a routable address of u and sends it back to the HLR at Step 3.3.

As indicated in the modified procedures, for an overflow user, the call setup procedures are more expensive than that for a normal mobile call setup. Therefore, it is desirable to reduce the possibility of overflow call setup. In [8], we proposed the *random replacement policy* where each record in the VLR is selected for replacement with the same probability. We observed that this policy is only suitable for an environment where the mobile users have relatively low call activities. Consider a mobile network with high and low call activity users. If the VLR is fully occupied by low call activity users, then the telephone circuits may be idle while the high call activity users (who need to make calls) are blocked because they cannot register into the VLR. Thus, replacing low call activity users by high call activity users will improve the utilization of the switch. In [9], we proposed the *inactive replacement policy* to reduce the overflow call setups in a VLR area where the call activities of visiting users vary significantly. In this policy, a period called *inactive threshold* is used to check if a user is not active. If a user has no call activity during the inactive threshold, then he/she is considered inactive and the VLR record can be selected for replacement. The inactive replacement policy attempts to select users with low call activities for replacement. Based on the same phi-

losophy, this paper proposes the *most idle replacement policy*. In this policy, when a replacement request arrives due to overflow registration or overflow call setup, the VLR record of the user who does not have call activities for the longest period (i.e., the most idle user) is selected for replacement. We develop analytic and simulation models to investigate the performance of this policy and compare it with the previously proposed approaches. Our study indicates that good performance can be achieved by the most idle replacement policy.

II. ANALYTIC MODEL FOR THE MOST IDLE REPLACEMENT POLICY

This section describes the analytic model for the most idle replacement policy. We assume that there are two classes of mobile users where class 1 users have low call activities and class 2 users have high call activities. Our model can be easily extended for the case where there are more than two classes of users. The details are omitted. The following assumptions are used in the model.

- The class 1 and class 2 user arrivals to a VLR area are Poisson processes [15] with rates $\lambda_{u,1}$ and $\lambda_{u,2}$, respectively.
- The call arrivals of class 1 and class 2 users are Poisson processes with rates $\lambda_{c,1}$ and $\lambda_{c,2}$, respectively. From the definitions of high and low call activities, we have $\lambda_{c,2} \gg \lambda_{c,1}$.
- The VLR residence time distributions for both classes are the same, which have a general distribution with distribution function $F_m(\cdot)$, mean $1/\lambda_m$ and variance V_m .

The Poisson assumptions on user arrivals and call arrivals can be easily relaxed to accommodate general distributions in our simulation experiments. Based on the above assumptions, we are interested in the probability P_{MR} that after an overflow registration or call setup, the user of the replaced record does not have any call activity before he/she leaves the VLR. Probability P_{MR} is derived as follows.

The users in a VLR area can be divided into two groups (see Fig. 1): the *VLR* group and the *overflow* group. The users in the *VLR* group have records in the VLR database, while those in the *overflow* group do not. The notations used in Fig. 1 are explained as follows.

- Random variables N_1 and N_2 are the numbers of users in the *VLR* and *overflow* groups when the system is in equilibrium.
- Random variable N is the number of users residing in the VLR area in the steady state, i.e., $N = N_1 + N_2$.
- p (q) is the long term proportion of class 1 (class 2) users in the *VLR* group. We have $p + q = 1$.
- p_2 (q_2) is the long term proportion of class 1 (class 2) users in the *overflow* group. We have $p_2 + q_2 = 1$.
- Random variable K_1 (K_2) is the number of class 1 (class 2) users residing in the VLR area (either in the *VLR* or the *overflow* groups) when the system is in equilibrium.
- p_1 (q_1) is the long term proportion that a class 1 (class 2) user is selected for replacement when a replacement request arrives. We have $p_1 + q_1 = 1$.

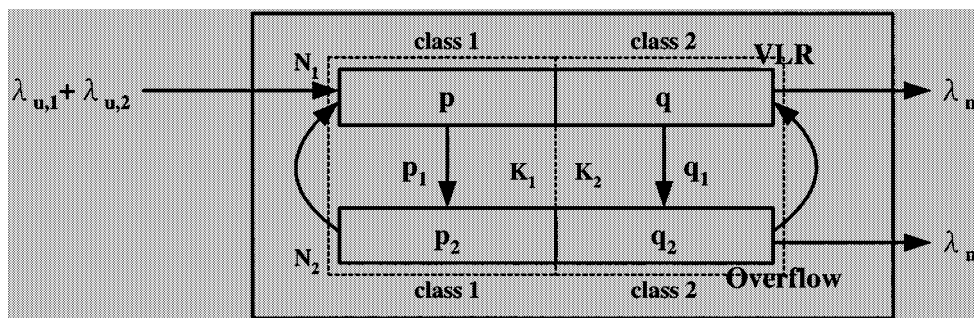


Fig. 1. The system diagram in equilibrium.

In the Appendix, we show that N , N_1 , N_2 , K_1 , and K_2 can be regarded as constants when the system reaches the steady state. In particular, we have [see (32) and (35)]

$$\begin{aligned} N &\simeq \frac{\lambda_{u,1} + \lambda_{u,2}}{\lambda_m}, & N_1 &\simeq L \\ N_2 &\simeq \frac{\lambda_{u,1} + \lambda_{u,2}}{\lambda_m} - L \\ K_1 &\simeq \frac{\lambda_{u,1}}{\lambda_m} \quad \text{and} \quad K_2 \simeq \frac{\lambda_{u,2}}{\lambda_m} \end{aligned} \quad (1)$$

where L is the size of the VLR database.

Since the input flow must be equal to the output flow for class 1 and class 2 users in the VLR group when the system is in equilibrium, we have the following two balance equations:

$$\begin{aligned} \overline{\lambda_{u,1}} &= \lambda_{u,1} + p_2 N_2 \lambda_{c,1} \\ &= p N_1 \lambda_m + a (\overline{\lambda_{u,1}} + \overline{\lambda_{u,2}}) p_1 \end{aligned} \quad (2)$$

$$\begin{aligned} \overline{\lambda_{u,2}} &= \lambda_{u,2} + q_2 N_2 \lambda_{c,2} \\ &= q N_1 \lambda_m + a (\overline{\lambda_{u,1}} + \overline{\lambda_{u,2}}) q_1 \end{aligned} \quad (3)$$

where a is the probability that an incoming user finds the VLR database full, and $\overline{\lambda_{u,1}}$ ($\overline{\lambda_{u,2}}$) is the arrival rate of class 1 (class 2) users to the VLR database. In equation (i) ($i = 2, 3$), the input flow includes the arrivals of class i users from other VLR areas (i.e., $\lambda_{u,i}$) and those from the overflow group due to call setups (i.e., $p_2 N_2 \lambda_{c,1}$ for class 1 users and $q_2 N_2 \lambda_{c,2}$ for class 2 users). On the other hand, the output flow includes the departures of class i users to other VLR areas (i.e., $p N_1 \lambda_m$ for class 1 users and $q N_1 \lambda_m$ for class 2 users) and those joining in the overflow group due to record replacement. The record replacements occur when the arrivals (i.e., $\overline{\lambda_{u,1}} + \overline{\lambda_{u,2}}$) find that the VLR database is full. Thus, the departure rates of class 1 and class 2 users due to record replacements are $a(\overline{\lambda_{u,1}} + \overline{\lambda_{u,2}})p_1$ and $a(\overline{\lambda_{u,1}} + \overline{\lambda_{u,2}})q_1$, respectively.

