Overflow Control for Cellular Mobility Database

Yi-Bing Lin, *Senior Member, IEEE*

***Abstract*—In a cellular phone system, the service area is partitioned into several location areas (LA's). Every LA is associated with a mobility database called visitor location register (VLR). When a mobile user enters an LA, the user must register to the VLR before receiving any cellular service. If the VLR is full, the registration procedure fails and the system cannot deliver services to the user under the existing cellular technology. To resolve this problem, we propose a VLR overflow control scheme to accommodate the incoming mobile users during VLR overflow. Our scheme only requires minor modifications to the existing cellular mobility management protocols. Particularly, no modification is made to the mobile phones.**

An analytic model is proposed to investigate the performance of the overflow control scheme. When exercising the scheme, the call setup procedure for an "overflow" user is more expensive than that for a "normal" user. Under the range of input parameters considered in our study, we show that even if the VLR overflow situation is serious, the overhead for exercising the overflow control scheme is very low.

***Index Terms*—Mobility database, overflow control, personal communication services, visitor location register.**

I. Introduction

CELLULAR or *mobile communications services* facilitate the exchange of information (voice, data, video, image, etc.) for mobile users independent of time, location, and access arrangement [1]–[3]. One of the most important issues in mobile communications is *mobility management*. To understand the mobility management issue, we first introduce the cellular system architecture (see Fig. 1). In this architecture, the cellular service area is covered by a set of *base stations*. The base stations are responsible for serving the calls to or from the mobile phones (in this figure, the mobile phones are mounted on the vehicles) located in their coverage areas. The base stations are connected to the *mobile switching centers* (MSC's) by land links. The MSC is a telephone exchange specially assembled for mobile communications, which interfaces the mobile phones (via base stations) and the public switched telephone network (PSTN). The base stations are grouped into location areas (LA's). The base stations in an LA are connected to the same MSC. When a mobile user moves from one LA (e.g., New York City) to another (e.g., Los Angeles), the cellular system should be informed of the current location of the user. Otherwise, it is impossible to deliver the services to this user. To support mobility management, protocols such as the EIA/TIA Interim Standard 41 (IS-41) [4] or global system for mobile communications (GSM) mobile application part (MAP) [5] have been defined for cellular systems. The IS-41 protocol is used in advanced mobile phone service (AMPS), IS-136 digital AMPS (DAMPS), and IS-95 code-division multiple-access (CDMA) cellular systems. The GSM MAP protocol is used in GSM, GSM 1800, and GSM 1900 systems. These two protocols follow a *two-level* database strategy in that they use a two-tier system of home and visited databases. When a user subscribes to the services of a cellular system, a record is created in the system's database called home location register (HLR). The 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, and validation period). The HLR is part of the cellular network and not under the control of the local exchange carriers. The HLR is connected to the PSTN via SS7 links. Every LA is associated with a database called visitor location register (VLR). The VLR is the location register other than the HLR used to retrieve information for handling of calls to or from a mobile user visiting the LA. When the mobile user visits an LA, a temporary record for the mobile user is created in the corresponding VLR. When the user leaves the LA, the corresponding VLR record is deleted. A VLR may control several MSC areas (and thus the LA's inside these MSC's). To simplify our discussion, we assume that every MSC covers an LA, and every VLR controls exactly one MSC (and thus one LA). Our results can be easily generalized to the cases where a VLR controls several MSC's, and an MSC controls several LA's. The details of the mobility management algorithms based on VLR and HLR will be elaborated in Section II. The algorithms for multiple LA's per VLR are given in [5].

In a cellular system, the number of the records in the HLR is the number of the customers in the system. When a customer first subscribes to the service, a permanent HLR record is created for the customer. Since mobile users of different cellular systems may visit an LA (e.g., the GSM users from England and Taiwan may visit Hong Kong), the number of the records in the VLR changes dynamically. Specifically, new records are created when users move in, and obsolete records are deleted when the corresponding users move out. Furthermore, the number of the records in the corresponding VLR may be larger than that of the HLR. The VLR may overflow if too many mobile users move into the LA in a short period. If the VLR is full when a mobile user arrives, the user fails to "register" in the database and thus cannot receive cellular service. This phenomenon is called *VLR overflow*. To resolve this problem, Section III proposes a VLR overflow control scheme that allows users to receive services when a VLR is full. The performance of the overflow control scheme is investigated in Section IV.

Manuscript received June 26, 1997; revised February 23, 1999. This work was supported in part by the National Science Council under Contract NSC88-2213-E009-079.

The author 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).

Publisher Item Identifier S 0018-9545(00)02542-1.

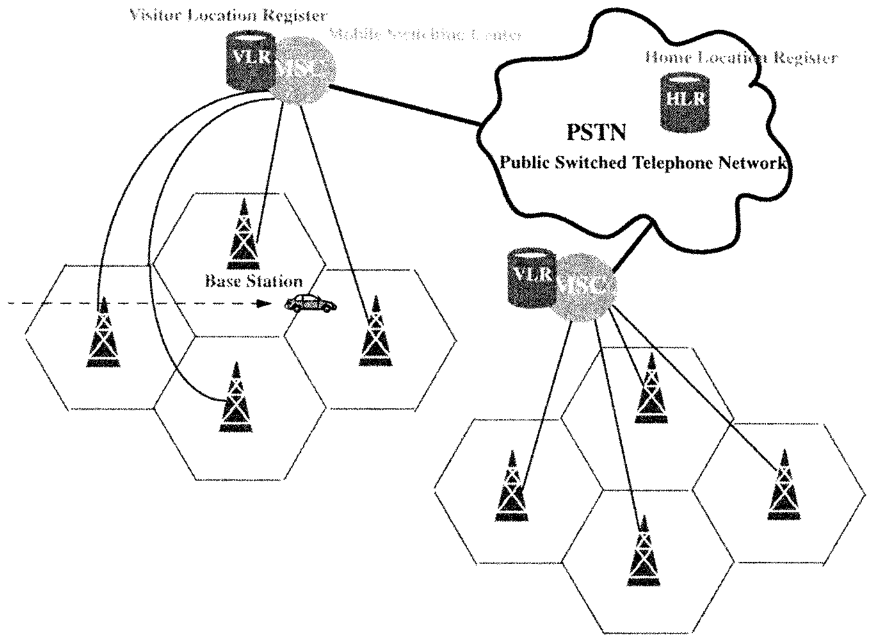


Fig. 1. An illustrative cellular system architecture.

II. Mobility Management in Existing Systems

This section describes mobility management for the existing cellular systems based on IS-41 [4] or GSM [5]. Four algorithms are described. Algorithm I (registration) and Algorithm II (cancellation) are used for location update. Algorithm III (call origination) and Algorithm IV (call termination) are used for call control. With the location update procedure, the HLR maintains the current LA address of every mobile user and guarantees that the mobile user has exactly one VLR record. The call origination procedure utilizes the VLR record to set up a call initiated by a mobile user. The call termination procedure utilizes both theHLR and VLR records to set up a call terminated at a mobile user.

Suppose that user $u1$ moves from location area LA1 to location area LA2. The registration procedure is performed between the HLR and $V2$ (the VLR for LA2), and the cancellation (deregistration) procedure is performed between the HLR and $V1$ (the VLR for LA1).

