Mobile Networks and Applications 2 (1997) 345–356

Traffic impacts of international roaming on mobile and personal communications with distributed data management

Jyhi-Kong Wey [a], Wei-Pang Yang [b] and Lir-Fang Sun [a]

[a] *Network Planning Laboratory, Telecommunication Laboratories, Chunghwa Telecom Co., Ltd., Ministry of Transportation and Communications, Taiwan, R.O.C.*

[b] *Institute of Computer and Information Science, National Chiao Tung University, Hsinchu, 300, Taiwan, R.O.C.*

In this paper, we propose four network interconnection scenarios and the related signaling aspects for the international roaming traffic in mobile and personal communications. With or without international gateway relay nodes summarized from the proposed scenarios, we also derive three international roaming network sets $\{\mathrm{IR}_1, \mathrm{IR}_2, \mathrm{IR}_3\}$ for the observed signaling traffic model with two-level databases. Based on the proposed performance metrics, analysis results show that under some reasonable assumptions the signaling traffic of international roaming is significantly impacted by the related gateway relay nodes and databases. In addition, the studied cache data management strategy at VLR can reduce the impacts of the signaling traffic of query operation, but the caching approach cannot improve the performance of update operation. Furthermore, the performance of proposed standalone STP scenario is better than that of integrated STP scenario for the international network connection.

1. Introduction

Roaming is one of the attractive features in mobile and personal communications networks (PCN). To provide this feature, the mobile and personal communications systems must keep track of information about the present location, the subscription information, and the service information associated to the subscriber. Roaming is not restricted within the whole coverage of a mobile and personal communications network, but it embraces all the interconnected communications networks across countries' border. This roaming feature is referred to international roaming.

As roaming is to mobile subscriber, so is mobility management to network. Roaming is supported by the mobility management in the network [9]. Roaming and mobility management are two folds of one thing. Signaling protocol system, Mobile Switching Centers (MSC), and distributed databases, such as Visitor Location Register (VLR) and Home Location Register (HLR) [14,16,25], are three major elements in the network to support the mobility management. The signaling protocol provides a reliable transportation vehicle for mobility management messages initiated from mobile switching centers and distributed databases in mobile and personal communications networks. The signaling protocol between nodes of the network or the international network is based on Signaling System 7 (SS7) [19,21,24].

The signaling traffic generated by various network architectures in mobile and personal communications has been discussed in [15,18,25–27]. We extend this signaling traffic study to an international roaming situation. Previous studies [18,20,22] also show that data management strategies have significantly traffic impacts on the signaling network and network databases. Regardless of the performance bottleneck at HLR database update or query load, or the signaling link performance, caching data in the down level of databases is considered as an efficient data management strategy in the two-level hierarchical strategy with two types of database (i.e., HLR and VLR) [18]. However, the interconnected network architectures for international mobile and personal communications are more complicate than the national wide architectures. Not only there are network coupling relations, but also the signaling traffic generation and relay are for the mobility-related signaling, such as call origination and call delivery. Therefore, the traffic impact of international roaming with distributed data management is a new issue in the teletraffic management of mobile and personal communications.

The outline of this paper is as follows. In section 2, we propose the international signaling network connection scenarios and related signaling aspects of roaming for the presentation and analysis of performance. In section 3, we describe the concerned cache data management strategy. We also present the international roaming database, and give the examples of the mobile originated and mobile terminated call to demonstrate the signaling traffic generated by routing and addressing. We then use this traffic information to quantify our analysis model. In section 4, a signaling traffic analysis model is formulated from international roaming scenarios with two-level hierarchical databases. Furthermore, with or without international gateway relay nodes some signaling traffic loads (e.g., databases′ waiting time, query and update response time, and total number of messages) are derived for analysis and comparison. Finally, section 5 provides conclusions.

Table 1
SPC coding.

Standard	ITU-T			ANSI		
Meaning	Zone	Area	SPC identification	Network	Cluster	Member number
Bits assignment	3 bits	8 bits	3 bits	8 bits	8 bits	8 bits
Example	4	150	6	255	255	1

2. International roaming network

Since SS7 is a CCS (Common Channel Signaling) [19] system which provides control and management in the PCN network, roaming in the PCN is based on the signaling messages transferred in the signaling link between signaling points which can be an MSC, VLR, and HLR in terms of the existing mobile network, e.g., Global System for Mobile communications (GSM). We assume MSC and VLR are functioning integrated; the terms MSC, VLR, and MSC/VLR are interchanged in this paper. In addition, the terms PCN, and mobile and personal communications networks are also used here interchangeably.

When two signaling points are capable of exchanging signaling messages between themselves through SS7 network, a signaling relation is defined between them. Roaming in the international PCN network is also based on the same signaling relation defined in the national PCN network. However, the SS7 signaling network can be classified into two functionally independent levels – the national levels and the international levels – for the network management and numbering plans in each network. Therefore, a signaling point can be a national signaling point, an international signaling point, or both. A signaling point is addressed by a signaling point code (*SPC*). An *SPC* can be 14 bits or 24 bits in the global SS7 signaling network. The contents and bits' assignment for SPC are shown in table 1. For the international standard of ITU-T, the 14 bits *SPC* is adopted and denoted by international *SPC* (*ISPC*) [4], e.g., $\{ISPC = 4\text{-}150\text{-}6\}$ is an MSC of the existing Taiwan's GSM network. If a signaling point serves as a national and international signaling point, it is called dual point code's signaling point.

2.1. Signaling aspects for roaming

For the signaling messages transported for roaming control, an SS7 signaling protocol stack is defined in the existing mobile communications [14,16,19,21]. The protocol stack is given in figure 1.

The Message Transfer Part (MTP) provides a reliable transfer and delivery of signaling information across the international signaling network. MTP consists of three levels of SS7 signaling protocol. They are referred to signaling data link, signaling link control, and signaling network function. With these three levels, connectionless signaling information will be transferred across the network to its desired destination. Therefore, a signaling point with MTP capability is called Signaling Transfer Point (STP).