From (2) and (3), we obtain

$$a = 1 - \frac{N_1 \lambda_m}{\lambda_{u,1} + \lambda_{u,2}}. \quad (4)$$

From (4), (2) and (3) can be combined as follows:

$$\begin{aligned} \lambda_{u,1} + p_2 N_2 \lambda_{c,1} &= p N_1 \lambda_m + \left(1 - \frac{N_1 \lambda_m}{\lambda_{u,1} + \lambda_{u,2}}\right) \\ &\quad \times (\overline{\lambda_{u,1}} + \overline{\lambda_{u,2}}) p_1 \end{aligned} \quad (5)$$

where $\overline{\lambda_{u,1}} + \overline{\lambda_{u,2}} = \lambda_{u,1} + p_2 N_2 \lambda_{c,1} + \lambda_{u,2} + q_2 N_2 \lambda_{c,2}$.

For class 1 users, the input flow to a VLR area (i.e., $\lambda_{u,1}$) must be equal to the output flow from that VLR area (i.e., $p N_1 \lambda_m + p_2 N_2 \lambda_m$). Thus, we have another balance equation

$$\lambda_{u,1} = p N_1 \lambda_m + p_2 N_2 \lambda_m. \quad (6)$$

To proceed with the derivation of P_{MR} , we need to introduce the concept of *quantile* [13]. The q th quantile of a random variable X or of its corresponding distribution is denoted by ζ_q and is defined as the smallest number ζ satisfying $F_X(\zeta) \geq q$. For example, the median of a random variable X is the 0.5th quantile and is denoted by $\zeta_{0.50}$.

Consider the timing diagrams in Fig. 2. Suppose that a mobile user u_i of class i enters the VLR at time t_0 and a replacement request occurs at time t_1 . Assume that the last call of u_i before t_1 arrived at time t_2 . Let $t_{m,i} = t_1 - t_0$ and $t_{c,i} = t_1 - t_2$. Since the replacement request stream is a Poisson process, the request occurring at t_1 is a random observer. Thus, from the excess life theorem [15], $t_{m,i}$ has the density function

$$r_{m,i}(t_{m,i}) = \lambda_m [1 - F_m(t_{m,i})] \quad (7)$$

where $F_m(\cdot)$ is the distribution function of VLR residence time. Similarly, by the excess life theorem, $t_{c,i}$ has the density function

$$r_{c,i}(t_{c,i}) = \lambda_{c,i} [1 - F_{c,i}(t_{c,i})] = \lambda_{c,i} e^{-\lambda_{c,i} t_{c,i}} \quad (8)$$

where $F_{c,i}(\cdot)$ is the distribution function of inter-call arrival time of class i users, which is exponential with rate $\lambda_{c,i}$. Let Z_i be the idle time of u_i seen by the replacement request at t_1 . If $t_{m,i} < t_{c,i}$ [(a) in Fig. 2], then Z_i is $t_{m,i}$. Otherwise, if $t_{m,i} > t_{c,i}$ [(b) in Fig. 2], then Z_i is $t_{c,i}$. Therefore, we have

$$Z_i = \min(t_{m,i}, t_{c,i}). \quad (9)$$

Let \bar{p} (\bar{q}) be the long term proportion of class 1 (class 2) users in the VLR group to all class 1 (class 2) users in the VLR area, then we have

$$\bar{p} = \frac{N_1 p}{K_1} = \frac{\lambda_m N_1 p}{\lambda_{u,1}} \quad \text{and} \quad \bar{q} = \frac{N_1 q}{K_2} = \frac{\lambda_m N_1 q}{\lambda_{u,2}}. \quad (10)$$

Since the idle time of users in the overflow group must be longer than all users in the VLR group, the \bar{p} th quantile $\zeta_{\bar{p}}^{(1)}$ of Z_1 represents the longest idle time of those class 1 users in the VLR group, and $F_{Z_1}(\zeta_{\bar{p}}^{(1)}) = \bar{p}$. Similarly, the \bar{q} th quantile $\zeta_{\bar{q}}^{(2)}$ of Z_2 represents the longest idle time of those class 2 users in the VLR group, and $F_{Z_2}(\zeta_{\bar{q}}^{(2)}) = \bar{q}$. Thus, $\zeta_{\bar{p}}^{(1)}$ of Z_1 must equal or at

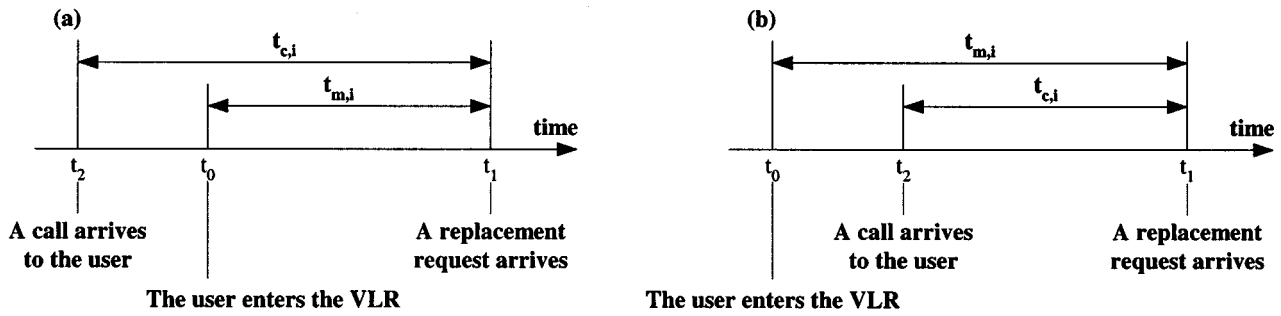


Fig. 2. Timing diagram I.

least approximate $\zeta_{\bar{q}}^{(2)}$ of Z_2 when the system is in equilibrium. Otherwise, for example, if $\zeta_{\bar{p}}^{(1)} > \zeta_{\bar{q}}^{(2)}$, then class 1 users would be selected for replacements for a long consecutive period (i.e., class 1 users will be replaced consecutively of order N times) and the system will not stay at stable state anymore. Based on the above discussion

$$\zeta_{\bar{p}}^{(1)} = \zeta_{\bar{q}}^{(2)}. \quad (11)$$

We consider two cases for (11).