Algorithm I. Registration (see Fig. 2):

Step 1) (Registration Request)

- Step 1.1) The mobile phone of $u1$ sends a registration message to $V2$.
- Step 1.2) $V2$ creates a temporary VLR record for $u1$.
- Step 1.3) $V2$ forwards the registration request to the HLR. This message reg_msg is MAP_UPDATE_LOCATION in GSM and is REGNOT in IS-41.

Step 2) (Registration Response)

- Step 2.1) The HLR updates the location of $u1$. This location information will be used to set up calls terminated at $u1$.
- Step 2.2) The HLR acknowledges the registration operation and sends $u1$'s profile to $V2$. This message reg_ack is MAP_UPDATE_LOCATION_ack in GSM and is regnot in IS-41. The profile provides necessary information for $u1$ to originate phone calls.
- Step 2.3) $V2$ sends an acknowledgement to the mobile phone.

Note 1: In GSM, the mobile phone sends the temporary mobile subscriber identity (TMSI) to $V2$ in Step 1) [5]. $V2$ then uses the TMSI to find the mobile phone's international mobile subscriber identity (IMSI). The IMSI is used by $V2$ to locate $u1$'s HLR.

Note 2: In Step 1), authentication (to check if $u1$ is a legal user) may be performed in separate message exchanges.

After $u1$ has moved from $V1$ to $V2$, $u1$'s VLR record in $V1$ is obsolete and is deleted as described in the following algorithm.

Algorithm II. Cancellation (see Fig. 3):

Step 1) The HLR sends a cancellation message to $u1$'s old VLR $V1$. This message cancel_msg is MAP_CANCEL_LOCATION in GSM and is REGCANC in IS-41.

Step 2) $V1$ deletes $u1$'s record.

Step 3) $V1$ sends a cancellation acknowledgement to the HLR. This message cancel_msg is MAP_CANCEL_LOCATION_ack in GSM and is regcanc in IS-41.

When $u1$ originates a call, the following algorithm is executed. Note that the authentication procedure is omitted to simplify our description.

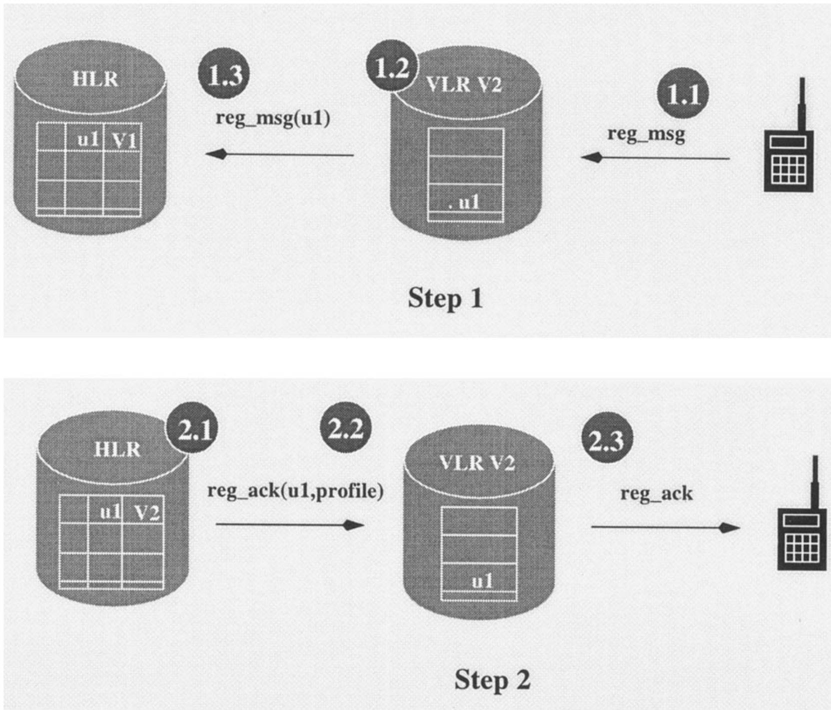


Fig. 2. The registration operation.

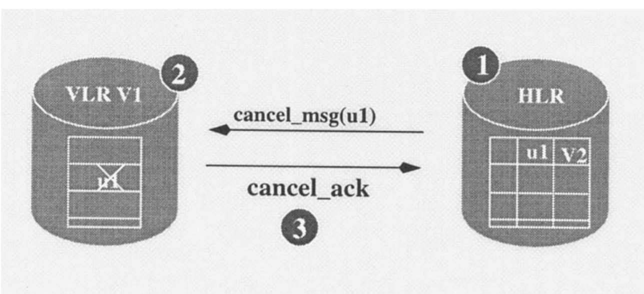


Fig. 3. The cancellation operation.

Algorithm III. Call Origination (see Fig. 4):

Step 1) The mobile phone sends the call origination request to the MSC.

Step 2) The MSC forwards the request to $V2$. The message call_req is MAP_SEND_INFO_FOR_OUTGOING_CALL in GSM and is ORREQ in IS-41.

Step 3) $V2$ checks the $u1$'s profile and grants the call request. The message call_ack is MAP_SEND_INFO_FOR_OUTGOING_CALL_ack in GSM and is orreq in IS-41.

Step 4) The MSC sets up the trunk according to the standard PSTN call setup procedure.

To deliver a call to $u1$, the following algorithm is executed. Without loss of generality, we assume that the calling party is a wireline user.

Algorithm IV. Call Termination (see Fig. 5):

Step 1) The calling party dials the phone number of $u1$. This phone number is mobile station ISDN number (MSISDN) in GSM or mobile identification number (MIN) in IS-41. The request is sent to the originating switch in PSTN. In GSM, this originating switch is a *Gateway MSC*.

Step 2) The originating switch sends a location query message to the HLR. This message loc_req is MAP_SEND_ROUTING_INFORMATION in GSM and is LOCREQ in IS-41.

Step 3) The HLR identifies the address of $u1$'s VLR (i.e., $V2$) and sends a query message to obtain the routing information. This message rout_req is MAP_PROVIDE_ROAMING_NUMBER in GSM and is ROUTREQ in IS-41.

Step 4) $V2$ creates the routable address of $u1$ and sends it back to the HLR. This routable address rout_addr is the mobile station roaming number (MSRN) in GSM and the temporary local directory number (TLDN) in IS-41. The message rout_ack is MAP_PROVIDE_ROAMING_NUMBER_ack in GSM and is routreq in IS-41.

Step 5) The HLR returns the routable address to the originating switch. This message loc_ack is MAP_SEND_ROUTING_INFORMATION_ack in GSM and is loc_req in IS-41.

Step 6) The originating switch sets up the trunk to the MSC based on the routable address.

Step 7) The MSC pages the mobile phone and the call path is established.