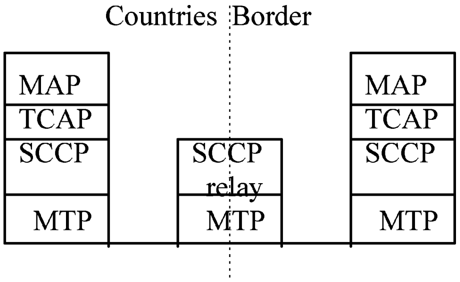

Figure 1. Signaling protocol stack for international roaming.

The addressing capability of MTP is limited to delivering a message to an adjacent node. The Signaling Connection Control Part (SCCP) extends the routing capability by providing an addressing capability in a global (international) signaling network. The translation function performed by the SCCP is to translate the SCCP address parameter from a global title to a point code and a subsystem number (*SSN*). The *SSN* is a local addressing information used by the SCCP to identify each of the SCCP users at a signaling node, e.g., MAP (Mobile Application Part), HLR, VLR, and MSC have *SSN* for the SCCP. The global title is an address, such as dialed digits in 800 service or Mobile Station (MS) ISDN number (*MSISDN*), which contains an implicit address information not routable by the MTP. For global titles, a translation capability called Global Title Translation (GTT) function [4] is required in the SCCP to translate the global title to a *DPC* (Destination Point Code) and an *SSN*. This GTT function can be performed at the originating signaling point of the message, or at an SCCP relay node.

The GTT function supports several numbering plans in the existing mobile communications [4,7], such as E.164 for ISDN/telephony numbering plan, E.212 for land mobile numbering plan, and E.214 for ISDN/mobile numbering plan. The detailed addressing and routing procedures using the GTT function and numbering plans will be described in the following sections.

The Transaction Capabilities Application Part (TCAP) is structured in the two sublayers, the component sublayer and the transaction sublayer, for dealing with the noncircuit-related signaling message. Essentially, TCAP provides a set of capabilities in order to exchange application in mobile and personal communications databases. The Mobile Application Part [9] is the specific protocol layer defined for the mobile application information, such as call registration and call handover signaling messages, exchanged between MSC, VLR, and HLR.

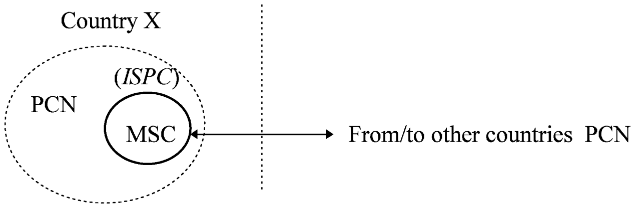

Figure 2. Scenario A: MSC with *ISPC*.

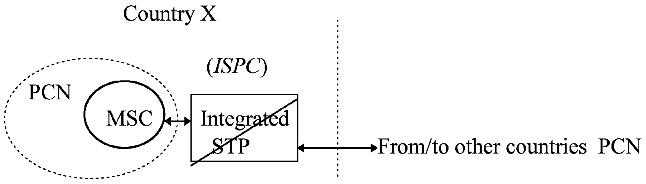

Figure 3. Scenario B: integrated STP with *ISPC*.

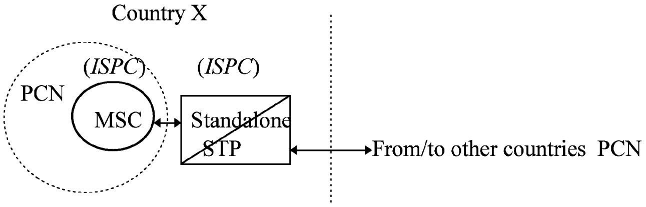

Figure 4. Scenario C: standalone STP with *ISPC*.

2.2. Signaling network interconnection

The signaling networks of PCN consists of signaling points and signaling links connecting the signaling points together. According to the definition of signaling networks and signaling modes, there are a number of scenarios on how signaling network interconnection to form international roaming networks. We propose some typical scenarios in the following.

Scenario A) MSC *with ISPC*
This scenario is shown in figure 2. The MSC with an *ISPC* has SCCP capabilities, and is directly connected to the international SS7 network. In this scenario, this MSC functions as a remote MSC of the interconnected countries' PCN. This MSC is a signaling point with dual point code. Therefore, a GTT function should be performed when mobility-related signaling messages originated or transported from this signaling point.

Scenario B) Integrated STP *with ISPC*
When an STP has SCCP function, it is referred to an integrated STP here. This scenario is shown in figure 3, the MSC interconnects an integrated STP which has *ISPC*. In this scenario, this integrated STP is a directly addressable signaling point by the interconnected countries' PCN. The GTT function should be performed at the integrated STP regardless of the GTT function performed at MSC. This integrated STP function is referred to an SCCP relay node in figure 1.

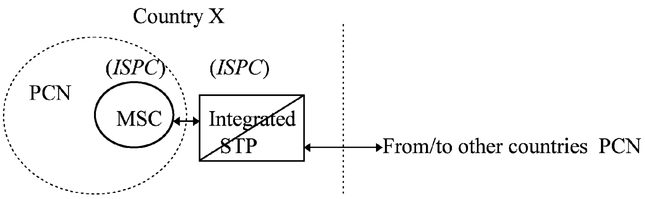

Figure 5. Scenario D: MSC and integrated STP, both with *ISPC*.

Scenario C) MSC *and standalone* STP *both with ISPC*
When an STP plays a role of *ISPC*, it is referred to a standalone STP here. This scenario is shown in figure 4, both the MSC and the standalone STP have *ISPC*. In this scenario, this MSC is a directly addressable signaling point by the interconnected countries' PCN. The GTT function should be performed at MSC. Furthermore, a standalone STP only plays the transparent signaling point, i.e., there is no GTT function performed in this node.

Scenario D) MSC *and integrated* STP *both with ISPC*
This scenario is shown in figure 5; the MSC interconnects an integrated STP. Both signaling points have *ISPC*. In this scenario, both the MSC and the integrated STP can be directly addressed by the interconnected countries' PCN. In addition, the GTT function can be performed at the MSC or the integrated STP depending on the engineering implementation.

Let IR $= \{(x, y)\}$ denote the international roaming network set where x is a scenario adopted by a country X's PCN, and y is a scenario adopted by a country Y's PCN. We have $x, y \in$ {Scenario A, Scenario B, Scenario C, Scenario D}. Therefore, there are totally 16 combinations of the international signaling point connection. For example, a (Scenario A, Scenario B) represents the network interconnection of figure 2 and figure 3.