Case 1. The VLR residence time is exponentially distributed with rate λ_m .

Case 2. The VLR residence time is lognormally [5], [13] distributed with mean $1/\lambda_m = e^{\mu+(1/2)\sigma^2}$ and variance $V_m = e^{2\mu+2\sigma^2} - e^{2\mu+\sigma^2}$, where μ is the alternative parameter and σ is the shape parameter. The lognormal distribution is selected because it can be used to approximate many other distributions as well as measured data [4], [6].

For exponentially distributed VLR residence time, we have density functions for Z_1 and Z_2 as follows:

$$f_{Z_1}(z_1) = (\lambda_{c,1} + \lambda_m) e^{-(\lambda_{c,1} + \lambda_m)z_1} \quad (12)$$

$$f_{Z_2}(z_2) = (\lambda_{c,2} + \lambda_m) e^{-(\lambda_{c,2} + \lambda_m)z_2}. \quad (13)$$

From (12) and (13), we have

$$\zeta_{\bar{p}}^{(1)} = - \left(\frac{1}{\lambda_{c,1} + \lambda_m} \right) \ln(1 - \bar{p})$$

and

$$\zeta_{\bar{q}}^{(2)} = - \left(\frac{1}{\lambda_{c,2} + \lambda_m} \right) \ln(1 - \bar{q}). \quad (14)$$

From (11) and (14)

$$- \left(\frac{1}{\lambda_{c,1} + \lambda_m} \right) \ln(1 - \bar{p}) = - \left(\frac{1}{\lambda_{c,2} + \lambda_m} \right) \ln(1 - \bar{q}). \quad (15)$$

We can use a numerical method (e.g., Newton–Raphson algorithm [11]) to solve p from (15), and then we can solve p_2 by (6). Finally, we can solve p_1 by (5).

For lognormally distributed VLR residence time with mean $e^{\mu+(1/2)\sigma^2}$ and variance $e^{2\mu+2\sigma^2} - e^{2\mu+\sigma^2}$, the distribution function is

$$\begin{aligned} F_m(x) &= \int_0^x \left(\frac{1}{y\sqrt{2\pi}\sigma} \right) \exp \left[-\frac{(\ln y - \mu)^2}{2\sigma^2} \right] dy \\ &= \Phi \left(\frac{\ln x - \mu}{\sigma} \right) \end{aligned} \quad (16)$$

where $\Phi(\cdot)$ is the standard normal distribution function. From (7)

$$r_{m,i}(t_{m,i}) = \lambda_m [1 - F_m(t_{m,i})] = \frac{\Phi \left(\frac{\mu - \ln t_{m,i}}{\sigma} \right)}{e^{\mu+(1/2)\sigma^2}}. \quad (17)$$

From (9), we have

$$\begin{aligned} \Pr(Z_1 > z) &= \Pr(t_{m,1} > z) \cdot \Pr(t_{c,1} > z) \\ &= e^{-\lambda_{c,1}z} \int_z^\infty \left[\frac{\Phi \left(\frac{\mu - \ln x}{\sigma} \right)}{e^{\mu+(1/2)\sigma^2}} \right] dx \end{aligned} \quad (18)$$

$$\begin{aligned} \Pr(Z_2 > z) &= \Pr(t_{m,2} > z) \cdot \Pr(t_{c,2} > z) \\ &= e^{-\lambda_{c,2}z} \int_z^\infty \left[\frac{\Phi \left(\frac{\mu - \ln x}{\sigma} \right)}{e^{\mu+(1/2)\sigma^2}} \right] dx. \end{aligned} \quad (19)$$

Substitute $\zeta_{\bar{p}}^{(1)}$ and $\zeta_{\bar{q}}^{(2)}$ for z s in (18) and (19), respectively

$$\begin{aligned} \Pr(Z_1 > \zeta_{\bar{p}}^{(1)}) \\ = 1 - \bar{p} &= e^{-\lambda_{c,1}\zeta_{\bar{p}}^{(1)}} \times \int_{\zeta_{\bar{p}}^{(1)}}^\infty \left[\frac{\Phi \left(\frac{\mu - \ln x}{\sigma} \right)}{e^{\mu+(1/2)\sigma^2}} \right] dx \end{aligned} \quad (20)$$

$$\begin{aligned} \Pr(Z_2 > \zeta_{\bar{q}}^{(2)}) \\ = 1 - \bar{q} &= e^{-\lambda_{c,2}\zeta_{\bar{q}}^{(2)}} \times \int_{\zeta_{\bar{q}}^{(2)}}^\infty \left[\frac{\Phi \left(\frac{\mu - \ln x}{\sigma} \right)}{e^{\mu+(1/2)\sigma^2}} \right] dx \end{aligned} \quad (21)$$

where

$$\begin{aligned} &\int_{\zeta_{\bar{p}}^{(1)}}^\infty \Phi \left(\frac{\mu - \ln x}{\sigma} \right) dx \\ &= \int_{\zeta_{\bar{p}}^{(1)}}^\infty \int_{\zeta_{\bar{p}}^{(1)}}^y \left(\frac{1}{y\sqrt{2\pi}\sigma} \right) \exp \left[-\frac{(\mu - \ln y)^2}{2\sigma^2} \right] dx dy \\ &= \int_{\zeta_{\bar{p}}^{(1)}}^\infty \left(\frac{y - \zeta_{\bar{p}}^{(1)}}{y\sqrt{2\pi}\sigma} \right) \exp \left[-\frac{(\mu - \ln y)^2}{2\sigma^2} \right] dy \\ &= A - \zeta_{\bar{p}}^{(1)} \int_{\zeta_{\bar{p}}^{(1)}}^\infty \left(\frac{1}{y\sqrt{2\pi}\sigma} \right) \exp \left[-\frac{(\mu - \ln y)^2}{2\sigma^2} \right] dy \\ &= A - \zeta_{\bar{p}}^{(1)} \Phi \left(\frac{\mu - \ln \zeta_{\bar{p}}^{(1)}}{\sigma} \right) \end{aligned} \quad (22)$$