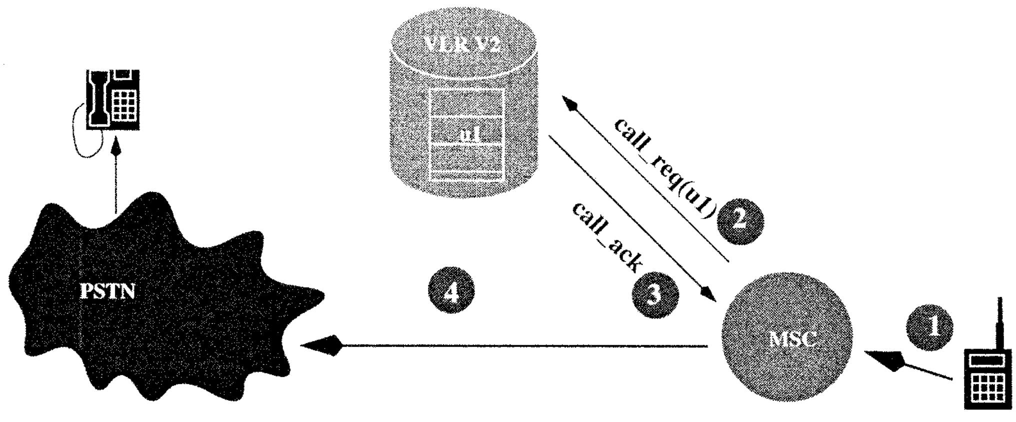


Fig. 4. The call origination operation.

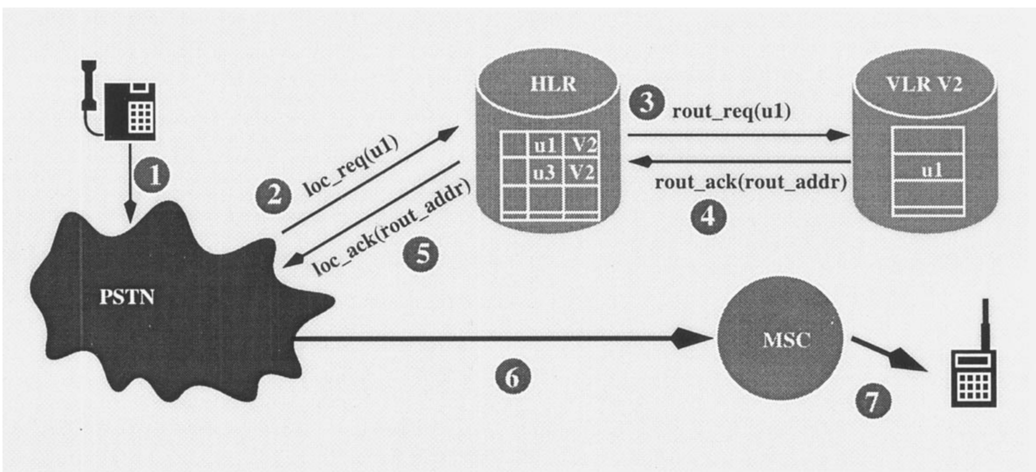


Fig. 5. The call termination operation.

III. Mobility Management in Overflow Systems

When a VLR is full, the incoming mobile users cannot register using Algorithm I and thus cannot receive cellular services. To resolve this problem, we propose the VLR overflow control algorithms O-I, O-II, O-III, and O-IV. Our approach allows new users to receive services when the VLR is full. In our overflow control scheme, an extra flag (one bit) is required in the HLR records. No modifications are made to the mobile phone.

Algorithm O-I. Registration (see Fig. 6): If $V2$ is not full, then Algorithm I is executed. If $V2$ is full, then the following steps are executed.

Step 1) (Registration Request)

- Step 1.1) This step is the same as that in Algorithm I.
- Step 1.2) The database is full. $V2$ follows a replacement policy to select a record to be deleted ($u3$ in Fig. 6). The storage for the deleted record is used to store the $u1$'s information. The selected user (i.e., $u3$) is called the *overflow user*. The replacement policy may be based on various heuristics. For example, $V2$ may select a record randomly, select the oldest record, or select an inactive record (i.e., the user has not had call activities recently). $V2$ may select $u1$ as the overflow user (i.e., $u3 = u1$) and do not create the VLR record for $u1$.
- Step 1.3) $V2$ forwards the registration request to the HLR with the indication that $u3$'s record is deleted due to database overflow.

Step 2) (Registration Response)

- Step 2.1) The HLR updates the location of $u1$ and sets the overflow flag in $u3$'s record (to indicate that $V2$ does not have a VLR record for $u3$). Note that $u3$ may be identical to $u1$ as pointed out in Step 1.2).
- Step 2.2) The HLR acknowledges the registration operation and sends $u1$'s profile to $V2$ (if $u1$ is the overflow user, then the message does not include the profile information).
- Step 2.3) $V2$ sends an acknowledgement to the mobile phone.

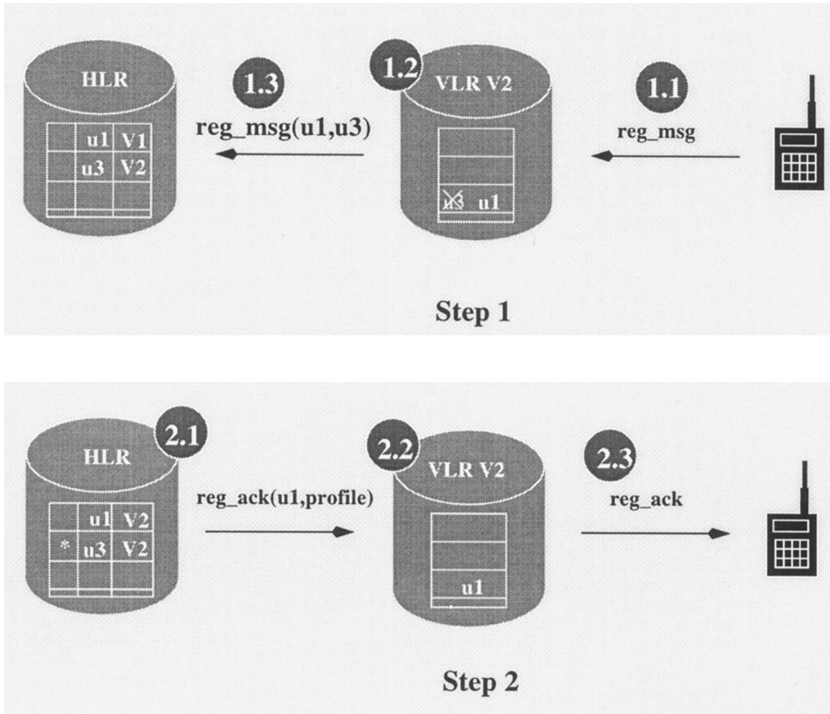


Fig. 6. The registration operation (overflow).

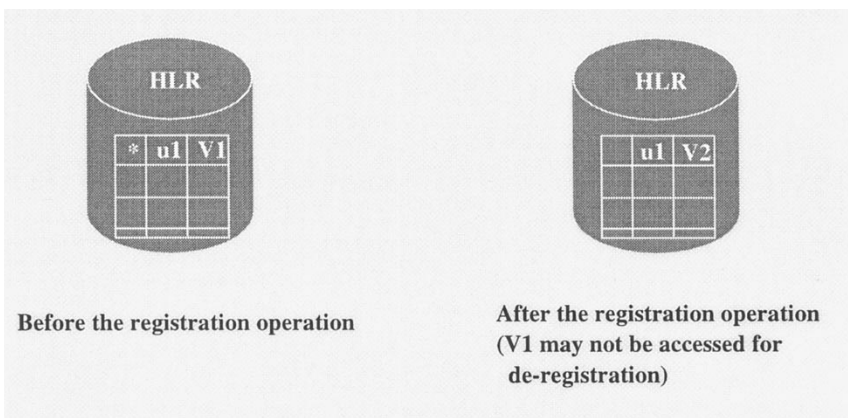


Fig. 7. The cancellation operation (overflow).