The Scenario A is referred to without international gateway relay nodes, and the Scenarios B, C and D are referred to with international gateway relay nodes for our later discussion.

3. Distributed databases

3.1. Efficient data management

A customer in the mobile and personal communications network requires the information stored in the databases for call origination and call delivery to a destination. Data management is essential in mobile and personal communications. Location tracking and registration, call delivery, authentication and verification, and service profile management are all data management issues. Many researchers [18,22,25,28] have devoted to studying the efficient database architecture for tracking customer locations and for improving the information retrieval efficiency in these architectures. An efficient database architecture significantly increases the mobility and call handling efficiency. Two kinds of data management in the mobile and personal communications databases are (i) call handling data management, and (ii) mobility management. The data management of call handling is relative to the information retrieval activities for call delivery and authentication. In addition, the data management of mobility management is relative to the location update and registration, and service profile management activities.

Distributed hierarchical database architecture is assumed to be efficient [1,14,23] in the data management of the mobile and personal communications. Two-level distributed hierarchical databases, VLRs and HLR, are adopted in GSM and DAMPS [6,25]. This distributed database approach with replicated data caching [18] at VLRs is proved efficient for automatic roaming and handover calls. The caching data approach is generally applied to computer systems [2]. Now, many researchers [18,22] have proved its feasibility and efficiency in personal communications.

Data replication is the basic method for data management in distributed databases. Replication can mean better performance and better availability for information retrieval in personal communications. However, the update propagation problem is introduced. A common scheme for dealing with this problem is referred to primary copy scheme, and is adopted in the existing mobile and personal communications, such as GSM, PACS, and PHS. One copy of each replicated data is designated as the primary copy stored at HLR. However, not all data are replicated at VLRs. Briefly, in personal communications VLR caches a set of partial replicated data, and the primary copy of data are stored in the centralized database HLR. The update operations are deemed to be complete as soon as the primary copy at HLR has been updated. Therefore, how large size of caching data at VLR directly resulting in the performance of information retrieval and update operation is the efficient data management issue in this study.

3.2. International roaming databases

International roaming databases involve the roaming-related data set in distributed databases. The following specific information is recommended in the international roaming databases [10]:

(i) the E.164 Country Code (*CC*) and National Destination Code (*NDC*) of the *MSISDN* number,

(ii) the E.212 Mobile Country Code (*MCC*) and Mobile Network Code (*MNC*) of the International Mobile Station Identity (*IMSI*) number,

(iii) the E.214 Country Code and Network Code (*NC*) of the Mobile Global Title (*MGT*) number,

(iv) the Mobile Station Roaming Number (*MSRN*),

(v) the *ISPC* used by two standalone/integrated STP nodes which the PCN is connected to, and

(vi) the type of exchange used as the two international gateway nodes which the PCN is connected to.

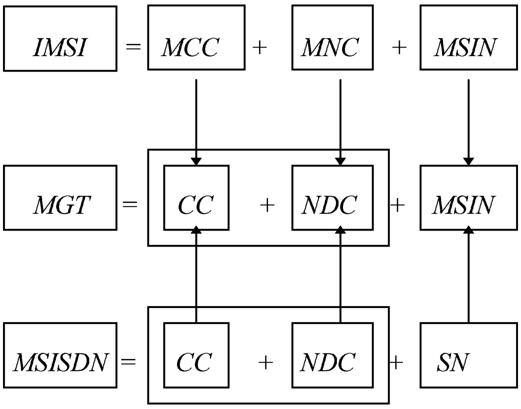

Figure 6. Mobile global title translated from *IMSI* and *MSISDN*.

The *MSISDN*, *MSRN*, *IMSI*, and *MGT* are the four most important numbers for international roaming. In principle, *MSISDN* should be possible for any subscriber of the ISDN or PSTN to call any Mobile Station (MS) in PCN. This presents $\{MSISDN = CC +$ national mobile number $= CC + NDC + SN$ (Serial Number)$\}$. Alternatively, an *MSRN* is temporally assigned to an MS by the VLR with which the MS is registered when a request from Gateway MSC (GMSC) via HLR. The GMSC uses this *MSRN* to route calls directed to an MS. The format of *MSRN* is identical to the *MSISDN*.

In addition, a unique *IMSI* shall be allocated to each mobile subscriber. The *IMSI* is composed of *MCC*, *MNC*, and *MSIN* (Mobile Subscriber Identification Number). The *MCC* uniquely identifies the country of domicile of the mobile subscribers. An *MNC* identifies the home PCN of the mobile subscribers, and the *MSIN* identifies the mobile subscribers within a PCN.

The *MSISDN* and *IMSI* numbers can be used as a global title address in the SCCP for signaling routing to the HLR of MSs. The translation scheme is given in figure 6. The $\{CC + NDC\}$ provides sufficient routing information [7] to address the exact *DPC* and *SSN* such as the database HLR. For example, the SCCP of MSC translates the *IMSI* to the *MGT*. Next, the SCCP of integrated STP translates the *MGT* to the international *SPC* of *DPC*.

3.3. Routing and addressing

Intra PCN roaming is easily found in previous studies [14,21]. The intra PCN call routing between the originating MSC and the terminating MSC is similar to the PSTN calls. Meanwhile, the routing information needed for a call is the technical issue of the database point code addressing (e.g., addressing VLR or HLR). The communications between VLR and HLR are to translate the Mobile GTT necessary to determine the *DPC* of each other [4].

However, routing and addressing between VLR and HLR for the international roaming case a GTT is necessary for each signaling message at the originating node and at all SS7 network border's nodes [11]. For demonstrating the routing and addressing scheme in the international roaming,

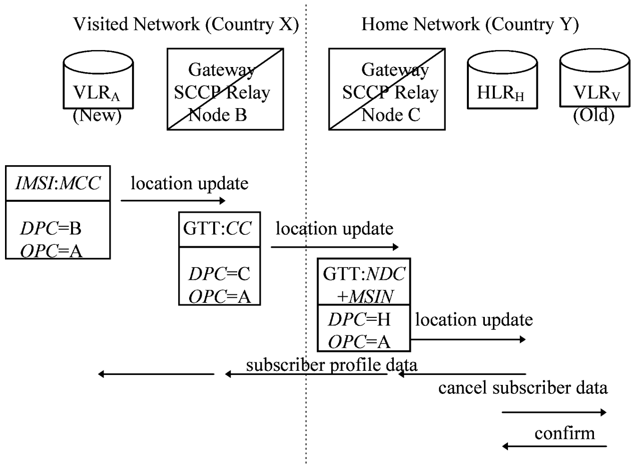

Figure 7. Registration and location update with mobile GTT function.