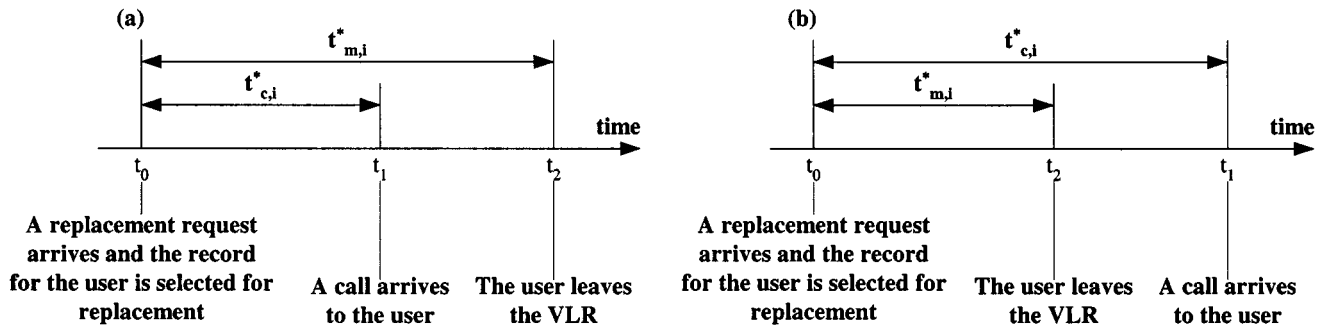


Fig. 3. Timing diagram II.

where

$$\begin{aligned}
 A &= \int_{\zeta_p^{(1)}}^{\infty} \left(\frac{1}{\sqrt{2\pi}\sigma} \right) \exp \left[-\frac{(\mu - \ln y)^2}{2\sigma^2} \right] dy \\
 &\quad (\text{let } z = \ln y) \\
 &= \int_{\ln \zeta_p^{(1)}}^{\infty} \left(\frac{1}{\sqrt{2\pi}\sigma} \right) \exp \left[-\frac{(\mu - z)^2}{2\sigma^2} \right] e^z dz \\
 &= \int_{\ln \zeta_p^{(1)}}^{\infty} \left(\frac{1}{\sqrt{2\pi}\sigma} \right) \exp \left[-\frac{(z - \mu - \sigma^2)^2}{2\sigma^2} \right] \\
 &\quad \times \exp \left[-\frac{\mu^2 - (\mu^2 + 2\mu\sigma^2 + \sigma^4)}{2\sigma^2} \right] dz \\
 &= \exp \left(\mu + \frac{1}{2}\sigma^2 \right) \Phi \left(\frac{\mu + \sigma^2 - \ln \zeta_p^{(1)}}{\sigma} \right). \quad (23)
 \end{aligned}$$

From (20), (22), and (23)

$$\begin{aligned}
 1 - \bar{p} &= e^{-\lambda_{c,1}\zeta_p^{(1)}} \left[\Phi \left(\frac{\mu + \sigma^2 - \ln \zeta_p^{(1)}}{\sigma} \right) \right. \\
 &\quad \left. - \left[\frac{\zeta_p^{(1)}}{\exp(\mu + \frac{1}{2}\sigma^2)} \right] \Phi \left(\frac{\mu - \ln \zeta_p^{(1)}}{\sigma} \right) \right]. \quad (24)
 \end{aligned}$$

From (11) and (21)–(23)

$$\begin{aligned}
 1 - \bar{q} &= e^{-\lambda_{c,2}\zeta_p^{(1)}} \left[\Phi \left(\frac{\mu + \sigma^2 - \ln \zeta_p^{(1)}}{\sigma} \right) \right. \\
 &\quad \left. - \left[\frac{\zeta_p^{(1)}}{\exp(\mu + \frac{1}{2}\sigma^2)} \right] \Phi \left(\frac{\mu - \ln \zeta_p^{(1)}}{\sigma} \right) \right]. \quad (25)
 \end{aligned}$$

From (10), we have

$$\begin{aligned}
 \lambda_{u,1}(1 - \bar{p}) + \lambda_{u,2}(1 - \bar{q}) &= \lambda_{u,1} + \lambda_{u,2} - \lambda_m N_1 \\
 &= \lambda_m N - \lambda_m N_1 \\
 &= \lambda_m N_2. \quad (26)
 \end{aligned}$$

Combining (24)–(26), we yield

$$\begin{aligned}
 \lambda_m N_2 &= \left(\lambda_{u,1} e^{-\lambda_{c,1}\zeta_p^{(1)}} + \lambda_{u,2} e^{-\lambda_{c,2}\zeta_p^{(1)}} \right) \\
 &\quad \times \left[\Phi \left(\frac{\mu + \sigma^2 - \ln \zeta_p^{(1)}}{\sigma} \right) - \left[\frac{\zeta_p^{(1)}}{\exp(\mu + \frac{1}{2}\sigma^2)} \right] \right. \\
 &\quad \left. \times \Phi \left(\frac{\mu - \ln \zeta_p^{(1)}}{\sigma} \right) \right]. \quad (27)
 \end{aligned}$$

Therefore, we can use a numerical method to solve $\zeta_p^{(1)}$ from (27), and then we can solve p by (24) and p_2 by (6). Finally, we can solve p_1 by (5).

After p_1 is obtained, we can compute P_{MR} as follows. Consider the timing diagrams in Fig. 3. Suppose that a replacement request arrives at time t_0 and the record for a mobile user u_i of class i is selected for replacement. Moreover, assume that the next call of u_i after t_0 arrives at time t_1 and u_i leaves the VLR at time t_2 . Let $t_{c,i}^* = t_1 - t_0$ and $t_{m,i}^* = t_2 - t_0$. From the excess life theorem, $t_{m,i}^*$ and $t_{c,i}^*$ have the density functions $r_{m,i}(t_{m,i}^*)$ and $r_{c,i}(t_{c,i}^*)$, respectively, where

$$\begin{aligned}
 r_{m,i}(t_{m,i}^*) &= \lambda_m [1 - F_m(t_{m,i}^*)] \\
 r_{c,i}(t_{c,i}^*) &= \lambda_{c,i} [1 - F_{c,i}(t_{c,i}^*)] = \lambda_{c,i} e^{-\lambda_{c,i} t_{c,i}^*}. \quad (28)
 \end{aligned}$$

Let $p_{3,i}$ be the probability that u_i will not make or receive calls before he/she leaves the VLR [i.e., (b) in Fig. 3]. Then

$$\begin{aligned}
 p_{3,i} &= \Pr[t_{m,i}^* < t_{c,i}^*] \\
 &= \int_{t_{m,i}^*=0}^{\infty} \int_{t_{c,i}^*=t_{m,i}^*}^{\infty} r_{c,i}(t_{c,i}^*) r_{m,i}(t_{m,i}^*) dt_{c,i}^* dt_{m,i}^* \\
 &= \left(\frac{\lambda_m}{\lambda_{c,i}} \right) [1 - f_m^*(\lambda_{c,i})] \quad (29)
 \end{aligned}$$

where $f_m^*(\cdot)$ is the Laplace Transform of the VLR residence time distribution.