Algorithm O-II. Cancellation: If $u1$ is not an overflow user at $V1$, then Algorithm II is executed to cancel $u1$'s VLR record in $V1$. If $u1$ is an overflow user at $V1$, then $u1$ does not have a record in $V1$. The cancellation operation simply resets the overflow flag of $u1$'s HLR record if $u1$ is not an overflow user in $V2$ (see Fig. 7).

The call origination for an overflow user is described below.

Algorithm O-III. Call Origination (see Fig. 8):

Step 1) The mobile phone sends the call origination request to $V2$ as described in Steps 1) and 2) in Algorithm III.

Step 2) $V2$ cannot find $u1$'s record and denies the call request.

Steps 3-4) The mobile phone initiates the registration procedure and Algorithm O-I is executed.

Steps 5-6) The mobile phone reissues the call origination request and Algorithm III is executed.

To deliver a call to an overflow user, the following algorithm is exercised.

Algorithm O-IV. Call Termination (see Fig. 9):

Step 1) (Location Query)

Step 1.1) The calling party dials the phone number of $u1$. The request is sent to the originating switch in PSTN.

Step 1.2) The originating switch sends a location query message to the HLR.

Step 1.3) The HLR identifies that $u1$ is an overflow user and sends a query message to obtain the

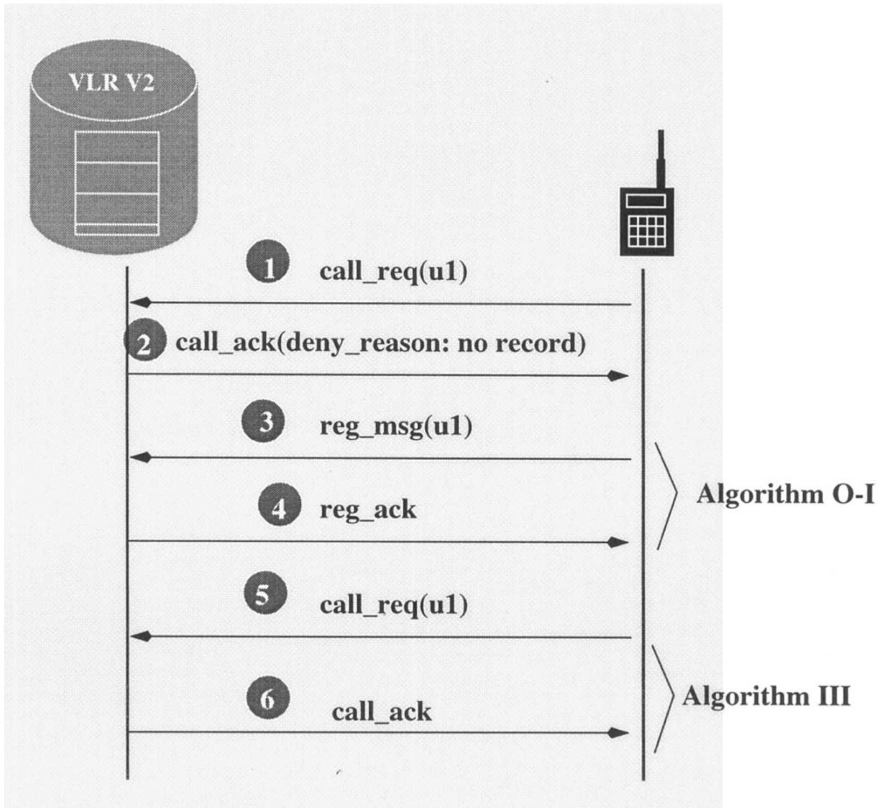


Fig. 8. The call origination operation (overflow).

routing information. The user profile information is attached in the message.

Step 2) (Location Response)

Step 2.1) If $V2$ is not full, then a record for $u1$ is created. If $V2$ is full, then a user record is deleted and is used to store $u1$'s information. $V2$ creates the routable address of $u1$ and sends it back to the HLR. If the VLR record is not available, the details of the routable address creation are described in [5]. If a record is replaced ($u3$ in Fig. 9), the replacement information is included in the message.

Step 2.2) The HLR returns the routable address to the originating switch. If a record is replaced, the overflow flags (for $u1$ and $u3$ in Fig. 9) are updated at the HLR.

Step 2.3) The originating switch sets up the trunk to the MSC based on the routable address.

Step 2.4) The MSC pages the mobile phone and the call path is established.

With Algorithms O-I–O-IV, an LA can accommodate unlimited number of mobile users as long as the number of simultaneous phone calls to these users is no larger than the size of the database (this situation never occurs in the real world).

IV. Modeling the Overflow Control Scheme

When exercising the overflow control scheme, we are interested in several performance issues.

Issue I. What is the stationary probability $p_{\rm ov}$ that the VLR overflow situation occurs?

Issue II. What is the probability α that the VLR record of a mobile user is never replaced due to VLR overflow?

Issue III. What is the probability β that the call activities of a user is not affected by VLR overflow? Note that an "affected" call connection (i.e., Algorithms O-III and O-IV) is more complicated than a "unaffected" call connection (i.e., Algorithms III and IV).

Issue III is different from Issue II. The call procedure is affected by VLR overflow if Algorithm O-III (for call origination) and Algorithm O-IV (for call termination) are exercised during call setup. After a VLR record replacement, the user may not originate or receive any call before moving out of the LA. Thus $\beta > \alpha$.

To address the above issues, we make the following assumptions.

Assumption I. Let N be the expected number of users in an LA. The expected value N can be obtained from operation, administration, and maintenance (OA&M) measurements in the cellular systems.

Assumption II. Let the residence time of a user in the LA have a general density function $f(t)$ with the Laplace transform $f^*(s) = \int_{t=0}^{\infty} f(t)e^{-st}\,dt$ and the mean $(1/\eta)$. The measurements for the arrival and the departure times of a user at an LA are available in the cellular systems. These data can be approximated by a general