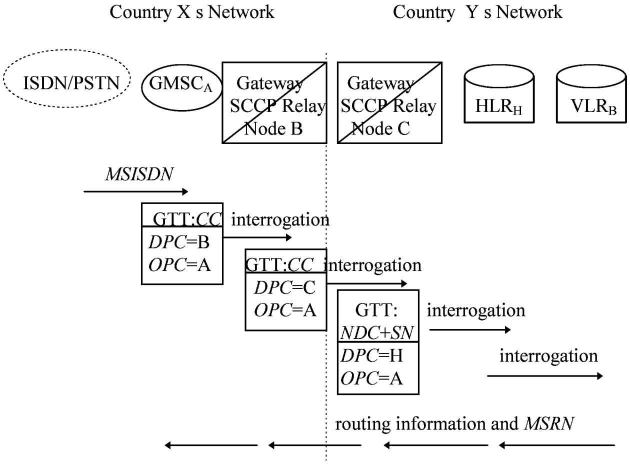

Figure 8. Mobile terminated call with mobile GTT function.

the examples of the mobile originated call and mobile terminated call are described as follows, respectively.

Example 3.1 (Mobile originated call). In this example, a mobile subscriber is assumed to roam in the foreign PCN, and IR $= \{$(Scenario B, Scenario B)$\}$ described in section 2.2 is adopted here. When the MS of this mobile subscriber is first activated, it initiates the location registration to the VLR in the visited PCN. This registration message is transferred to the VLR by a location update message [9]. Upon receiving this registration message, the VLR checks whether this MS is already registered or not. If this MS is not registered, VLR checks the *IMSI* of this registration message to determine the exact home HLR to which the MS is belonged. Next, the visited VLR communicates with the home HLR using the *MGT* derived from the *IMSI*. Consequently, the *MGT* is analyzed by the SCCP routing function. The location update message is determined to be transmitted to a node in a foreign country by the *CC* of *MGT*. Therefore, the *SPC* of the gateway SCCP relay node takes charge of the translation from the originating node to the subsequent SS7 network is determined. This GTT function is performed repeatedly in all intermediate SCCP relay nodes until the home HLR can be addressed by the $\{NDC+MSIN\}$. The GTT function performed by the SCCP is to translate the SCCP address parameter from a global title to a point code and a subsystem number. Upon receiving this location update message, the HLR, in turn, sends subscriber profile data on the basis of the received visited VLR number by the MTP transportation function. Next, the home HLR cancels the subscriber profile data cached in the old VLR.

The subsequent mobile originated call will query the visited VLR database and the home HLR by means of the same GTT function described in the registration scenario above. A simplified procedure of the registration and location update with GTT function is shown in figure 7. We use rectangle box to demonstrate the GTT function performed at each node. The upper portion of rectangle box shows the key parameter to be translated by the GTT. The lower portion of rectangle box shows the *DPC* and *OPC* (Origination Point Code) determined by the GTT.

Example 3.2 (Mobile terminated call). An ISDN/PSTN subscriber dials the *MSISDN* of called subscriber. The exchange of ISDN/PSTN shall route this call to a GMSC that is used to perform interrogation of HLR for this mobile subscriber. The GMSC determines an appropriate HLR to interrogate the present location of the called subscriber by performing GTT on an *MSISDN*. In turn, HLR interrogates the visited VLR for a routable address and an *MSRN*. Next, HLR responds the GMSC with the routing information and *MSRN*. The GTT function performed in the above procedure is shown in figure 8. Consequently, this call is routed from the GMSC to the exact destination of MSC in the visited foreign country.

4. Traffic impact analysis

In this section, a signaling traffic performance is derived from the international roaming scenarios with two-level hierarchical databases, such as VLR and HLR. This study derives the waiting time of VLR and HLR first. Next, database query response time, database update response time, and transaction messages between each VLR and HLR are observed with or without international gateway relay nodes.

4.1. Analytical model

According to the examples of the mobile originated call and mobile terminated call, HLR functions as a centralized database in the international roaming network, and VLRs are the localized databases distributed in remote sites. HLR and VLRs are assumed first-come, first-served basis (FIFO) as well as all databases with $M/G/1$ queueing service model. Also, this study assumes that each VLR caches the replicated data of HLR with probability p. Therefore, the query at HLR with probability $(1 - p)$ when mobile originated call initiates.

The handover number of each call is also a very important parameter in the databases' traffic analysis. For observing the impact of handover, a varying probability q is taken in this study. By the way, in order to simulate various kinds of traffic patterns the parameters p and q are used simultaneously to compose the predefined cache size probability at VLR and subscriber moving probability.

In terms of database activity, the signaling traffic described in [9] and figures 7 and 8 can be summarized as two kinds of mobility-related signaling traffic – query and update. The mobility-related signaling consists of registration, deregistration, call origination, and call delivery. With $\mathrm{M/G/1}$ queueing model [17] for HLR and VLR, all query and update arrivals form Poisson process with average arrival rate, such as $\lambda_{\mathrm{VLR}}^{\mathrm{Query}}$ and $\lambda_{\mathrm{VLR}}^{\mathrm{Update}}$. The HLR and VLR service times are assumed with the same gamma distribution density function [12]. Let gamma variable T, we have

$$f_T(t) = \frac{t^{\tau-1}}{\Gamma(\tau)\xi^{\tau}} \mathrm{e}^{-t/\xi}, \quad t \geqslant 0, \tag{1}$$

where

$$\Gamma(\tau) = \int_0^{\infty} \mathrm{e}^{-x} x^{\tau-1}\,\mathrm{d}x, \quad \tau > 0.$$

