19

Analysis of Performance and Fault-Tolerance of Database Management Schemes for PCS

Dipak Ghosal,
Bellcore,
331 Newman Springs Road,
Red Bank, NJ 07701-7020.

Abstract

This paper presents the performance and fault-tolerance analysis of four different placement schemes for the HLR (Home Location Register) and the VLR (Visitor Location Register) databases in the signalling network. Based on a mean value analysis it has been shown that distributing the HLR and the VLR database may be required for high degree of penetration of PCS usage. A simple distributed scheme where the HLR is maintained in the Service Control Point (SCP) and VLR is co-located (physically or logically) in the SP (Signalling Point) can significantly reduce the impact of PCS signalling on the existing signalling network. Performance of distributed architectures with replication depends largely on how the databases are replicated. The results show that performance gains are achievable when the replication is based on the roaming characteristics of the mobile station. Fault-tolerance of the different placement strategies is based on a simple failure and recovery model of the databases. The results show that the system can be made more robust against database failures with a distributed and replicated database placement strategy if all the distributed replicated copies are always consistent with each other. This may incur additional performance penalties. The analysis is based on a number of simplifying assumptions and is only intended for a comparative study of different database placement schemes.

S. V. Raghavan et al. (eds.), *Computer Networks, Architecture and Applications*

1 INTRODUCTION

Personal Communication Services (PCS) require significant intelligence in the signalling network to keep track of the mobile subscribers and appropriately route the calls. This study focuses on the performance and fault-tolerance issues of the databases that are required to provide PCS in a high density and highly mobile subscriber community.

In this study we consider a network supporting PCS providers offering cellular services to end customers. The mobile network considered in this study consists of wireless mobile terminals/stations (MS) which have radio access capability with base stations (BS). The type of terminal equipment and its radio access capability depends largely on the mobility of the user [Cox, 1990; Steel, 1990]. The service area consists of cells, each of which is serviced by a base station. A number of base stations is served by a controller such as a Mobile Switching Center (MSC) which is assumed to be co-located in the SP and provides access to the Public Switched Telephone Network (PSTN) [Kathy et al., 1992; CCITT, 1988]. The call control and routing are based on information maintained in two databases, namely the Home Location Register (HLR) and the Visitor Location Register (VLR). These are databases which store user profiles related to the nature of the service to which the user has subscribed. The HLR holds the permanent subscriber parameters while the VLR contains the database of subscriber parameters within a particular location area. In this study, we are interested in the placement strategies of these databases in the signalling network and to determine their impact on the performance and fault-tolerance of the PCS.

The signalling network considered in this study is the Common Channel Signalling (CCS) network based on the Signalling System Number 7 protocol [Bellcore, 1989]. The CCS network consists of the Signalling Points (SPs) interconnected via Signalling Transfer Points (STPs). The network database is provided in the Service Control Points (SCPs) which are connected to the STPs. SCPs are centralized databases which receive queries from the Signalling Points (SPs) and return the appropriate information. The 800 services and Calling Card are examples of two services that are provided by the SCP [Homa et al., 1992; Skoog et al., 1990].

The issue of SCP query and update rates for PCS is also studied in [Lo et al., 1992] in which the authors have considered support of end-to-end PCS with the subscriber data co-located in the regional SCP. Based on a Monte-Carlo simulation it has been shown that the mean real-time query rate can be up to an order of magnitude higher than the estimated load on the SCP due to 800 services. Furthermore, the real-time update rate is comparable to the query volume.

Both the above results show that providing region wide end-to-end PCS service solely based on the performance level of current SCP technology may impose limitations on the degree of penetration and mobility of the mobile terminal users. The effect of relocating databases to the SPs has been addressed in [Kathy et al., 1991]. In [Kathy et al., 1991] the authors have quantified the SS7 traffic load associated with the terminal mobility when Pan European standard GSM is used to provide PCS. Based on a detailed analysis of the SS7 signalling messages required for GSM, the authors have determined the impact of GSM PCN and GSM cellular with respect to normal POTS traffic. The results show that the placement of the VLR has impact on the additional traffic that must be handled by the SS7 network due to PCS.

This paper extends the work reported in [Kathy et al., 1992] in a number of ways. First it provides a more detailed analysis of four different placement strategies which involve both the HLR and the VLR. In [Kathy et al., 1992] only the placement of the VLR is considered while the HLR is assumed to be always outside the MSC. Second, in this paper we consider a more accurate characterization of the traffic mixes and environments. Finally, this study also provides analysis of the fault-tolerance of the system under different database placement strategies.

In this paper we investigate the following four placement strategies for HLR and VLR.

1. HLR and VLR in SCP.
2. HLR in SCP and VLR in SP.
3. Master copy of HLR in SCP and cached copy in SP. VLR in SP.
4. Master copy of HLR in SCP and cached copies replicated in SPs. VLR in SP.

Clearly these strategies provide different performance and fault-tolerance capability for personal communication services. When HLR and VLR are co-located in the SCP, many call set-up/tear-down messages and handovers and location update messages require queries to the SCP. This will require a large number of messages to be exchanged across the signalling network. Furthermore, since the SCP is a centralized database from the services and protocol perspectives, any congestion and/or isolation of the SCP can severely affect the personal communication services to a large user community. Moving the VLR to the SP can improve performance but not the fault-tolerance of the system [Gray et al., 1993]. Maintaining master copy of HLR in SCP and cached copies replicated in one or more SPs along with the VLR can improve fault-tolerance but only at the cost of performance. Thus there is a trade-off between the performance and fault tolerance for the different placement strategies. This study is an attempt to quantify this trade-off.