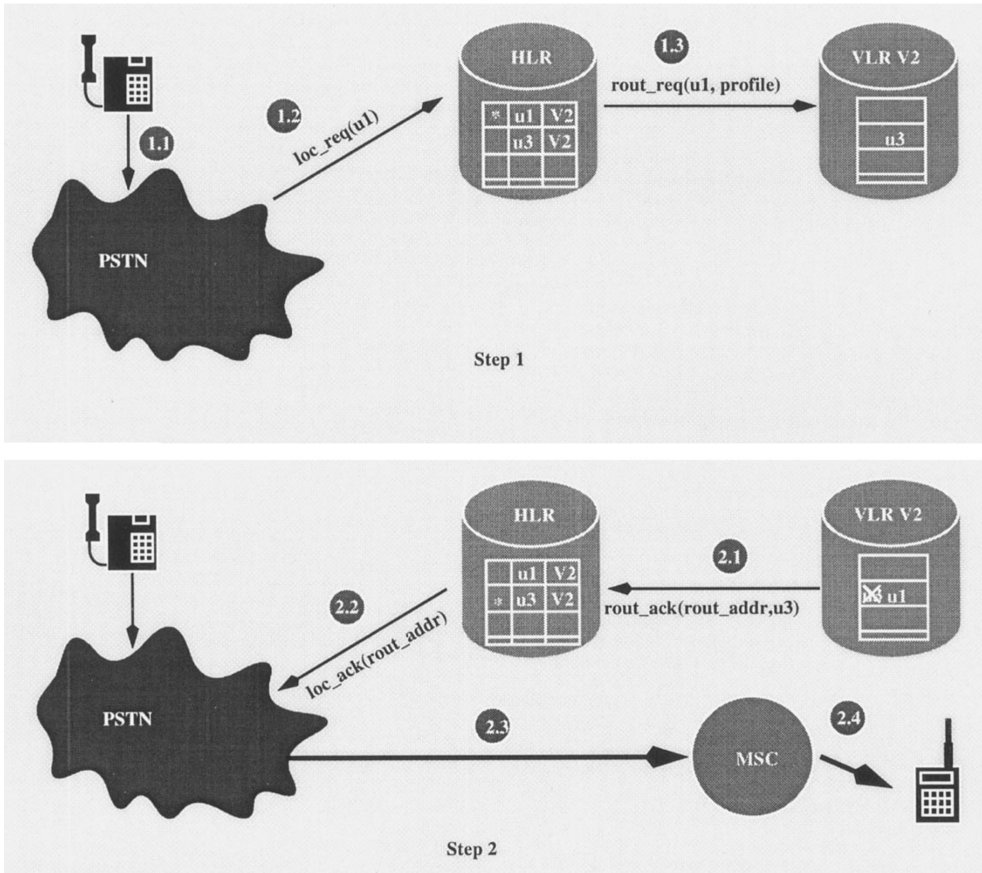


Fig. 9. The call termination operation (overflow).

density function $f(t)$ using the standard statistic methods.

Assumption III. Let the call arrivals to a user be a Poisson distribution with the arrival rate λ. The Poisson call arrivals are a widely used assumption in telecommunication modeling, which has been validated in PSTN.

Assumption IV. Let M be the size of the VLR.

Assumption V. Assume that during the short periods (i.e., the residence time of a user) the population is stationary. We will addresses the mobility effect through the population distribution (1). This technique significantly reduces the complexity of the analytic analysis, which has been shown reasonable in performance [6].

Assumption VI. If a registered mobile user u_1 (a nonoverflow user) leaves the VLR, there are two approaches to handle the cancellation procedure. In the first approach, the HLR selects an overflow user u_2 (if any) and sends its information to the VLR through the cancellation message (u_1's deregistration). The VLR then creates the VLR record for u_2 using the storage for u_1. Thus, if the number of users in the VLR is larger than M, we guarantee that the VLR is always full. The advantage is that when call origination/termination occurs to u_2, Algorithm III/IV is used without overflow handling cost.

In the second approach, the VLR simply deletes the record for u_1 without further action. When the next new user u_3 arrives, the VLR creates a VLR record for u_3 using the reclaimed storage. The advantage of this approach is that the VLR may not be considered overflow even if the number of user in the area is larger than M. In the following analysis, the first approach is considered.

A. Issue I: Derivation for p_{ov}

A VLR area overflows if the number of users in the area is larger than the size of the VLR database. The probability p_{ov} was computed in our previous work [6]–[8]. Specifically, we derived the LA population distribution or the probability π_n that there are n users in the LA. The probability π_n is derived as follows. In the steady state, the rate that mobile phones move into a cell equals to the rate that they move out of the cell. The arrivals of mobile phones to an LA can be viewed as being generated from N input streams that have the same general distribution with arrival rate η. The net input stream to an LA can be approximated as a Poisson process with arrival rate $\lambda^* = N\eta$. Thus, the distribution for the mobile phone population can be modeled by an $M/G/\infty$ queue with arrival rate λ^* and the service rate η. From [9], the steady-state probability π_n is derived as

$$\pi_n = \left(\frac{\lambda^*}{\eta}\right)^n \frac{e^{-\frac{\lambda^*}{\eta}}}{n!} = \frac{N^n e^{-N}}{n!}. \tag{1}$$

Since VLR overflow occurs when there are more than M users in the LA, we have

$$p_{\text{ov}} = \sum_{n=M+1}^{\infty} \pi_n. \tag{2}$$

Based on (2), Fig. 10 plots p_{ov} as a function of M/N for different N values. The figure indicates that p_{ov} is large if $M < N$ and $p_{\text{ov}} < 0.2\%$ if $M > 1.3N$. The figure also shows that for the same M/N value, larger N results in smaller p_{ov}.

B. Issue II: Derivation for α

The probability α (i.e., the probability that the VLR record of a mobile user is never replaced due to VLR overflow) is derived as follows. We make an approximation assumption that during the short periods (i.e., the residence time of a user) the population is stationary and then address the mobility effect through the population distribution (1). This technique significantly reduces the complexity of analytic analysis and has been shown reasonably accurate [6]. Suppose that there are $n > M$ users in the LA. Suppose that the VLR records are randomly selected for replacement at Steps 1) and 2) in Algorithm O-I. Then the probability q that a VLR record is selected for replacement is

$$q = \frac{1}{M}.$$

From Assumption III, the rate of the calls that result in VLR record replacement is

$$\mu = (n - M)\lambda. \tag{3}$$

Let $\Pr[K = k, t, \mu]$ be the probability that there are k call arrivals with VLR record replacement during a period t. From Assumption III, $\Pr[K = k, t, \mu]$ has a Poisson distribution

$$\Pr[K = k, t, \mu] = \frac{(\mu t)^k}{k!} e^{-\mu t}.$$

Let $\alpha(t, \mu)$ be the probability that a VLR record is not selected for replacement during the period t. Then

$$\begin{aligned}\alpha(t,\mu) &= \sum_{k=0}^{\infty} (1-q)^k \Pr[K = k, t, \mu] \\ &= \sum_{k=0}^{\infty} \frac{[\mu(1-q)t]^k}{k!} e^{-\mu t} \\ &= e^{-q\mu t}.\end{aligned}$$

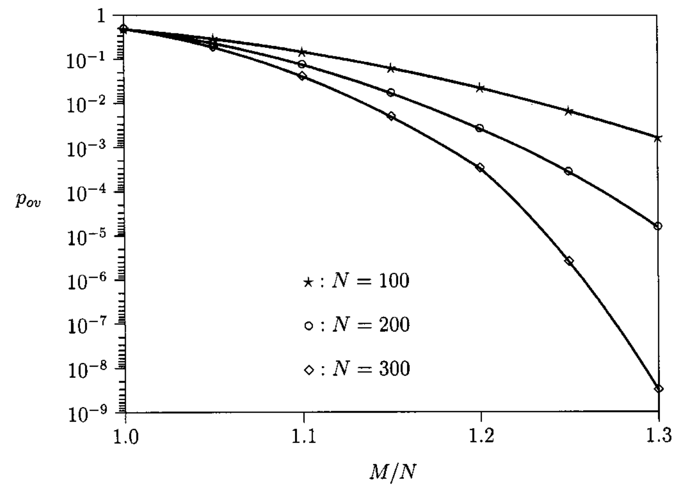


Fig. 10. The stationary probability p_{ov} for VLR overflow.