Since p_1 is the proportion of class 1 users out of being replaced users, P_{MR} can be computed as follows:

$$P_{MR} = p_1 p_{3,1} + (1 - p_1) p_{3,2}. \quad (30)$$

The analytic model for the random replacement policy is similar to (but much simpler than) the most idle replacement model. The reader is referred to [9] for the details of the analytic model for random replacement.

The analytic model has been validated against the simulation experiments. Table I compares the analytic and simulation results, where $\lambda_{c,2} = 50\lambda_m$ and $\lambda_{u,1} = \lambda_{u,2} = 1000\lambda_m$. The table indicates that for exponential VLR residence time with $\lambda_m = 0.2$, the error rate is less than 0.02%. For lognormal VLR residence time with $\lambda_m = 0.2$ and $V_m = 1/\lambda_m^2$, the error rate is less than 1.2%. It is clear that the analytic analysis is consistent with the simulation results.

III. NUMERICAL RESULTS

This section investigates the performance of the overflow record replacement policies. In the numerical examples, the VLR residence time distributions for both classes are the

TABLE I
COMPARISON BETWEEN THE ANALYTIC AND SIMULATION RESULTS

Exponential $\lambda_m = 0.2$						
$\lambda_{c,1}/\lambda_m$	$10^{-3.0}$	$10^{-2.6}$	$10^{-2.2}$	$10^{-1.8}$	$10^{-1.4}$	$10^{-1.0}$
$P_{MR}(Analytic)$	99.9001%	99.7494%	99.373%	98.4398%	96.1713%	90.9091%
$P_{MR}(Simulation)$	99.8989%	99.761%	99.3769%	98.4381%	96.1722%	90.8984%
Lognormal $\lambda_m = 0.2, V_m = 1/\lambda_m^2$						
$\lambda_{c,1}/\lambda_m$	$10^{-3.0}$	$10^{-2.6}$	$10^{-2.2}$	$10^{-1.8}$	$10^{-1.4}$	$10^{-1.0}$
$P_{MR}(Analytic)$	99.9%	99.7496%	99.3744%	98.4478%	96.2153%	91.1256%
$P_{MR}(Simulation)$	99.9113%	99.7931%	99.4676%	98.6732%	96.7%	92.1765%

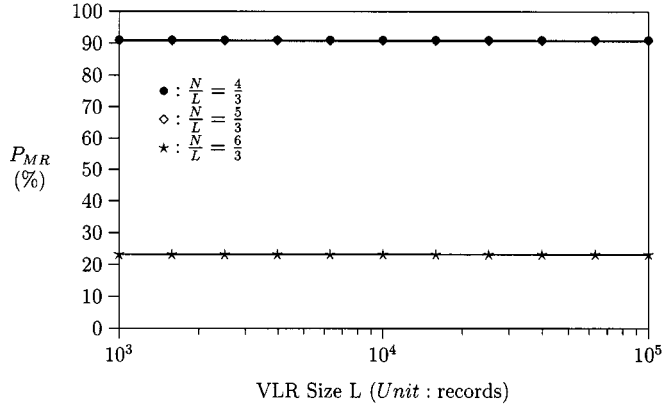


Fig. 4. Effects of N/L on P_{MR} ($\lambda_{c,1} = 0.1\lambda_m$, $\lambda_{c,2} = 50\lambda_m$, $\lambda_{u,1} = \lambda_{u,2} = 1000\lambda_m$).

same, which have a Gamma distribution with mean $1/\lambda_m$ and variance V_m . In most experiments, we assume that the total user arrival rate λ_u (i.e., $\lambda_{u,1} + \lambda_{u,2}$) to a VLR area is a fixed value $2000\lambda_m$. Fig. 4 plots P_{MR} as functions of N/L where $\lambda_{c,1} = 0.1\lambda_m$, $\lambda_{c,2} = 50\lambda_m$, and $\lambda_{u,1} = \lambda_{u,2} = 1000\lambda_m$. This figure shows that for a fixed N/L , P_{MR} values are the same for various sizes of the VLR database. Thus, in the remainder of this section, we only consider $L = 1500$. The same results can be obtained for larger database sizes with the same N/L ratio.

Effects of the Replacement Policies: Figs. 5–9 show that the most idle replacement policy significantly outperforms the random replacement policy in most cases. Define

$$\theta = \frac{\lambda_{u,1}}{\lambda_{u,1} + \lambda_{u,2}}.$$

Both policies have a similar performance when $\lambda_{u,1}$ is very small (e.g., $\theta = 0.1$ in Fig. 8) and the overflow situation is very serious. In such a circumstance, there is only a small number of class 1 users in the VLR area. Under serious overflow situation, it is likely that all class 1 users have already become overflow users, and class 2 users (in the VLR group) will be selected for replacement in an overflow registration or call setup. As a result, small P_{RR} and P_{MR} are expected, which are insignificantly affected by the replacement algorithms. The performance of the inactive replacement policy is between the most idle and the random replacement policies when V_m is not large. For large V_m , the inactive replacement policy may outperform the most idle replacement policy as will be elaborated in Fig. 5.

Effects of Variance V_m : Fig. 5 plots P_{RR} , P_{IR} (for the inactive replacement policy), and P_{MR} as functions of V_m where

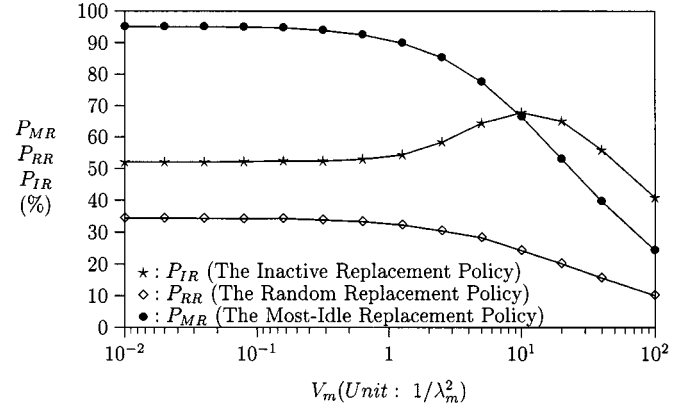


Fig. 5. Effects of variance V_m on P_{RR} , P_{IR} and P_{MR} ($\lambda_{c,1} = 0.1\lambda_m$, $\lambda_{c,2} = 50\lambda_m$, $\theta = 0.5$).