This paper is organized as follows. Section 2 gives a detailed description of the system analyzed in this paper. The various messages that are required for mobile call set-up, handovers and location updates are discussed in this section. The assumptions made in the analysis are discussed in Section 3. One of the key parameters of the system is the model of the MS and characterization of their mobility. This is also discussed in Section 3. A summary of the various parameters of the system is also tabulated in this section.

Section 4 discusses the four placement strategies and their performance impact on the signalling network. The fault-tolerance of the system under different placement strategies is discussed in Section 5. Finally Section 6 concludes this paper and points out some future research directions.

2 SYSTEM DESCRIPTIONS

This study is based on the network shown in Figure 1. We consider a total service area A_{sa} which is divided into L equal sized location areas of size A_{la}. Each location area is serviced by a controller referred to as the Mobile Switching Centre (MSC) which is co-located in the Signalling Point (SP) which also provides wire-line access services.

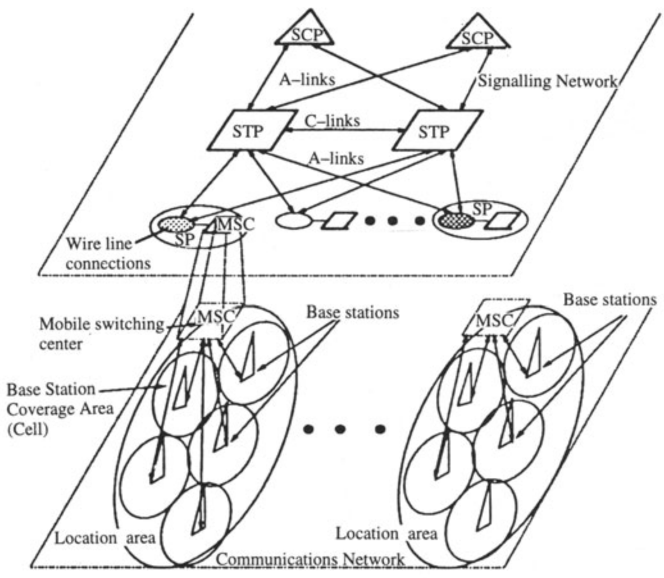

Figure 1: Internetworking between the wireless and the signalling networks.

Each location area is divided into equal sized cells each of area A_{cell}. Each cell is served by a base station which provides the radio interface to the wireless mobile terminals/stations (MS) within the coverage area of the base station. The signalling network is the Common Channel Signalling (CCS) network based on the SS7 protocol [Bellcore, 1989; Skoog et al., 1990]. It consists of SPs connected via A-links to a Signalling Transfer Point (STP) pair which in turn is connected by C-links. The STP is a packet switch that routes SS7 signalling messages between SPs. The SS7 messages are call control messages for call set-up and disconnect and network control messages for maintaining the correct and stable operation of the signalling network. The network database is provided in the Signalling Control Points (SCPs) which is connected to the STP pair through A-links.

As mentioned in the introduction, call control and routing in the mobile network is performed based on two databases, namely, the Home Location Register (HLR) and the Visitor Location Register (VLR) [CCITT, 1988]. The HLR contains the permanent and temporary data about each subscriber. The permanent data includes the current services the subscriber receives and temporary data includes a pointer to the current VLR which services the subscriber. There is one VLR for each location area and it contains a list of MS stations that are currently registered in that location area. For each MS, the VLR also contains a copy of the permanent data stored in the HLR. In this study we will assume that each location area is served by one VLR and one VLR serves only one location area.

Because the MS can roam from one location to another both the VLR and the HLR need to be updated with the most recent location of the MS. This study focuses on the placement strategies of these databases in the signalling network and their impact on the performance and fault-tolerance of PCS. In order to perform a detailed analysis of database location strategies it is necessary to understand the various messages that are needed for call set-up/tear-down, location update and handovers. These are described in the following sub-sections with respect to the IS-41 protocol [Homa et al., 1992].

2.1 Call Set-up/Tear-down

Call set-up with or from a mobile station, in general, involves three major tasks: *terminal control, routing* and *call control.* Terminal control includes set-up functions between the mobile station and the controlling MSC. These include set-up radio channels and authentication and validation procedures and may require access of the VLR in which the mobile station is currently registered. Routing tasks are required to find the current location of the mobile station so that the call can be appropriately routed. This information is obtained from the HLR of the mobile station. Finally, call control functions relate to the messages that are exchanged between the controlling SPs to set-up the call.

Figure 2 shows the messages that are exchanged for a call set-up originating at a MS and terminating at a wire-line equipment. The originating SP sends a call set-up request to the controlling VLR which invokes application processes to manage the call set-up with the mobile station. The primary function of these processes involves authenticating the originating mobile station and providing communication privacy. This requires messages

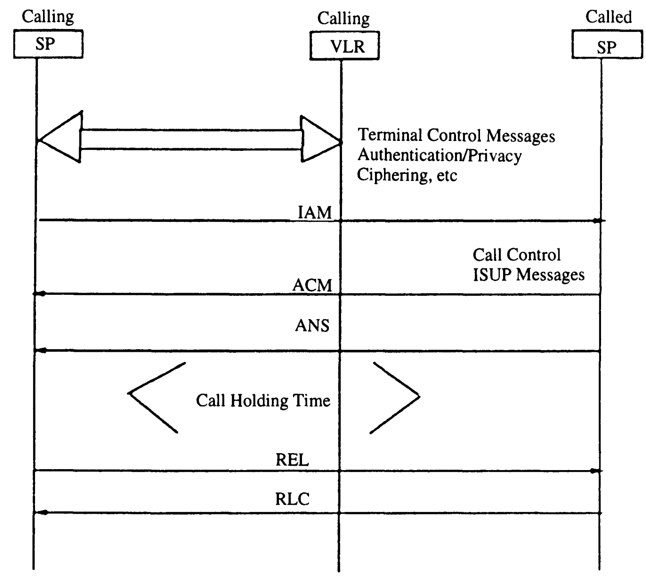

Figure 2: Messages exchanged for call set-up originating at a MS.