Then, the mean and variance of service time T are $E[T] = \mu_T = \tau\xi$ and $E\{T - E[T]\}^2 = \sigma_T^2 = \tau\xi^2$. Furthermore, the waiting time W with mean arrival rate λ at a database HLR or VLR can be derived from the Pollaczek–Khinchin formula [17,22] and (1),

$$W = \frac{\lambda(\sigma_T^2 + \mu_T^2)}{2(1 - \lambda\mu_T)} = \frac{\lambda(\tau+1)\tau\xi^2}{2(1 - \lambda\tau\xi)}. \tag{2}$$

The previous description in this section indicates the traffic projection for our later analysis. This traffic projection is easily understood since it is observed in the long period from the existing mobile network [29]. However, the performance requirements of the database system for the mobile and personal communications are controversy. With the stringent real-time response requirements for database activities, a real-time performance of the distributed databases is critical. Generally, the real-time performance in these distributed databases is highly dependent on the database system architecture, data model, and information retrieval algorithm. Most of databases are designed as mated pairs with duplicate architecture [28] for new telecommunication services. Under the specific traffic projection, [30] predicts that

$$\Pr(\text{database response time for queries } \leqslant 150 \text{ ms}) > 0.98. \tag{3}$$

Actually, the predicted database response time in (3) should be the sum of database service time and waiting time in our study. By using a generalization theorem of Chebyshev's inequality [13], the mean service time in (3) can be roughly estimated as 3 ms.

Another reference value can be found in [28] who claims that the SCP capacity for 800 service is 110 calls per second. Furthermore, the GSM performance objectives for databases under the specific reference load [8] are very close to the value described above. Therefore, we adopt a reasonable nominal value for database service time requirements in our study.

To observe the impact of the international roaming signaling traffic between HLR and VLRs, we make four kinds of measurement in the study. They are:

(i) the waiting time of signaling message in HLR and VLR for various tails of gamma distribution (e.g., $\tau = 1.0$ for exponential distribution and $\tau = 3.0$ for uniform distribution [22]),

(ii) the average query response time of signaling message between VLR and HLR,

(iii) the average update response time of signaling message between VLR and HLR, and

(iv) the expected numbers of signaling message per second exchanged between VLR and HLR [2].

4.2. Without international gateway relay nodes

Assume we have no international gateway relay nodes that are corresponding to the international roaming network set $\mathrm{IR}_1 = \{(x, y) \mid (x, y) = (\text{Scenario A, Scenario A})\}$ in section 2.2; in this case databases are the potential bottlenecks here. At a centralized database HLR, there are three types of request. Two types of request are query operations from remote databases and from the GMSC (i.e., PSTN call delivery), respectively. One type of request is update operation for each VLR. Since the combined flow of requests also forms a Poisson process with the sum of each arrival rate in $\mathrm{M/G/1}$ model. From table 2, the service time for HLR is computed as follows:

$$\mu_{\mathrm{HLR}} = \left[\frac{\lambda_{\mathrm{HLR}}^{\mathrm{QueryfPSTN}} + \lambda_{\mathrm{HLR}}^{\mathrm{Queryfmobile}}}{\sum \lambda_{\mathrm{HLR}}}\right] T_{\mathrm{HLR}}^{\mathrm{Query}} + \left[\frac{\lambda_{\mathrm{HLR}}^{\mathrm{Update}}}{\sum \lambda_{\mathrm{HLR}}}\right] T_{\mathrm{HLR}}^{\mathrm{Update}}, \tag{4}$$

where $\sum \lambda_{\mathrm{HLR}}$ is the sum of various arrival rates at HLR. Similarly, there are four types of request at each localized database VLR. Two types of request are query operations from the local MS and the HLR, respectively. Another two types of request are update operations from the MS registration, call origination, and deregistration. Therefore, the mean service time for each VLR is computed from the following expression:

$$\mu_{\mathrm{VLR}} = \left[\frac{\lambda_{\mathrm{VLR}}^{\mathrm{Query}} + \lambda_{\mathrm{VLR}}^{\mathrm{QueryfPSTN}}}{\sum \lambda_{\mathrm{VLR}}}\right] T_{\mathrm{VLR}}^{\mathrm{Query}} + \left[\frac{\lambda_{\mathrm{VLR}}^{\mathrm{Update}}}{\sum \lambda_{\mathrm{VLR}}}\right] T_{\mathrm{VLR}}^{\mathrm{Update}} + \left[\frac{\lambda_{\mathrm{VLR}}^{\mathrm{Dereg}}}{\sum \lambda_{\mathrm{VLR}}}\right] T_{\mathrm{VLR}}^{\mathrm{Dereg}}, \tag{5}$$

Table 2
Parameters in the analysis.

Parameter	Description	Proposed value
N	number of VLR	8
p	probability of subscriber data cached at VLR	range from 0 to $1/N$
q	handover probability	range from 0 to 0.5
$\lambda_{\mathrm{VLR}}^{\mathrm{Query}}$	query arrival rate at VLR	0.075 queries per ms
$\lambda_{\mathrm{VLR}}^{\mathrm{QueryfPSTN}}$	query arrival rate from PSTN at VLR	$0.55(\lambda_{\mathrm{VLR}}^{\mathrm{Query}}/0.45)$
$\lambda_{\mathrm{VLR}}^{\mathrm{Update}}$	update arrival rate at VLR	0.075 updates per ms
$\lambda_{\mathrm{VLR}}^{\mathrm{Dereg}}$	deregistration update rate at VLR	$\sum q\lambda_{\mathrm{VLR}}^{\mathrm{Update}}$
$T_{\mathrm{VLR}}^{\mathrm{Query}}$	query processing time at VLR	10 ms
$T_{\mathrm{VLR}}^{\mathrm{Update}}$	update processing time at VLR	20 ms
$T_{\mathrm{VLR}}^{\mathrm{Dereg}}$	deregistration processing time at VLR	$T_{\mathrm{VLR}}^{\mathrm{Update}}$
$\lambda_{\mathrm{HLR}}^{\mathrm{QueryfPSTN}}$	query arrival rate from PSTN at HLR	$\lambda_{\mathrm{VLR}}^{\mathrm{QueryfPSTN}}$
$\lambda_{\mathrm{HLR}}^{\mathrm{Queryfmobile}}$	mobile originating query arrival rate at HLR	$\sum(1-p)\lambda_{\mathrm{VLR}}^{\mathrm{Query}}$
$\lambda_{\mathrm{HLR}}^{\mathrm{Update}}$	update arrival rate at HLR	$\sum q\lambda_{\mathrm{VLR}}^{\mathrm{Update}}$
$T_{\mathrm{HLR}}^{\mathrm{Query}}$	query processing time at HLR	10 ms
$T_{\mathrm{HLR}}^{\mathrm{Update}}$	update processing time at HLR	30 ms