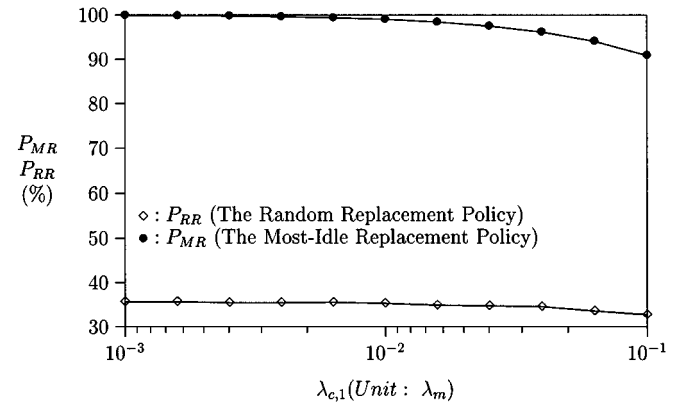
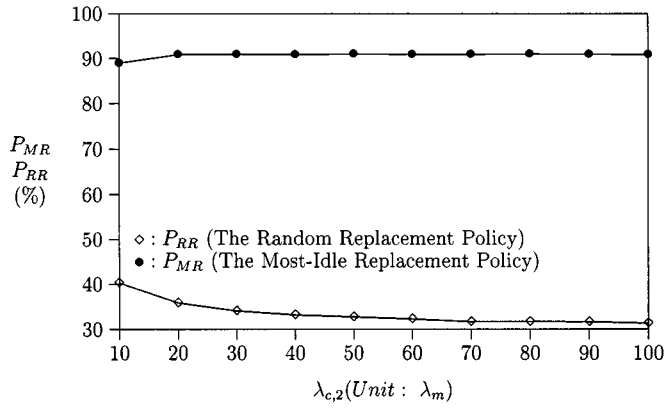
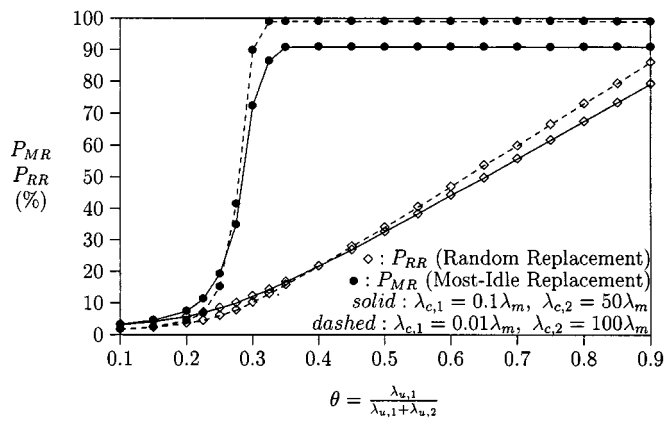
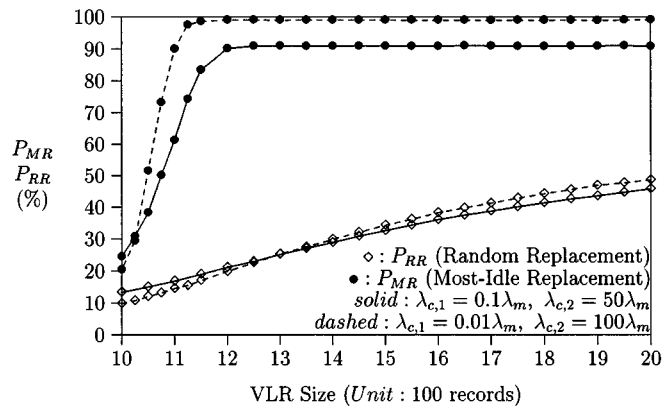


Fig. 6. Effects of $\lambda_{c,1}$ on P_{RR} and P_{MR} ($\lambda_{c,2} = 50\lambda_m$, $\theta = 0.5$).

$\lambda_{c,1} = 0.1\lambda_m$, $\lambda_{c,2} = 50\lambda_m$, and $\theta = 0.5$. We note that both P_{RR} and P_{MR} decrease as V_m increases. That is, for Gamma distributed VLR residence time, as V_m increases, more long $t_{m,i}^*$ are observed [see (28)], and $p_{3,i} = \Pr[t_{m,i}^* < t_{c,i}^*]$ decreases accordingly. Consequently, probabilities P_{RR} and P_{MR} decrease as V_m increases. On the other hand, P_{IR} not only depends on $p_{3,i}$ but also on the probability p_1 that a class 1 user is selected for replacement when a replacement request arrives. In [9] we showed that p_1 increases as V_m increases. The combined effects of $p_{3,i}$ and p_1 result in the phenomenon that P_{IR} increases and then decreases as V_m increases. This figure also indicates that $P_{MR} > P_{IR}$ for small V_m , and the reverse result is observed for large V_m . In the remainder of this section, we only compare P_{MR} and P_{RR} . A detailed analysis of the inactive replacement policy was given in [9] and will not be repeated


 Fig. 7. Effects of $\lambda_{c,2}$ on P_{RR} and P_{MR} ($\lambda_{c,1} = 0.1\lambda_m$, $\theta = 0.5$).

 Fig. 8. Effects of $\lambda_{u,i}$ on P_{RR} and P_{MR} .

 Fig. 9. Effects of the VLR size on P_{RR} and P_{MR} ($\theta = 0.5$).

in this paper. Also, in the remainder of this section, $V_m = 1/\lambda_m^2$ is considered.

Effects of $\lambda_{c,1}$: Fig. 6 plots P_{RR} and P_{MR} as functions of $\lambda_{c,1}$ where $\lambda_{c,2} = 50\lambda_m$ and $\theta = 0.5$. The figure indicates intuitive results that P_{RR} and P_{MR} decrease as $\lambda_{c,1}$ increases. The nonintuitive results are that both P_{RR} and P_{MR} are not affected by $\lambda_{c,1}$ when $\lambda_{c,1}$ is sufficiently small (namely, $\lambda_{c,1} < 0.01\lambda_m$). When $\lambda_{c,1} < 0.01\lambda_m$, the probability $p_{3,1}$ that a replaced class 1 user will not make or receive calls before he/she leaves the VLR will approach 1 [see (29)]. In this case, P_{RR} and P_{MR} will be only dominated by $\lambda_{c,2}$ [see (30)].