Let $\alpha(\mu)$ be the probability that the VLR record for a user is not replaced (under the condition that (3) holds). Since the LA residence time of a user has a general density function $f(t)$ (Assumption II), $\alpha(\mu)$ can be derived as follows:

$$\begin{aligned}\alpha(\mu) &= \int_{t=0}^{\infty} f(t)\alpha(t,\mu)\, dt \\ &= \int_{t=0}^{\infty} f(t) e^{-q\mu t}\, dt \\ &= f^*(q\mu).\end{aligned} \tag{4}$$

From (3), $\alpha(\mu) = \alpha((n-M)\lambda) = f^*(q(n-M)\lambda)$. From (1) and an argument similar to the one for deriving (2), we have

$$\begin{aligned}\alpha &= 1 - \sum_{n=M+1}^{\infty} \pi_n [1 - \alpha(q(n-M)\lambda)] \\ &= 1 - \sum_{n=M+1}^{\infty} \pi_n [1 - f^*(q(n-M)\lambda)].\end{aligned} \tag{5}$$

If f is an exponential density function, then

$$f^*(s) = \frac{\eta}{s + \eta} \tag{6}$$

and (5) is rewritten as

$$\alpha = 1 - \sum_{n=M+1}^{\infty} \frac{\pi_n (n-M)\lambda}{(n-M)\lambda + \eta}. \tag{7}$$

Based on (7), the solid curves in Fig. 11 illustrates α as a function of M/N. Fig. 11(a) plots the α curves when $\eta = 0.1\lambda, \lambda$ and 10λ (i.e., 10, 1, and 0.1 calls for a mobile user are expected during the user's stay in an LA), respectively.

In this figure, $N = 100$. The figure indicates that α increases as η/λ increases. This result is consistent with our intuition that if call arrivals are infrequent, the VLR records are unlikely to be replaced. We also observe that $\alpha > 99.6\%$ when $M > 1.3N$.

Fig. 11(b) plots the α curves for $N = 100$, 200, and 300, respectively. In this figure, $\eta = 0.1\lambda$. The figure shows that for a fixed M/N value, α increases as N increases.

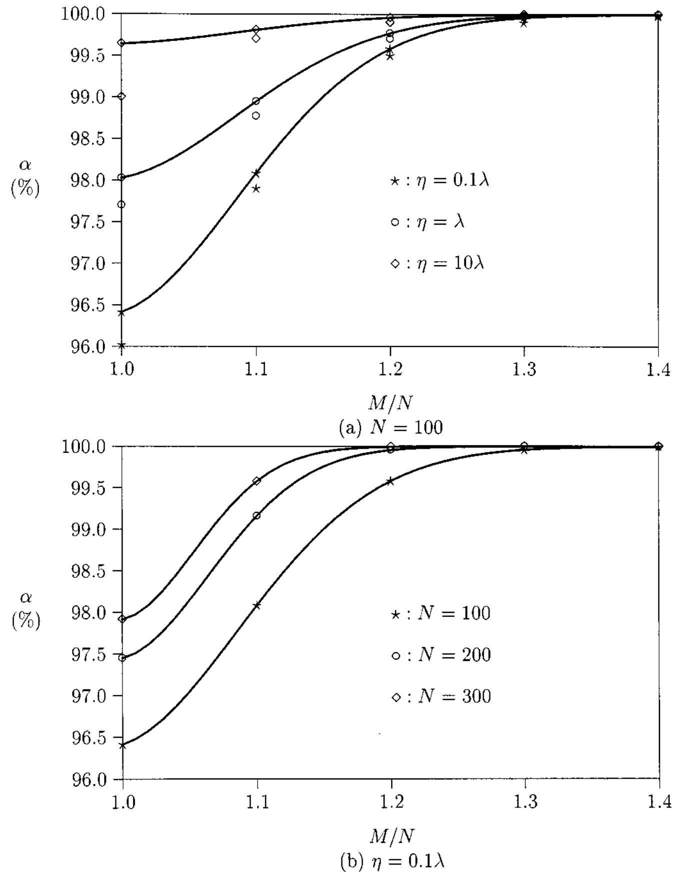


Fig. 11. The probability α that the VLR record of a mobile user is never replaced due to VLR overflow.

C. Issue III: Derivation for β

The probability β (i.e., the probability that the call activities of a user is not affected by VLR overflow) is derived as follows. Suppose that there are $n > M$ users in the LA. Starting from time $t = 0$, let $\beta(t_1, k)$ be the density function that a VLR record x is replaced at time t_1, and there are $k-1$ replacements to other VLR records during the period $[0, t_1)$. Then

$$\beta(t_1, k) = (1-q)^{k-1} q \left[\frac{(\mu t_1)^{k-1}}{(k-1)!} \right] \mu e^{-\mu t_1}. \tag{8}$$

On the right-hand side of (8), the term $(1-q)^{k-1}$ means that the first $k-1$ replacements are made to other VLR records. The kth replacement is made to record x with probability q. From Assumption III, the interarrival times for calls (that will result in VLR record replacements) have an exponential distribution with rate $\mu = (n-M)\lambda$ [cf., (3)], and the sum of the k call interarrival times has an Erlang density function given in the remaining part of the right-hand side of (8).

Consider the timing diagram in Fig. 12. Suppose that a user enters the LA at time zero and leaves the LA at time t. The VLR record of the user is replaced at time t_1. After t_1, the first call to/from the user occurs at time $t_1 + t_2$. The call setup procedure for this call is not affected by VLR overflow if $t_1 + t_2 > t$. Suppose that there are k record replacements in the VLR during the period $[0, t_1]$. Let $\theta(\mu, k)$ be the probability that the call

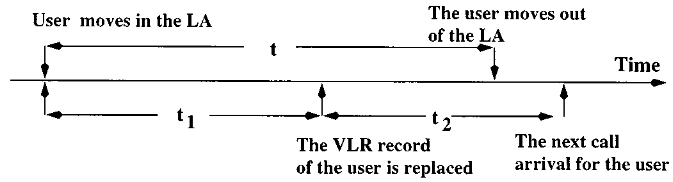


Fig. 12. The timing diagram for deriving β.

setup procedure will be affected by VLR overflow (i.e., $t_1 + t_2 < t$). From Appendix A, we have

$$\begin{aligned} \theta(\mu, k) &= (1-q)^{k-1} q \left\{ 1 - \sum_{l=0}^{k-1} \left[\frac{(-\mu)^l}{l!} \right] \frac{d^l f^*(s)}{ds^l} \bigg|_{s=\mu} \right. \\ &\quad - \left(\frac{\mu}{\mu - \lambda} \right)^k \left\{ f^*(\lambda) - \sum_{l=0}^{k-1} \left[\frac{(\lambda - \mu)^l}{l!} \right] \right. \\ &\quad \left. \left. \times \left[\frac{d^l f^*(s)}{ds^l} \bigg|_{s=\mu} \right] \right\} \right\} \end{aligned} \tag{9}$$

and

$$\beta = 1 - \sum_{n=M+1}^{\infty} \sum_{k=1}^{\infty} \theta((n-M)\lambda, k) \pi_n. \tag{10}$$