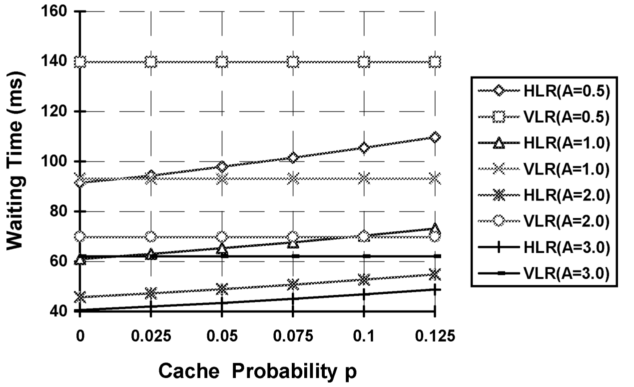

Figure 9. Waiting times at HLR and VLR with $\lambda = 0.05$ and $q = 0.5$. (A denotes τ.)

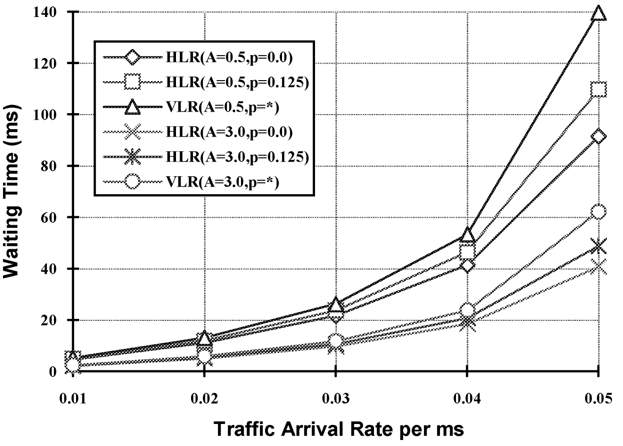

Figure 10. Waiting times at HLR and VLR with $q = 0.5$. (A denotes τ.)

where $\sum \lambda_{\mathrm{VLR}}$ is the sum of various arrival rates at VLR. Consequently, waiting times at the HLR and VLR are easily derived from (4), (5) and (2).

Assume the mobile subscriber penetration rate 10 per cent in the twenty millions' population areas. Let the traffic per subscriber be 0.03 Erlangs in the busy hour, and the average holding time 100 seconds with the 45 per cent of total calls originated from mobiles and 55 per cent of total calls terminated to mobiles [29]. Consequently, the traffic 0.075 calls/ms for 250,000 subscribers are generated from each of eight VLRs [29]. The SS7 network transmission time is assumed to be a constant $T_{\mathrm{Trans}} = 8.75$ ms where an average 70 bytes of signaling message transmitted with 64 kbps speed is taken. In addition, the signaling data link propagation time $T_{\mathrm{P}} = 5$ ms is assumed here. As described in section 2, the PCN network architectures for analysis are divided into without or with international gateway relay nodes. The analysis parameters in this study are summarized in table 2.

Figures 9–11 show the analysis results of waiting times at HLR and VLR under various traffic parameters. Figure 9 demonstrates waiting times at HLR and VLR for various τ values and the cache probabilities p with fixed traffic arrival rate $\lambda = 0.05$ and handover probability $q = 0.5$. The results show that waiting times at VLR are invariant with the cache probability p, but waiting times at HLR linearly slow increase with p. Actually, caching information at VLR improves the performance of HLR. Also, waiting times at VLR are more sensitive with small *shape* parameter τ of gamma distribution. This is because under the same mean value the variance of gamma variable with small τ is larger than the variance of gamma variable with large τ.

High traffic arrival rate under high handover probability $q = 0.5$ also significantly affects the waiting times at VLR with the independence of cache probability p shown in figure 10. Also, in any case of traffic arrival rate caching data at VLR cannot completely reduce the waiting time

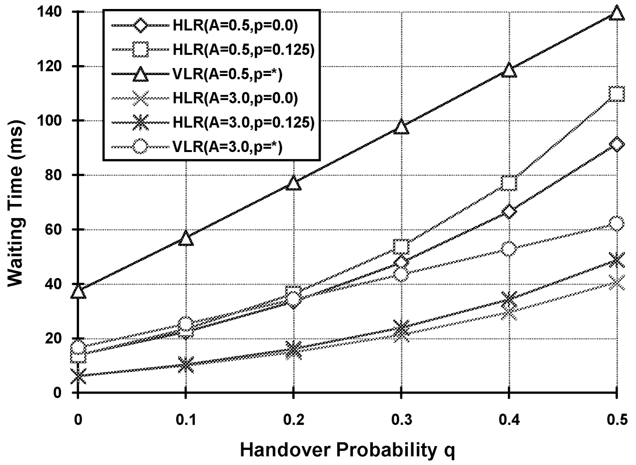

Figure 11. Waiting times at HLR and VLR with $\lambda = 0.05$. (A denotes τ.)

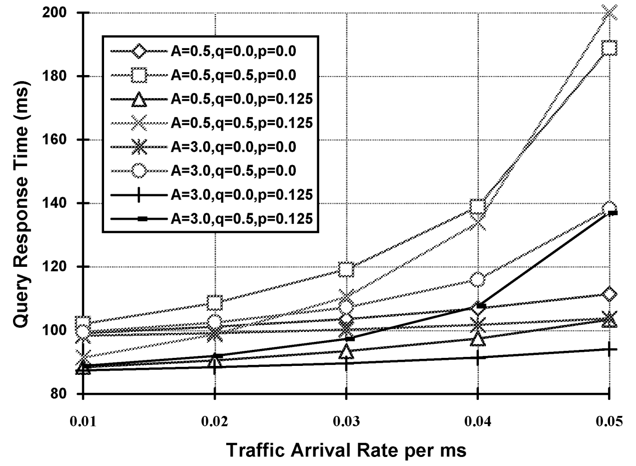

Figure 12. Query response times with various traffic arrival rates. (A denotes τ.)

performance at HLR. Figure 11 shows that high handover probability q also contributes to the waiting times at HLR and VLR, but the increasing slopes of waiting times are approximately to linear under traffic arrival rate $\lambda = 0.05$ condition. The reason for this phenomena is resulted from the update operation at VLR and HLR all with linear coefficient q.