Effects of $\lambda_{c,2}$: Fig. 7 plots P_{RR} and P_{MR} as functions of $\lambda_{c,2}$ where $\lambda_{c,1} = 0.1\lambda_m$ and $\theta = 0.5$. The figure shows that P_{RR} decreases as $\lambda_{c,2}$ increases and P_{MR} increases as $\lambda_{c,2}$ increases. In random replacement policy, every record in the VLR is selected for replacement with the same probability. Thus, as $\lambda_{c,2}$ increases, it is more likely that the next call for the replaced class 2 user arrives before he/she leaves the VLR. Therefore P_{RR} decreases as $\lambda_{c,2}$ increases. In contrast, for the most idle replacement policy, as $\lambda_{c,2}$ increases, it is less likely that class 2 users are selected for replacement. Consequently, P_{MR} increases as $\lambda_{c,2}$ increases.

Effects of $\lambda_{u,1}$ ($\lambda_{u,2}$): Fig. 8 plots P_{RR} and P_{MR} as functions of the ratio $\theta = (\lambda_{u,1}/\lambda_{u,1} + \lambda_{u,2})$. When θ is large, it is more likely that class 1 users are selected for replacement. Thus, it is obvious that both P_{RR} and P_{MR} increase as the ratio θ increases. We also note that P_{MR} increases rapidly for $0.2 < \theta < 0.35$. For most idle replacement, when $\theta < 0.2$, there are only a small number of class 1 users in the VLR database which can be selected for replacement. Once $\theta > 0.2$, the probability p_1 that a class 1 user is selected for replacement will increase exponentially as θ increases. When θ reaches 0.35, there are already sufficient class 1 users for replacements. In this case, large P_{MR} is expected, and increasing θ does not improve P_{MR} performance.

Effects of the VLR Size: Fig. 9 plots P_{RR} and P_{MR} as functions of the VLR size where $\theta = 0.5$. Fig. 9 shows that P_{RR} increases linearly as the VLR size increases. On the other hand, P_{MR} increases rapidly as the VLR size increases from 1000 to 1200 records. After 1200 records, P_{MR} is only slightly affected by the change of the VLR size. That is, when the VLR size is less than a threshold (e.g., 1200 records in our example), the overflow situation will be very serious and class 2 users will also be selected for replacements. Therefore, increasing of the VLR size significantly reduces the occurrences of overflow replacements for class 2 users. When the VLR size is larger than the threshold, there are enough class 1 users for replacement, and increasing the VLR size does not improve P_{MR} .

IV. CONCLUSION

This paper studied the overflow problem of a mobility database, specifically the VLR, in a PCS network. Under the existing PCS technology, when the VLR is full, the incoming mobile users cannot receive services. Through record replacement, we can offer telecommunication services to overflow users (i.e., those users who have no associated records in the VLR). An important criterion of record replacement policies is to reduce the possibility of overflow operations in the future. To achieve this goal, we proposed the most idle replacement policy to replace the record that has no call activity for the longest period. We compared the most idle replacement policy with the previously proposed policies (inactive replacement and random replacement). Our study indicated that the most idle replacement policy significantly outperforms the random replacement policy. Both inactive and most idle policies show good performance. In most scenarios investigated in this paper, the most idle replacement policy can reduce over 90% of the overflow call setups.

As a final remark, we note that in a real mobile network, the records of a huge VLR are physically stored in several separated databases. The most idle policy will need to access all databases for a replacement while the inactive policy only needs to access some databases independently. In this case, the most idle policy is not as efficient as the inactive replacement policy. To overcome this disadvantage, we may exercise the most idle policy in the individual databases in parallel and then select the most idle record from the most idle records in the individual databases.

APPENDIX DERIVATIONS FOR \bar{N} , N_1 , K_1 , AND K_2

When the system is in equilibrium, let random variable N_1 be the number of users in the VLR group and random variable N_2 be the number of users in the overflow group. Then the number of users in the VLR area is $N = N_1 + N_2$. Let random variables K_1 and K_2 be the numbers of class 1 and class 2 users in the VLR area when the system is in equilibrium. It is clear that

$$N_1 \leq L. \quad (31)$$

Since a VLR area can be regarded as an $M/G/\infty$ queueing system, N is a Poisson random variable with mean $N^* = (\lambda_{u,1} + \lambda_{u,2})/\lambda_m$ [3], where $\lambda_{u,1} + \lambda_{u,2}$ is the user arrival rate from other VLR areas and $1/\lambda_m$ is the mean VLR residence time. Similarly, treating class 1 and class 2 users separately, K_1 and K_2 are Poisson random variables with means $\lambda_{u,1}/\lambda_m$ and $\lambda_{u,2}/\lambda_m$, respectively. When $\lambda_{u,1}/\lambda_m$ and $\lambda_{u,2}/\lambda_m$ are both of order N^* , from the Chebyshev's inequality [14], we have

$$N = \frac{\lambda_{u,1} + \lambda_{u,2}}{\lambda_m} + O_p(\sqrt{N^*})$$

$$K_1 = \frac{\lambda_{u,1}}{\lambda_m} + O_p(\sqrt{N^*})$$

and

$$K_2 = \frac{\lambda_{u,2}}{\lambda_m} + O_p(\sqrt{N^*})$$

where

$$X_n = O_p(f(n)) \stackrel{\text{def}}{=} \forall \epsilon > 0, \exists m > 0 \text{ such that} \\ \Pr\left(\left|\frac{X_n}{f(n)}\right| < m\right) > 1 - \epsilon \text{ for all } n \text{ [7].}$$

If N^* is sufficiently large

$$N \simeq \frac{\lambda_{u,1} + \lambda_{u,2}}{\lambda_m}, \quad K_1 \simeq \frac{\lambda_{u,1}}{\lambda_m} \quad \text{and} \quad K_2 \simeq \frac{\lambda_{u,2}}{\lambda_m}. \quad (32)$$

Consider the loss model $M/G/L/L$ where L is the size of the VLR database and is of order N^* . In this model, if the Poisson user arrivals from other VLR areas find that the VLR database is full, then they join the overflow group and never go back to the VLR group. Let random variable \bar{N} be the number of users in this loss model in equilibrium. From [15, Th. 5.7.4], we know that \bar{N} has a truncated Poisson distribution with parameter N^* which is truncated at L , and therefore

$$\bar{N} = L + O_p(\sqrt{N^*}). \quad (33)$$

In (33), $O_p(\sqrt{N^*})$ has a negative value. Compare the loss model with the original system, and we have

$$\bar{N} \leq N_1 \quad (34)$$

because the overflow users in the original system may go back to the VLR group, while those in the loss model do not. From (31), (33), and (34)

$$\bar{N} \leq N_1 \leq L \implies N_1 = L + O_p(\sqrt{N^*}).$$

When N^* is sufficiently large, we have

$$N_1 \simeq L. \quad (35)$$

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments.