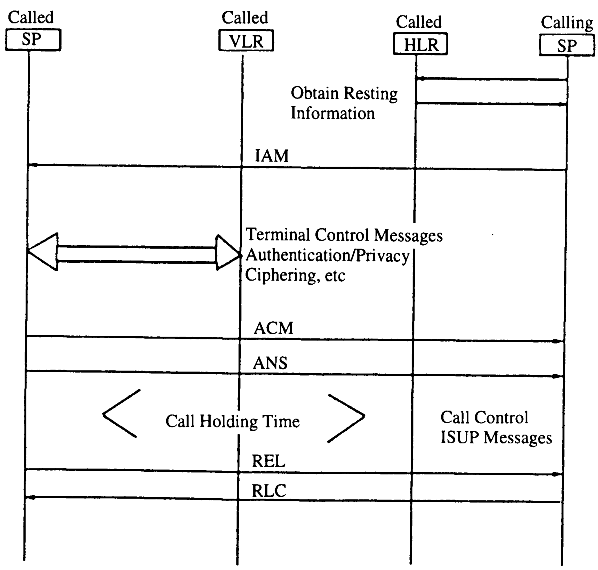

Figure 3: Messages exchanged for call-set terminating at a MS.

to be exchanged between the VLR and the SP. The number of such messages depends on the algorithms used for session key agreement and terminal authentication. In this study we assume that 4 messages are required in each direction to achieve terminal control functions. Note that since this is only a comparative study, an exact enumeration of various messages will not significantly alter the conclusions of this study.

The call control functions are achieved by the standard SS7 call set-up messages, namely Initial Address Message (IAM), Address Complete Message (ACM) and Answer Message (ANS). The call is cleared by the Release(REL) and Release Clear (RLC) messages. If the called number is also a mobile station, then the routing function must be invoked to determine the controlling VLR of the terminating mobile station. This is done by querying the HLR which contains a pointer to the VLR where the terminating mobile station is registered. This is achieved by the **loqreq** and **locreq acknowledge** messages as shown in Figure 3. The remainder of the call set-up involves terminal control functions for the terminating mobile station and call control functions between the controlling SPs.

2.2 Location update

Location update is required whenever the mobile station crosses location areas. Location update involves three main functions; 1) updating the location information maintained in the HLR (i.e., the pointer to the VLR), 2) registering the mobile station with the VLR which controls the new location area, and 3) cancelling the registration in the previous VLR. Figure 4 shows the different messages which achieve these tasks. Note that the new VLR obtains the subscriber parameters and authorization parameters from the HLR using the *profreq* and *qualreq* messages respectively.

2.3 Handovers

There are three types of handovers[CCITT, 1988];

1. Handovers between radio channels of the same radio port. This arises primarily due to interference or disturbances in the radio channels;
2. Handovers between base radio ports within the same SP; and
3. Handovers between radio ports of different SPs.

In this study we will assume that handovers of types 1 and 2 will be handled by the controlling SP and do not require any access to the signalling network. Inter-SP handovers, i.e., handovers of type 3, is similar to a call set-up and provides a through connection between the SPs involved in the handover. Inter-MSC handovers based on the IS-41 MAP use dedicated facilities between MSCs and thus do not require ISUP capabilities. The messages are exchanged between the MSCs involved in the handover and consists of 3 invoke and 2 response messages. These are ***hand-off measure request, facilities directive*** and ***mobile channel*** invoke messages and ***hand-off measure request*** and ***facilities directive*** response messages. In this study we will assume that inter-SP handovers require 4 inter-SP messages in each direction and one query to the VLR database.

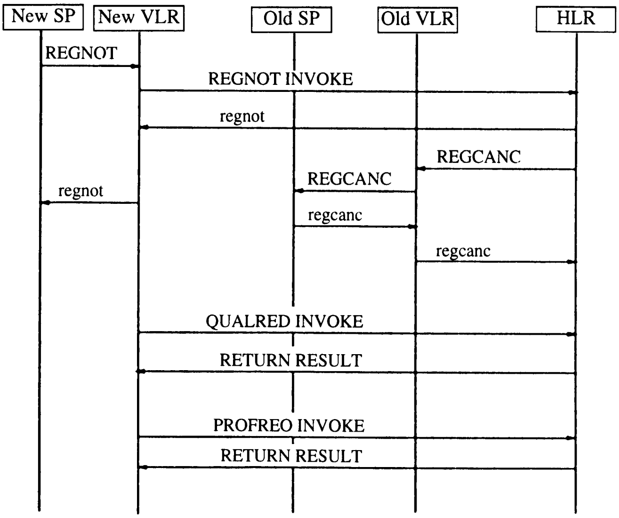

Figure 4: Messages exchanged during location update.

3 MODEL AND ASSUMPTIONS

In order to simplify the analysis we will make the following assumptions regarding the network;

Assumption 1 : We will assume that Figure 1 represents the entire network, i.e., all calls originate and terminate with the subscribers connected to the network.

Assumption 2 : We will assume that each location area is served by one VLR and one VLR serves only one location area.