For exponential LA residence times, $f^*(s)$ in (9) is replaced by (6) to yield

$$\theta(\mu, k) = \frac{\lambda q \mu^k (1-q)^{k-1}}{(\eta + \lambda)(\mu + \eta)^k}. \tag{11}$$

Substituting (11) into (10), we have

$$\beta = 1 - \sum_{n=M+1}^{\infty} \frac{\lambda^2 q (n-M) \pi_n}{(\eta + \lambda)[\eta + (n-M)\lambda q]}. \tag{12}$$

Based on (12), Table I lists β as a function of M/N for various N and η/λ values. The figure indicates that under the range of input parameters considered in our study, $\beta > 99.5\%$ for $M > N$. In other words, the results indicate that even if the VLR overflow probability is high (e.g., $p_{\text{ov}} \simeq 48\%$ for $M = N$ in Fig. 10), the call setup procedure is only slightly affected ($(1-\beta) < 0.5\%$) when the VLR overflow algorithm is exercised.

V. Conclusion

This paper studied the mobility database overflow problem in a cellular system. If the mobility database (specifically, the VLR) for a cellular service LA is full, the incoming mobile users cannot receive services under the existing cellular phone technology. To allow these "overflow" users to receive cellular services, we proposed a VLR overflow control scheme. The scheme can be easily adopted in the existing cellular systems such as GSM and AMPS with only minor modifications. Particularly, no modifications are made to the mobile phones.

TABLE I
THE PROBABILITY β THAT THE CALL ACTIVITIES OF A USER ARE NOT AFFECTED BY VLR OVERFLOW

M/N	1.0	1.1	1.2	1.3	1.4
$\eta = 0.1\lambda$	99.673	99.840	99.967	99.997	99.999
$\eta = 1.0\lambda$	99.980	99.990	99.998	99.999	100.00
$\eta = 10\lambda$	99.999	99.999	100.0	100.0	100.0

(a) $N = 100$

$\eta = 0.1\lambda$	99.878	99.963	99.998	100.0	100.0
$\eta = 1.0\lambda$	99.993	99.997	99.999	100.0	100.0
$\eta = 10.0\lambda$	99.999	100.0	100.0	100.0	100.0

(b) $N = 200$

$\eta = 0.1\lambda$	99.932	99.987	99.999	100.0	100.0
$\eta = 1.0\lambda$	99.996	99.999	100.0	100.0	100.0
$\eta = 10.0\lambda$	99.999	100.0	100.0	100.0	100.0

(c) $N = 300$

We also studied three performance issues for VLR overflow. Let N be the expected number of mobile users in an LA and M be the size of the corresponding VLR. Under the range of input parameters we considered, the following results were observed.

- The VLR overflow probability is less than 0.2% for $M > 1.3N$.
- When the overflow control scheme is exercised, a mobile user's VLR record may be replaced. We showed that this "replacement" probability is less than 0.4% for $M > 1.3N$.
- In the VLR overflow control scheme, the call activities of a "replaced" user are affected (and extra effort is required in the call setup procedure) if, after the VLR record replacement, the user has call activity before moving out of the LA. Our study indicated that the call activities of a mobile user is unlikely to be affected (e.g., a user is affected with probability less than 0.4% if $M > N$).

The last observation merits further discussion. When $M = N$, about 50% of the incoming users to the LA cannot receive service under existing cellular technology. By exercising the overflow control scheme, all the mobile users in the LA continue to receive the service. The extra overhead for the scheme is very low: most of the call setups (with a probability larger than 99.6%) to a user is not affected by the overflow control scheme.

This work assumed that every VLR is associated with an LA and the replacement policy is uniformly random. One of the research directions is to extend the analytic model for the system with multiple LA's per VLR and to consider various replacement policies.

APPENDIX I
THE DERIVATION FOR $\theta(\mu, k)$

Consider the timing diagram in Fig. 12. Suppose that a user enters the LA at time zero and leaves the LA at time t. The VLR record of the user is replaced at time t_1. After t_1, the first call to/from the user occurs at time $t_1 + t_2$. The call setup procedure for this call is not affected by VLR overflow if $t_1 + t_2 > t$. Suppose that there are k record replacements in the VLR during the period $[0, t_1]$. Let $\theta(\mu, k)$ be the probability that the call setup procedure will be affected by VLR overflow (i.e., $t_1 + t_2 < t$). Since the density functions for t, t_1, and t_2 are $f(t), \beta(t_1, k)$, and $\lambda e^{-\lambda t_2}$, respectively, we have

$$\begin{aligned}\theta(\mu, k) &= \int_{t_1=0}^{\infty}\int_{t=t_1}^{\infty}\int_{t_2=0}^{t-t_1} f(t)\beta(t_1,k)\lambda e^{-\lambda t_2}\,dt_2\,dt\,dt_1\\ &= \int_{t_1=0}^{\infty}\int_{t=t_1}^{\infty} f(t)\beta(t_1,k)\left[1-e^{-\lambda(t-t_1)}\right]dt\,dt_1\\ &= \int_{t_1=0}^{\infty}\int_{t=t_1}^{\infty}(A-B)\,dt\,dt_1 \qquad (13)\end{aligned}$$

where $A = f(t)\beta(t_1, k)$ and $B = f(t)\beta(t_1, k)e^{-\lambda(t-t_1)}$. From (8) and after rearrangement

$$\begin{aligned}\int_{t_1=0}^{\infty}\int_{t=t_1}^{\infty} A\,dt\,dt_1 &= \int_{t=0}^{\infty}\int_{t_1=0}^{t} f(t)(1-q)^{k-1}q\\ &\quad\times\left[\frac{(\mu t_1)^{k-1}}{(k-1)!}\right]\mu e^{-\mu t_1}\,dt_1\,dt\\ &= \int_{t=0}^{\infty} f(t)(1-q)^{k-1}q\\ &\quad\times\left\{\int_{t_1=0}^{t}\left[\frac{(\mu t_1)^{k-1}}{(k-1)!}\right]\mu e^{-\mu t_1}\,dt_1\right\}dt\\ &= \int_{t=0}^{\infty} f(t)(1-q)^{k-1}q\\ &\quad\times\left[1-\sum_{l=0}^{k-1}\frac{(\mu t)^l}{l!}e^{-\mu t}\right]dt\\ &= \int_{t=0}^{\infty}(A_1-A_2)\,dt \qquad (14)\end{aligned}$$

where

$$A_1 = f(t)(1-q)^{k-1}q$$

and

$$A_2 = f(t)(1-q)^{k-1}q\left[\sum_{l=0}^{k-1}\frac{(\mu t)^l}{l!}e^{-\mu t}\right].$$