The international roaming traffic impacts on the VLR and the HLR are the observed key points in this study. Of course, the query and update response times are deserved to evaluate the traffic impacts on the international networking. Based on the databases activities and mobility-related signaling traffic shown in figures 7 and 8, and table 2, the query response time of distributed databases is easily derived as the weighted sum of query in the VLR and HLR, respectively:

$$Q_{\text{rsp}} = p\left[W_{\text{VLR}} + T_{\text{VLR}}^{\text{Query}}\right] + (1-p)\left[2(T_{\text{Trans}} + T_{\text{P}}) + W_{\text{HLR}} + T_{\text{HLR}}^{\text{Query}}\right], \quad (6)$$

where W_{VLR} and W_{HLR} are waiting times at VLR and HLR.

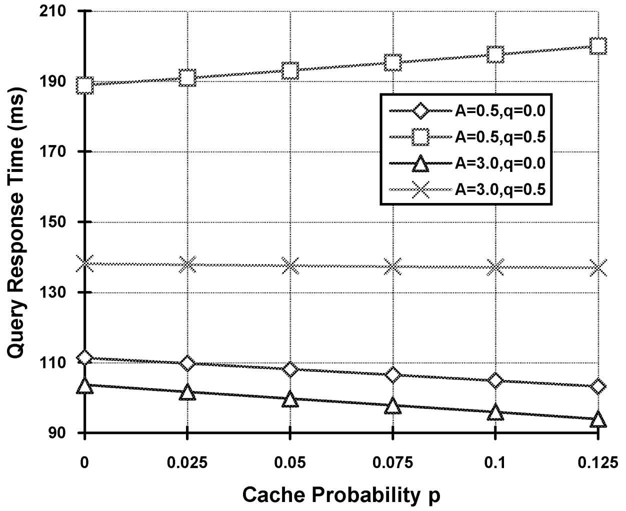

Figure 13. Query response times with $\lambda = 0.05$ and various cache probabilities. (A denotes τ.)

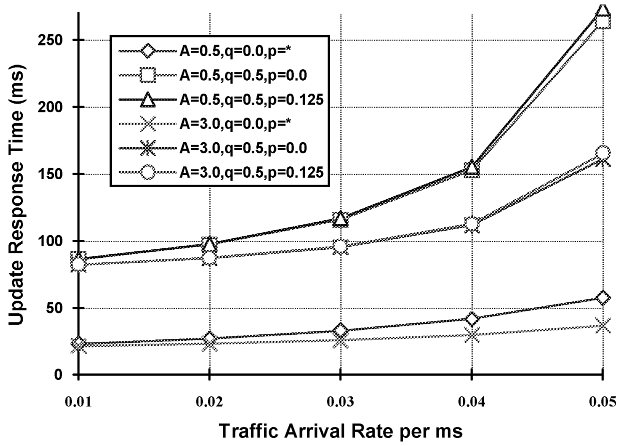

Figure 14. Update response times with various traffic arrival rates. (A denotes τ.)

Analysis results shown in figures 12 and 13 obviously confirm that the cache probability p and the handover probability q simultaneously dominate the performance of query response time with high traffic arrival rate. Furthermore, the query response time is more sensitive with small *shape* parameter τ of gamma distribution. Figure 13 reveals the cache effect on the query response time. Even though the handover probability is increasing dramatically, the performance of query response time increases slowly by means of caching data at VLR. Again, we can find that the linear coefficient p on (6) results in the approximate linear effect on the query response time of figure 13. Similarly, the update response time expression of distributed databases is the sum of updates in the VLR and the updates in the HLR multiplied by the handover probability

$$U_{\text{rsp}} = \left[W_{\text{VLR}} + T_{\text{VLR}}^{\text{Update}}\right] + q\left[2(T_{\text{Trans}} + T_{\text{P}}) + W_{\text{HLR}} + T_{\text{HLR}}^{\text{Update}}\right]. \quad (7)$$

Analysis results confirm that the impacts of cache probability p and handover probability q on the update response

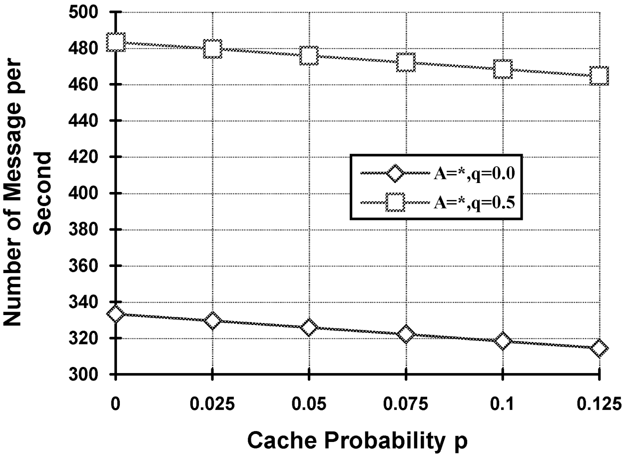

Figure 15. Total numbers of message exchanged between each VLR and HLR per second with $\lambda = 0.05$. (A denotes τ.)

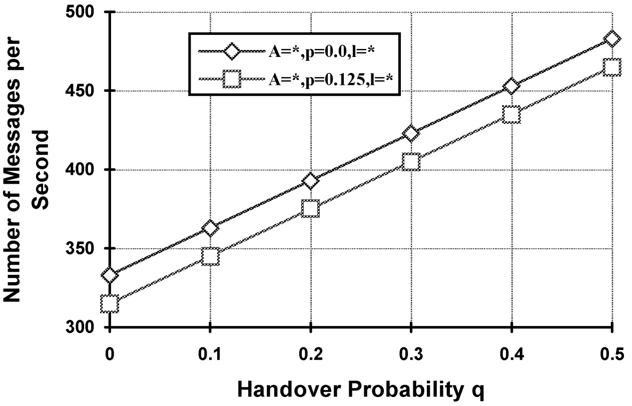

Figure 16. Total numbers of message exchanged between each VLR and HLR per second. (A denotes τ, 1 denotes λ.)

time are similar to those on the query response time. Therefore, we only demonstrate the update response time with various traffic arrival rates in figure 14. One of the important results indicates that the size of cache probability improves the update response time very little. Therefore, the lines shown in figure 14 almost exist by pairs. Another result is that the update response time exist almost double size compared with the query response time under each traffic arrival rate.