Assumption 3 : We will assume that the cells in a location area are identical in size.

Let A_{la} denote the size of the location area in square kilometers and N denote the number of subscribers in the location area. Thus, N also represents the size (in terms of number of lines) of the SP serving the location area.

3.1 Traffic mixes and environments

Let Λ denote the number of calls per hour serviced by an SP in an existing POTS (Plain Old Telephone System). These calls can be approximately[1] divided into three components: 1) Outgoing calls which constitute 40% of the overall calls, 2) Incoming calls which also constitute 40% of the overall calls and 3) Intra-SP calls which originate and terminate in the SP constitute 20% of the calls. We will assume the following three types of SPs:

METRO : An end-office in a major metropolitan area with 50,000 to 70,000 lines serving 3.59 calls per line per hour in a HDBH (High Day Busy Hour) period.

Single System City : An end-office in a medium size town with 20,000 to 50,000 lines serving 2.9 calls per line per hour during a HDBH period.

Suburban : An end-office residing in the outskirts of a metropolitan area with 10,000 to 30,000 lines generating 1.9 calls per line per hour in a HDBH period.

Let λ_w (λ_m) denote the number of calls per hour made by a wireline (mobile) subscriber. Also, let t_{hw} (t_{hm}) denote the mean call holding time for a wireline (mobile) call.

3.2 Performance measure

In the CCS network, A-links connect the SPs with the STP pair. In this study we assume that there is one 56 Kbps link in each direction from an SP to an STP. We will use the utilization of the A-link in the outgoing direction as our performance measure. Note that one performance objective is to maintain the A-link utilization below 40%. The A-link utilization, U, assuming only POTS call and 30 bytes per message, is given by

$$\begin{aligned} U &= \Lambda \times \textit{fraction of o/g POTS calls per sec} \\ &\quad \times \textit{messages per call} \times \textit{bits per message} \\ &= \frac{\Lambda \times 0.4 \times 30 \times 8 \times 5}{2 \times 36 \times 56 \times 10^5} \end{aligned} \qquad (1)$$

where Λ denotes the calls per hour serviced by the SP. From the above relation we can obtain the A-link utilization for different SP calling rates. In the following analysis, this will be used as the base case for comparing the various database placement schemes.

3.3 Mobility model

Let ξ denote the penetration of PCS usage. This implies that in a location area of N users, ξN are mobile stations and $(1-\xi)N$ are wireline subscribers. Furthermore, for Λ

[1] For exact numbers it is necessary to account for the mis-dials and calls which terminate due to a busy signal

Table 1: Number of outgoing messages per call type for HLR and VLR both in SCP. WL : Wireline, MS : Mobile station

Call types	Outgoing calls	Intra-SP calls
WL → WL	5	0
WL → MS	10	5
MS → WL	9	4
MS → MS	14	9

Table 2: Number of outgoing messages per call type for HLR and VLR both in SCP.

Event types	Handover	Location Update
Outgoing messages	9	2

calls per hour served by an SP, the proportions of different call types assuming that a person calls any other person with equal probability are as follows;

$$\begin{aligned} MS \rightarrow MS &= \xi^2 \Lambda \\ MS \rightarrow Wireline &= (1-\xi)\xi\Lambda \\ Wireline \rightarrow wireline &= (1-\xi)^2\Lambda \\ Wireline \rightarrow MS &= (1-\xi)\xi\Lambda \end{aligned} \tag{2}$$

In order to evaluate the impact due to PCS signalling, it is necessary to determine the number of handovers and location updates. Let λ_{ho} and λ_{lu} denote the number of handovers and location updates per hour. These are obtained based on the simple flow model described in [Thomas et al., 1988; Kathy et al., 1991]. Based on the assumption that the MSs are uniformly distributed on the surface of the cell and the direction of the movement is uniformly distributed between $[0, 2\pi]$, it can be shown that the number of boundary crossings per hour, $R_{cross}(area)$, is given by

$$R_{cross}(area) = V\rho L/\pi \tag{3}$$

where V is the mean velocity, ρ is the density of mobile station in the area and L is the length of the boundary of the area.

Now, the inter-SP handovers are location area crossings while the mobile station is active, i.e., there is a call in progress. Thus

$$\lambda_{ho} = R_{cross}(A_{la}) \times s \tag{4}$$

where s is equal to the fraction of an hour a mobile station is active. The latter depends on the call holding time and the number of calls per hour made by the mobile station and is given by

$$s = \frac{\lambda_m \times t_{mw}}{3600} \quad (5)$$

Note that the above equation gives the number of handovers out of the area. By the fluid flow approximation it is also equal to the number of handovers into the area. Based on the same idea, the number of location updates per hour is given by

$$\lambda_{lu} = R_{across}(A_{la}) \times (1 - s) \quad (6)$$

This is because the number of location updates due to location area departures is equal to the number of location area crossings while the terminal is idle.

4 DATABASE PLACEMENT STRATEGIES

In this section we consider the various database placement strategies and analyze their performance impact on the signalling network.

4.1 VLR and HLR in SCP

In this scheme the SCP maintains both the HLR for each subscriber and the VLR for each location area. As a result all terminal control and routing messages which require access to the VLR or the HLR must be sent to the SCP. Note that in this scheme even for intra-SP calls in which at least one mobile station is involved will require access to the SCP. In order to determine the link utilization for this placement strategy, it is necessary to quantify the number of messages in the outgoing direction. This is shown in Table 1.