Thus

$$\int_{t=0}^{\infty} A_1\,dt = \int_{t=0}^{\infty} f(t)(1-q)^{k-1}q\,dt = (1-q)^{k-1}q \qquad (15)$$

and

$$\begin{aligned}&\int_{t=0}^{\infty} A_2\,dt\\ &= \int_{t=0}^{\infty} f(t)(1-q)^{k-1}q\left[\sum_{l=0}^{k-1}\frac{(\mu t)^l}{l!}e^{-\mu t}\right]dt\\ &= \sum_{l=0}^{k-1}(1-q)^{k-1}q\mu^l\int_{t=0}^{\infty}\frac{t^l}{l!}f(t)e^{-\mu t}\,dt\\ &= (1-q)^{k-1}q\left\{\sum_{l=0}^{k-1}\left[\frac{(-\mu)^l}{l!}\right]\left.\frac{d^l f^*(s)}{ds^l}\right|_{s=\mu}\right\}. \qquad (16)\end{aligned}$$

Substituting (15) and (16) into (14), we have

$$\begin{aligned}\int_{t_1=0}^{\infty}\int_{t=t_1}^{\infty} A\,dt\,dt_1 &= (1-q)^{k-1}q \\ &\quad \times \left\{1-\sum_{l=0}^{k-1}\left[\frac{(-\mu)^l}{l!}\right]\left.\frac{d^l f^*(s)}{ds^l}\right|_{s=\mu}\right\}. \qquad (17)\end{aligned}$$

From (13) and after rearrangement

$$\begin{aligned}&\int_{t_1=0}^{\infty}\int_{t=t_1}^{\infty} B\,dt\,dt_1 \\ &= \int_{t=0}^{\infty}\int_{t_1=0}^{t} f(t)\beta(t_1,k)e^{-\lambda(t-t_1)}\,dt_1\,dt \\ &= \int_{t=0}^{\infty} f(t)e^{-\lambda t}\left[\int_{t_1=0}^{t}\beta(t_1,k)e^{\lambda t_1}\,dt_1\right]dt \\ &= \int_{t=0}^{\infty} f(t)e^{-\lambda t}(1-q)^{k-1}q\left(\frac{\mu}{\mu-\lambda}\right)^k \\ &\quad \times \left[1-\sum_{l=0}^{k-1}\frac{[(\mu-\lambda)t]^l}{l!}e^{-(\mu-\lambda)t}\right]dt \\ &= (1-q)^{k-1}q\left(\frac{\mu}{\mu-\lambda}\right)^k\left\{f^*(\lambda)-\sum_{l=0}^{k-1}\int_{t=0}^{\infty}f(t)\right. \\ &\quad \left.\times\, e^{-\lambda t}\left\{\frac{[(\mu-\lambda)t]^l}{l!}\right\}e^{-(\mu-\lambda)t}\,dt\right\} \\ &= (1-q)^{k-1}q\left(\frac{\mu}{\mu-\lambda}\right)^k\left\{f^*(\lambda)-\sum_{l=0}^{k-1}\left[\frac{(\mu-\lambda)^l}{l!}\right]\right. \\ &\quad \left.\times\int_{t=0}^{\infty}t^l f(t)e^{-\mu t}\,dt\right\} \\ &= (1-q)^{k-1}q\left(\frac{\mu}{\mu-\lambda}\right)^k\left\{f^*(\lambda)\right. \\ &\quad \left.-\sum_{l=0}^{k-1}\left[\frac{(\lambda-\mu)^l}{l!}\right]\left[\left.\frac{d^l f^*(s)}{ds^l}\right|_{s=\mu}\right]\right\}. \qquad (18)\end{aligned}$$

Substituting (17) and (18) into (13), we have

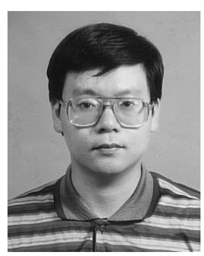
$$\begin{aligned}&\theta(\mu,k) \\ &= (1-q)^{k-1}q \\ &\quad \cdot\left\{1-\sum_{l=0}^{k-1}\left[\frac{(-\mu)^l}{l!}\right]\right. \\ &\quad \times \left.\frac{d^l f^*(s)}{ds^l}\right|_{s=\mu}-\left(\frac{\mu}{\mu-\lambda}\right)^k \\ &\quad \cdot\left\{f^*(\lambda)-\sum_{l=0}^{k-1}\left[\frac{(\lambda-\mu)^l}{l!}\right]\left[\left.\frac{d^l f^*(s)}{ds^l}\right|_{s=\mu}\right]\right\}\Bigg\}. \qquad (19)\end{aligned}$$

ACKNOWLEDGMENT

The author would like to thank the anonymous reviewers for their valuable comments which significantly improved the quality of this work. Specifically, Assumption VI was established with the assistance of the reviewers.

REFERENCES

[1] W. C. Y. Lee, *Mobile Cellular Telecommunications Systems*. New York: McGraw-Hill, 1995.
[2] ——, *Mobile Communications Engineering*. New York: McGraw-Hill, 1997.
[3] D. C. Cox, “Wireless personal communications: What is it?,” *IEEE Personal Commun. Mag.*, pp. 20–35, Apr. 1995.
[4] “Cellular intersystem operations (Rev. C),” EIA/TIA, Tech. Rep. IS-4, 1995.
[5] “Mobile application part (MAP) specification, version 4.8.0,” ETSI/TC, Tech. Rep., Recommendation GSM 09.02, 1994.
[6] Y.-B. Lin, “Modeling techniques for large-scale PCS networks,” *IEEE Commun. Mag.*, vol. 35, Feb. 1997.
[7] Y.-B. Lin and A. Noerpel, “Implicit deregistration in a PCS network,” *IEEE Trans. Veh. Technol.*, vol. 43, no. 4, pp. 1006–1010, 1994.
[8] Y.-B. Lin and W. Chen, “Impact of busy lines and mobility on call blocking in a PCS network,” *Int. J. Commun. Syst.*, vol. 9, pp. 35–45, 1996.
[9] L. Kleinrock, *Queueing Systems: Volume I—Theory*. New York: Wiley, 1976.



Yi-Bing Lin (S'80–M'96–SM'96) received the B.S.E.E. degree from National Cheng Kung University, Taiwan, R.O.C., in 1983 and the Ph.D. degree in computer science from the University of Washington, Seattle, in 1990.

From 1990 to 1995, he was with the Applied Research Area, 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), Hsinchu, Taiwan. In 1996, he was appointed as Deputy Director of the 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. He is an Associate Editor of the IEEE NETWORK. He was a Guest Editor of the IEEE TRANSACTIONS ON COMPUTERS Special Issue on Mobile Computing.

Dr. Lin is an Associate Editor of *SIMULATION Magazine*, an Area Editor of the *ACM Mobile Computing and Communication Review*, a Columnist of *ACM Simulation Digest*, a member of the editorial board of the *International Journal of Communications Systems*, a member of the editorial board of *ACM/Baltzer Wireless Networks*, a member of the editorial board of *Computer Simulation Modeling and Analysis*, Guest Editor for the *ACM/Baltzer MONET* Special Issue on Personal Communications, and an Editor of the *Journal of Information Science and Engineering*. He was the 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, and Publicity Chair of ACM Sigmobile. He received the 1997 Outstanding Research Award from the National Science Council, Taiwan, and the Outstanding Youth Electrical Engineer Award from CIEE, Taiwan.