REFERENCES

- [1] 3GPP, "3rd Generation Partnership Project; Technical Specification Group Services and Systems Aspects; Architectural Requirements for Release 1999," Technical Rep. 3G TS 23.121 version 3.3.0 (2000-03), 3GPP, 1999.
- [2] ETSI/TC, "Mobile application part (MAP) Specification, Version 4.8.0," Technical Rep. Recommendation GSM 09.02, ETSI, 1994.
- [3] D. Gross and C. M. Harris, *Fundamentals of Queueing Theory*, 3rd ed. New York: Wiley, 1998.
- [4] C. Jedrzycki and V. C. M. Leung, "Probability distribution of channel holding time in cellular telephony systems," in *IEEE Vehic. Technol. Conf. VTC 1996*, vol. 1, pp. 247-251.
- [5] N. L. Johnson, *Continuous Univariate Distributions—2*. New York: Wiley, 1970.
- [6] J. Jordan and F. Barcelo, "Statistical modeling of transmission holding time in PAMR systems," in *Proc. IEEE Globecom '97*, vol. 1, 1997, pp. 121-125.
- [7] E. L. Lehmann, *Elements of Large-Sample Theory*. New York: Springer-Verlag, 1999.
- [8] Y.-B. Lin, "Overflow control for cellular mobility database," *IEEE Trans. Vehic. Technol.*, vol. 49, pp. 520-530, 2000.
- [9] —, "Eliminating overflow for large-scale mobility databases in cellular telephone networks," *IEEE Trans. Computers*, vol. 50, Apr. 2001.
- [10] Y.-B. Lin and I. Chlamtac, *Wireless and Mobile Network Architectures*. New York: Wiley, 2000.
- [11] M. J. Maron and R. J. Lopez, *Numerical Analysis a Practical Approach*, 3rd ed. Wadsworth, 1991.
- [12] Mehrotra, *GSM System Engineering*: Artech, 1997.
- [13] A. M. Mood, F. A. Graybill, and D. C. Boes, *Introduction to the Theory of Statistics*, 3rd ed. New York: McGraw-Hill, 1974.
- [14] S. M. Ross, *Introduction to Probability Models*, 5th ed. New York: Academic, 1993.
- [15] —, *Stochastic Processes*, 2nd ed. New York: Wiley, 1996.



Hui-Nien Hung received the B.S.Math. degree from National Taiwan University, Taiwan, R.O.C., in 1989, the M.S.Math. degree from National Tsing-Hua University, Taiwan, R.O.C., in 1991, and the Ph.D. degree in statistics from The University of Chicago in 1996.

Since 1996, he has been an Associate Professor at the Institute of Statistics, National Chiao Tung University, Taiwan, R.O.C. His current research interests include applied probability, financial calculus, bioinformatics, and mathematical statistics.



Yi-Bing Lin (SM'96) received the B.S.E.E. degree from National Cheng Kung University in 1983 and the 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 a Professor of the Department of Computer Science and Information Engineering (CSIE), National Chiao Tung University (NCTU). In 1996, he was appointed Deputy Director of the Microelectronics and Information Systems Research Center, NCTU. During the period 1997–1999, he was elected Chairman of CSIE, NCTU. He authored the book *Wireless and Mobile Network Architecture* (co-author with Imrich Chlamtac; New York: Wiley). His current research interests include design and analysis of personal communications services network, mobile computing, distributed simulation, and performance modeling.

Dr. Lin is an Associate Editor of *IEEE Network*, an editor of *IEEE JOURNAL OF SELECTED AREAS OF COMMUNICATION: Wireless Series*, an editor of the *IEEE Personal Communications Magazine*, an editor of *Computer Networks*, an area editor of *ACM Mobile Computing and Communication Review*, a columnist of *ACM Simulation Digest*, an editor of *International Journal of Communications Systems*, an editor of *ACM/Baltzer Wireless Networks*, an editor of *Computer Simulation Modeling and Analysis*, and an editor of *Journal of Information Science and Engineering*. He was also 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, Guest Editor for the ACM/Baltzer MONET special issue on Personal Communications, a Guest Editor for *IEEE TRANSACTIONS ON COMPUTERS: Special Issue on Mobile Computing*, and a Guest Editor for *IEEE Communications Magazine* special issue on Active, Programmable, and Mobile Code Networking. He received Outstanding Research Awards from the National Science Council, R.O.C., in 1998 and 2000 and the 1998 Outstanding Youth Electrical Engineer Award from CIEE, R.O.C. He is an Adjunct Research Fellow of Academia Sinica.

Dr. Lin is an Associate Editor of *IEEE Network*, an editor of *IEEE JOURNAL OF SELECTED AREAS OF COMMUNICATION: Wireless Series*, an editor of the *IEEE Personal Communications Magazine*, an editor of *Computer Networks*, an area editor of *ACM Mobile Computing and Communication Review*, a columnist of *ACM Simulation Digest*, an editor of *International Journal of Communications Systems*, an editor of *ACM/Baltzer Wireless Networks*, an editor of *Computer Simulation Modeling and Analysis*, and an editor of *Journal of Information Science and Engineering*. He was also 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, Guest Editor for the ACM/Baltzer MONET special issue on Personal Communications, a Guest Editor for *IEEE TRANSACTIONS ON COMPUTERS: Special Issue on Mobile Computing*, and a Guest Editor for *IEEE Communications Magazine* special issue on Active, Programmable, and Mobile Code Networking. He received Outstanding Research Awards from the National Science Council, R.O.C., in 1998 and 2000 and the 1998 Outstanding Youth Electrical Engineer Award from CIEE, R.O.C. He is an Adjunct Research Fellow of Academia Sinica.



Nan-Fu Peng received the B.S. degree in applied mathematics from National Chiao Tung University, Hsinchu, Taiwan, in 1981, and the Ph.D. degree in statistics from the Ohio State University in 1989.

He is now an Associate Professor of the Institute of Statistics, National Chiao Tung University. His research interests include Markov chains, population dynamics, and queueing theory.



Shun-Ren Yang received the B.S.C.S.I.E. and M.S.C.S.I.E. degrees from National Chiao Tung University (NCTU), Hsinchu, Taiwan, R.O.C., in 1998 and 1999, respectively. He is currently working toward the Ph.D. degree at NCTU.

His research interests include design and analysis of a personal communications services network, computer telephony integration, mobile computing, and performance modeling.