The above table is obtained from the figures in Section 2 which illustrate the messages that are exchanged during call set-up. For example the number of outgoing messages for an outgoing call from a wireline to a mobile station consists of 1 message to the HLR to obtain the SP of the terminating mobile station and 5 POTS message to set up the call. Note that the proportion of different call types can be found using Equation (2).

The other factor which will affect the link utilization is the number of messages required to perform handovers and location updates. From the call flows described in Sections 2.2 and 2.3 and making the fluid flow assumption that the number of handovers (location updates) due to mobile stations roaming out of the location area is equal to the number of handovers (location updates) due to mobile stations roaming into the location area, it is easy to obtain Table 2 which shows the number of outgoing messages for every handover and location update.

Let $\lambda_{o/g}$ and λ_{in} denote the number of outgoing and intra-SP calls per hour, respectively. Using Table 1 and Table 2 and assuming a mean message length of 30 bytes it can be shown that the link utilization U is given by

Table 3: Number of outgoing messages per call type for HLR in SCP and VLR in SP

Call types	Outgoing calls	Intra-SP calls
WL → WL	5	0
WL → MS	6	1
MS → WL	5	0
MS → MS	6	1

Table 4: Number of outgoing messages for handover and location updates for HLR in SCP and VLR in SP.

Event types	Handover	Location Update
Outgoing messages	8	4

$$U = \frac{\lambda_{o/g}\times 30\times 8}{36\times 56\times 10^5}[5(1-\xi)^2 + 15(1-\xi)\xi + 10\xi^2] + \frac{\lambda_{in}\times 30\times 8}{36\times 56\times 10^5}[9(1-\xi)\xi + 9\xi^2] \frac{\lambda_{ho}\times 30\times 8\times 9}{36\times 56\times 10^5} + \frac{\lambda_{lu}\times 30\times 8\times 2}{36\times 56\times 10^5} \quad (7)$$

where $\lambda_{o/g} = \frac{\Lambda\times 0.4}{2}$ and $\lambda_{in} = \frac{\Lambda\times 0.2}{2}$. Since PCS signalling messages will have higher average message length, assuming a lower number of bytes per message gives advantage to this placement scheme.

4.2 HLR in SCP and VLR in SP

In this scheme the HLR for each subscriber is stored in the SCP and the VLR for each location area is stored in the SP. We assume that the VLR contains a copy of the subscriber profile and can perform the authentication, validation and ciphering procedures. The HLR for each subscriber contains the permanent copy of the subscriber data and a pointer to the current VLR in which the subscriber is currently registered. This placement strategy significantly reduces the number of messages for call set-up as shown in Table 3.

Note that the intra-SP calls that terminate in a mobile station require accesses to the HLR to obtain the current VLR of the terminating mobile station. This is an assumption and may not be true under certain implementations. In this study we assume that when the SP determines the called number to that of a mobile station it obtains its current location from the HLR. Searching the local VLR can save an access to the HLR if the mobile station is currently registered in the same VLR.

The number of handovers and location update messages are shown in Table 4. The

Table 5: Number of outgoing messages per call type for HLR replicated in SCP and SP and VLR in SP

Call types	Outgoing calls	Intra-SP calls
WL → WL	5	0
WL → MS	7	p_1
MS → WL	5	0
MS → MS	6	p_1

number of handover messages decreases because the VLR is co-located in the SP. The number of location update messages increases because the messages between the two VLRs must be sent over the signalling network. Following the same approach as before, the link utilization can be written down as follows:

$$U = \frac{\lambda_{o/g}\times 30\times 8}{36\times 56\times 10^5}\,[5(1-\xi)^2 + 11(1-\xi)\xi + 6\xi^2] + \frac{\lambda_{in}\times 30\times 8}{36\times 56\times 10^5}\,[(1-\xi)\xi + \xi^2] \\ \frac{\lambda_{ho}\times 30\times 8\times 8}{36\times 56\times 10^5} + \frac{\lambda_{lu}\times 30\times 8\times 4}{36\times 56\times 10^5} \tag{8}$$

where $\lambda_{o/g}$ and λ_{in} are defined as before.

4.3 Master copy of HLR in SCP and cached copy in SP. VLR in SP

In this scheme each mobile station is assigned to a *home location area* and the permanent subscriber parameters of all subscribers with the same home location area are kept in a HLR database co-located in the SP. The HLR databases in each SP are referred to as the cached copies of the master HLR database which is maintained in the SCP. The VLR for each location area is also maintained in the SP of that location area. The home location area for a subscriber can be chosen to be the location area in which the mobile station is most often registered. The mobile station identifier of mobile stations with the same home location area has a unique identifier that can be recoginzed by all SPs for the purpose of routing the calls.

When the called number belongs to a mobile station, the calling SP queries the HLR in the home location area of the called mobile station to obtain the location of the controlling VLR. When the mobile station roams out of the home location area, both the cached HLR in the home location area and the master copy of HLR in the SCP are updated. However, note that HLR is not used in obtaining the routing information and thus can be updated at a much slower rate than the cached HLR and this rate will depend on the consistency that needs to be maintained between the master copy and the cached copy.

Table 6: Number of outgoing messages for handover and location updates for HLR replicated in SCP and VLR in SP.

Event types	Outgoing messages
Handover	8
Updating cached HLR	5.33
Updating master HLR	δ

The number of outgoing messages for intra-SP and outgoing calls are shown in Table 5. Note that the number of messages for outgoing calls terminating at a mobile station is greater than the previous case. The simple way to figure this out is to realize that in this scheme obtaining the routing information involves two SPs while in the previous case only the calling SP and the SCP is involved. In this case most intra-SP calls will be completed without requiring access to the signalling network. The only exception will be the intra-SP calls which terminate at mobile stations that belong to a different home location area. For these calls the routing information needs to be obtained from the home HLR which must be queried via the signalling network. The fraction of these calls is denoted by p_1.

The number of outgoing messages for location update and handover are shown in Table 6. The number of location update messages depends on the roaming characteristics of the mobile station with respect to the home location area. If the mobile station roams in two location areas which are different from the home location area then 8 outgoing messages are required per location update. This assumes that for every location area crossing in a particular direction by a subscriber there is another location area crossing by a similar subscriber in the opposite direction. On the other hand of one of the location areas is the home location area then the number of outgoing messages is 4. We assume that a mobile station roams randomly and uniformly in three location areas of which one is the home location area. Based on this assumption the mean number of location update messages is 5.33. The rate at which the master copy of the HLR is updated will be denoted by δ per hour.

Based on the above tables we derive the link utilization which is given by

$$
\begin{aligned}
U = & \frac{\lambda_{o/g}\times 30\times 8}{36\times 56\times 10^5}\,[5(1-\xi)^2 + 12(1-\xi)\xi + 7\xi^2] \\
& + \frac{\lambda_{in}\times 30\times 8}{36\times 56\times 10^5}\,[p_1(1-\xi)\xi + p_1\xi^2] \\
& \frac{\lambda_{ho}\times 30\times 8\times 8}{36\times 56\times 10^5} + \frac{(\lambda_{lu}+\delta)\times 30\times 8\times 5.33}{36\times 56\times 10^5}
\end{aligned}
\qquad (9)
$$

Table 7: Number of outgoing messages for handover and location updates for HLR replicated in SCP and neighboring SPs and VLR in SP.

Event types	Outgoing messages
Handover	9
Updating 2 HLRs	3.33
Updating 3 HLRs	2.0
Updating master HLR	δ

4.4 Master copy of HLR in SCP and cached copies of replicated in SPs. VLR in SP

This scheme is the same as before except that the HLR is cached into a number of nearby location areas. The VLR is still co-located in the SP. In this scheme the location update involves updating all the cached copies and updating the master copy at a certain rate. Note that if the HLR is properly replicated then the probability of a call to a MS with a different home location area, denoted by p_2, can be reduced significantly. The number of location update messages also depends on the number of replications. Assuming as before that the mobile station roams in three location areas, the number of outgoing location update messages depends on the number of replication as shown in Table 7.

The link utilization is given by

$$U = \frac{\lambda_{o/g}\times 30\times 8}{36\times 56\times 10^5}\,[5(1-\xi)^2 + 12(1-\xi)\xi + 7\xi^2] + \frac{\lambda_{in}\times 30\times 8}{36\times 10^5}\,[p_2(1-\xi)\xi + p_2\xi^2$$
$$\frac{\lambda_{ho}\times 30\times 8\times 8}{36\times 56\times 10^5} + \frac{(\lambda_{lu}+\delta)\times 30\times 8\times 3.33|2.0}{36\times 56\times 10^5} \tag{10}$$

4.5 Results and discussion

Table 8 lists the link utilization as a function of the call serving rate for different database placement strategies. The various parameters are shown below the table.

The velocity of the mobile station is based on the characteristics of a "student pedestrain" and is obtained from [kathy et al., 1992]. The size of the location area is assumed to be 2.25 square kilometers. This would correspond to a high density metro area. In this experiment we assume a switch with 70000 subscribers handling x (O+I) calls per hour per subscriber. The different calling rates were obtained by varying x. When the HLR and the VLR are co-located in the SCP, the link utilization is almost double that of the existing POTS network. As a result, the link utilization becomes greater than the required 40% limit for almost half the call handling rate that can be supported in the existing of penetration of mobile users. Let x denote the probability that the originating call is an intra-SP call. From the above one can easily establish the following four different call types and their respective probability of occurrence.

$$
\begin{aligned}
MS \rightarrow MS(Intra - SP) &= \xi x \\
MS \rightarrow Wireline(Intra - SP) &= (1 - \xi)x \\
MS \rightarrow wireline(Inter - SP) &= (1 - \xi)1 - x \\
MS \rightarrow MS(Inter - SP) &= (1 - x)\xi
\end{aligned} \tag{11}
$$

Note that based on the statistics of the various traffic mixes it can be easily established that x is equal to 0.3333.

Since the SCPs operate in a replicated mode, for the SCPs to be unavailable, both SCPs must be in the failed state. Based on a simple analysis it can be shown that the probability that at least one SCP is in the active state, i.e., availability of the SCPs, η_{scp} is given by

$$\eta_{scp} = 1 - \frac{\gamma^2}{\gamma^2+\gamma\beta+\beta^2} \tag{12}$$

If both the HLR and the VLR are co-located in the SCP, then the probability that a mobile station can initiate a call is equal to the probability that at least one SCP is in the active state. Thus the availability for this case is given by

$$\eta_{case1} = 1 - \frac{\gamma^2}{\gamma^2+\gamma\beta+\beta^2} \tag{13}$$

Next, consider Case 2 in which the HLR is in the SCP and the VLR is in the SP. We make the assumption that if the VLR has failed then call control and routing functions can be performed by the HLR database. Note that this will require extra messages to be exchanged which has not been modeled here. Based on this assumption it can be shown that the probability that an MS station can initate a call can be written as

$$
\begin{aligned}
\eta_{case2} = \xi(1 - \frac{\gamma^2}{\gamma^2+\gamma\beta+\beta^2}) + (1 - \xi) \\
[1 - (\frac{\gamma^2}{\gamma^2+\gamma\beta+\beta^2}\frac{\gamma}{\gamma+\beta})]
\end{aligned} \tag{14}
$$

The above equation is dervied by considering all the possible call types and noting that $\eta_{vlr} = \frac{\gamma}{\gamma+\beta}$ and η_{scp} is given by Equation(12).

Similarly, when the HLR and the VLR are cached in the SP, it can be shown that

$$
\begin{aligned}
\eta_{case3} = x[1 - (\frac{\gamma^2}{\gamma^2+\gamma\beta+\beta^2}\frac{\gamma}{\gamma+\beta})] \\
+(1 - x)(1 - \xi)[1 - (\frac{\gamma^2}{\gamma^2+\gamma\beta+\beta^2}\frac{\gamma}{\gamma+\beta})] \\
(1 - x)\xi(1 - \frac{\gamma^2}{\gamma^2+\gamma\beta+\beta^2})
\end{aligned} \tag{15}
$$

Table 8: Utilization as a function of the calling rate for different placement strategies. Parameters: $\xi = 0.25, N = 70000, A_{la} = 2.25, V = 5.17, L = 1.5, t_{hm} = 180, p_1 = 0.3, p_2 = 0.005, \delta = 1/hr$

Calling rate	70000	126000	182000	238000
POTS	0.083	0.150	0.217	0.283
HLR and VLR in SCP	0.166	0.281	0.396	0.511
HLR in SCP, VLR in SP	0.137	0.211	0.285	0.358
HLR cached in SP	0.155	0.230	0.305	0.380
HLR cached in 2 SPs	0.132	0.208	0.283	0.360
HLR cached in 3 SPs	0.118	0.194	0.270	0.346

POTS network. Moving the VLR into the SP results in a 25% to 30% reduction in the link utilization for the same calling rate. However, caching the HLR in the SP does not result in any improvement. In fact, there is a marginal increase in the A-link utilization. The main problem with this scheme is that the number of messages for outgoing calls increases without a compensating decrease in the messages for the intra-SP calls. Finally, replicating the HLR in one or more SPs results in a decrease in utilization for the same calling rate. Replication reduces the number of location update messages and also messages for intra-SP calls terminating at a mobile station. Note that in order to reduce the number of location update messages the replication must be correlated with the roaming characteristics of the mobile station.

The effect of different degrees of penetration of PCS usage is shown in Figure 5. From the figure it is clear that a higher degree of penetration will make distributed and/or replicated architectures a more attractive alternative. The results also show that a simple distributed scheme where the VLR is co-located in the SP and the HLR is maintained in the SCP can perform as well as more complex distributed and replicated architectures. Also note that the gains due to replication increase when the databases are replicated based on the roaming characteristics of the mobile station.

Finally, the effect of mobility is shown in Table 9 which shows the link utilization for different velocity in kilometers per hour (kph). The increase in velocity results in a higher number of crossings and hence more location updates and hence higher gains due to replication.

5 FAULT-TOLERANCE ANALYSIS

We consider a very simple model to study the fault-tolerance of the various database placement strategies. We make the following assumptions;

Table 9: A-link utilization as a function of the mobility for different placement strategies. Parameters: $\xi = 0.40, N = 70000, A_{la} = 9.00, \lambda_m = 3.0, L = 3.0, t_{hm} = 180, p_1 = 0.3, p_2 = 0.005, \delta = 1/hr$

Velocity (V kph)	5	11	17	23	29
POTS	0.25	0.25	0.25	0.25	0.25
HLR and VLR in SCP	0.55	0.58	0.61	0.64	0.68
HLR in SCP, VLR in SP	0.32	0.37	0.42	0.47	0.52
HLR cached in SP	0.34	0.40	0.46	0.53	0.59
HLR cached in 2 SPs	0.33	0.37	0.41	0.45	0.50
HLR cached in 3 SPs	0.32	0.35	0.38	0.41	0.44

- We consider only the failure of the databases. We assume that the inter-failure time and the recovery time are (negative) exponentially distributed random variables with mean rates γ and β, respectively.
- We assume that the two SCPs are operated in the replicated mode (as opposed to shared mode). This implies that both SCPs have the same information.

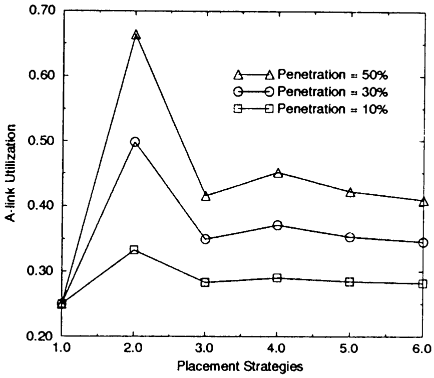

Figure 5: The link utilization for different placement strategies plotted for different degrees of penetration. X-axis : 1 → POTS, 2 → HLR and VLR in SCP, 3 → HLR in SCP and VLR in SPs, 4 → HLR cached in SPs, 5 → HLR cached in 2 SPs, 6 → HLR cached in 3 SPs. Parameters : $\xi = 0.40$, $N = 70000$, $A_{la} = 9.00$, $\lambda_m = 3.0$, $L = 3.0$, $t_{hm} = 180$, $p_1 = 0.3$, $p_2 = 0.005$, $\delta = 1/hr$.

We consider fault-tolerance only with respect to PCS. In view of this we define *availability* to be the probability that an MS can initiate a call. As before ξ denotes the degree

Table 10: Unavailability (1 - Availability) as a function of the failure rates for different placement strategies. Parameters : $\xi = 0.25, \beta = 1, x = 0.3333$.

γ	Unavailability		
	HLR and VLR in SCP	HLR in SCP, VLR in SP	HLR Cached in SP
5.0e-01	1.4e-01	0.72e-01	0.635e-01
5.0e-02	2.4e-03	0.7e-03	0.49e-03
5.0e-03	2.5e-05	0.63e-05	0.43e-05
5.0e-04	2.5e-07	0.6e-07	0.42e-07
5.0e-05	2.5e-09	0.6e-09	0.4e-09

Table 10 shows the availability of the three cases for different failure rates with $\beta = 1$ and $\xi = 0.25$. From the results it is clear that the availability of case 3 is the highest. This is not very surprising since replication in general results in higher availability. Based on the same analysis it can also be shown that if the HLR is replicated in the neighbouring SPs, then the availability can be further improved.

One of the key assumptions which leads to the above results is that the call control and routing functions can be performed by the HLR in case the VLR has failed. This may not be true for multiple reasons, e.g., the HLR and VLR may not be consistent with each other. If this assumption is not valid then it can be shown using a very similar analysis that availability when both the HLR and VLR is kept in the SCP is the highest. This suggest that in order to achieve high fault-tolerance it is necessary to maintain tight synchronization between the HLR and VLR databases. This implies higher performance penalties for higher degrees of replication.

6 CONCLUDING REMARKS

In this paper we have addressed performance and fault-tolerance of database management schemes to provide PCS in the existing signalling network. Specifically, we consider the placement of the VLR and th HLR databases in the existing signalling network. Four different schemes have been studied reflecting different degrees of distribution and replication. Based on a mean value analysis it has been shown that distributing the HLR and the VLR databases may be required for high degree of penetration of PCS usage. Simple distributed schemes where the VLR is co-located in the SP and the HLR is maintained in the SCP can significantly reduce the impact of PCS signalling on the existing signalling network. Performance of distributed architectures with replication depends largely on how the databases are replicated. The results show that performance gains are achievable when the replication is based on the roaming characteristics of the mobile station.

Fault tolerance of the different placement strategies was based on a simple failure and recovery model of the databases. The results show that the claim that the network can be made more robust against database failures with a distributed and replicated database placement strategy is valid only if all the distributed replicated copies are always consistent with each other. This may incur additional performance penalties.

There are a number of important future research directions. First, the mobility model assumed in this study is very simplistic and may not hold for different service areas. Realistic models are needed to quantify the number of location updates more accurately. Second, this study is based on a homogeneous signalling network. The database issues addressed in this paper should be analyzed for the case when the signalling network provides accesses to all different classes of PCS providers. Finally, this study is based on a mean value analysis or in other words, the "sunny day" scenario. It is important to study the performance and fault-tolerance of different database management schemes under failure and/or congestion conditions with temporal and spatial correlations.

7 REFERENCES

[Awerbuch et al., 1989] Baruch Awerbuch and David Peleg, "Online Tracking of Mobile Users", MIT Technical Report, October 1989.

[Bellcore, 1989] "Bell Communication Research Specification of Signalling System No 7", TR-NPL-000246, Issue 1,1985, reissued in June 1991.

[Cox, 1990] D. C. Cox, "Personal Communications – Viewpoint", IEEE Communications Magazine, November 1990.

[CCITT, 1988] "Public Land Mobile Network : Interworking with ISDN and PSTN", CCITT Recommendations Q.1000 – Q.1032, November 1988.

[Homa et al., 1992] Jonathan Homa and Steve Harris, "Intelligent Network Requirements for Personal Communication Services", IEEE Communications Magazine, February 1992, pp. 70–76.

[Jabbari, 1992] Bijan Jabbari, "Intelligent Network Concepts in Mobile Communications", IEEE Communications Magazine, February 1992, pp. 64–69.

[Kathy et al., 1992] K. S. Meier-Hellstern, E. Alonso and D. R. Neil, "The Use of SS7 and GSM to Support High Density Personal Communications", Winlab Workshop, April 1992.

[Kathy et al., 1991] K. S. Meier-Hellstern and E. Alonso, "Signalling System No 7 Messaging in GSM", Winlab Technical Report No. 25, December 1991.

[Lo et al., 1992] C. N. Lo, R. S. Wolff and R. C. Berhardt, "Expected Network Database Transaction Volume to Support Personal Communications Services", UPT Conference in Italy, January 1992.

[Minhas, 1987] H. Minhas, "GSM Signalling on the Radio Path : A Dimensioning Study", Proceedings International Conference on Digital Land Mobile Radio Communications, Venice 1987.

[Raymond, 1991] Paul-Andre Raymond, "Performance Analysis of Cellular Networks", IEEE Transactions on Communications, Vol. 39, No. 12, December 1991, pp. 1787–1793.

Skoog et al., 1990] A. R. Modarressi and R. A. Skoog, "Signalling System No. 7 : A Tutorial", IEEE Communications Magazine, July 1990, pp 19–35.

[Steel, 1990] R. Steel, "Deploying Personal Communications Networks", IEEE Communications Magazine, pp. 12–15, September 1990.

[Thomas et al., 1988] R. Thomas, H. Gilbert and G. Mazziotto, "Influence of the Movement of the Mobile Station on the Performance of a Radio Mobile Cellular Network", Proc. of 3rd Nordic Seminar, paper 9.4, Copenhagen, September 1988.

[Gray et al., 1993] Jim Gray and Andreas Reuter, "Transaction Processing : Concepts and Techniques", Morgan Kaufman Publishers, 1993.