Consequently, the expected numbers of signaling message per second exchanged between each localized database VLR and centralized database HLR is the sum of all signaling messages of query and update between VLR and HLR:

$$M_{M\text{sgn}o} = 2(1000)\left[\lambda_{\text{VLR}}^{\text{QueryfPSTN}} + (1-p)\lambda_{\text{VLR}}^{\text{Query}} + 2q\lambda_{\text{VLR}}^{\text{Update}}\right]. \tag{8}$$

Figures 15 and 16 show the analysis results under various handover probability sizes and cache probability size. The results obviously indicate that:

(i) the larger size of cache probability p becomes the smaller volume of message exchanged between each VLR and HLR with linearly monotonic slow decreasing relation,

(ii) the larger size of handover probability q becomes the larger volume of message exchanged between each VLR and HLR with linearly monotonic fast increasing relation,

(iii) the size of cache probability positively reduces the total numbers of message required for the international roaming traffic,

(iv) the handover probability size contributes to the total numbers of message required for the international roaming traffic definitely and significantly, and

(v) the pace differences of two lines in these two figures are mainly introduced from the size of handover probability $q = 0.0$ or 0.5, respectively.

4.3. *With international gateway queueing delay*

The standalone STP has no SCCP relay function. The Scenario C, and the combination of Scenarios A and C also belong to this case. The message delay through a standalone STP is specified as the cross-STP delay. The cross-STP delay is composed of the mean processor handling time and the mean signaling link delay. The mean signaling link delay is dependent on the message length. We investigate the mean processor handling time and the mean signaling link delay in the gateway relay nodes.

The service time (processing time), waiting time, and the signaling link delay in the HLR and VLR are concerned in this study. In addition, the message transfer time, T_{msg}, between HLR and VLR can be modeled with the following equation to reflect the actual response time:

$$T_{\text{msg}} = T_{\text{ms}} + \left[\sum_{i=1}^{n+1} T_{\text{P}_i} + \sum_{i=1}^{n} T_{\text{C}_i}\right] + T_{\text{mr}},$$

where T_{ms} is the time from the message entering the MAP to leaving the signaling link control of MTP in figure 1; T_{mr} is the time from the message entering the signaling link control of MTP to leaving the MAP; T_{P_i} is the signaling data link propagation time; T_{C_i} is the cross time of gateway relay nodes; and n is the total number of gateway relay nodes. We assume $n = 2$ and $T_{\text{msg}} = 77$ ms in this analysis.

For the international roaming network set

$$\begin{aligned}\text{IR}_2 = \big\{(x, y) \mid (x, y) \in \big\{&(\text{Scenario A}, \text{Scenario C}),\\ &(\text{Scenario C}, \text{Scenario A}),\\ &(\text{Scenario C}, \text{Scenario C})\big\}\big\},\end{aligned}$$

(6) and (7) can be modified as follows:

$$\begin{aligned}Q'_{\text{rsp}} &= p\left[W_{\text{VLR}} + T_{\text{VLR}}^{\text{Query}}\right]\\ &\quad + (1-p)\left[2(T_{\text{msg}} + W_{\text{HLR}} + T_{\text{HLR}}^{\text{Query}}\right], \end{aligned}\tag{9}$$

$$\begin{aligned}U'_{\text{rsp}} &= \left[W_{\text{VLR}} + T_{\text{VLR}}^{\text{Update}}\right]\\ &\quad + q\left[2(T_{\text{msg}} + W_{\text{HLR}} + T_{\text{HLR}}^{\text{Update}}\right]. \end{aligned}\tag{10}$$

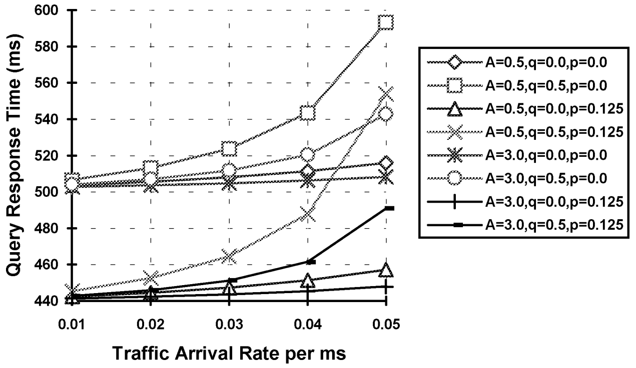

Figure 17. Query response times with various traffic arrival rates and gateway queueing delays. (A denotes τ.)

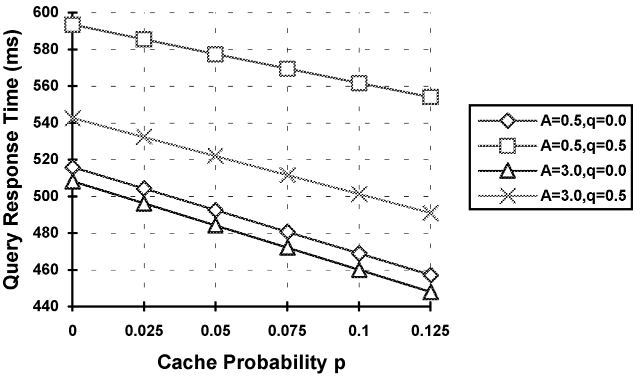

Figure 18. Query response times with $\lambda = 0.05$, various cache probabilities, and gateway queueing delays. (A denotes τ.)

Results shown in figures 17 and 18 obviously confirm that the cache probability and the handover probability simultaneously dominate the performance of query response time with high traffic arrival rate. As two cluster of lines demonstrated in figure 17, the cache probability almost has the constant time contribution to the query response time. The caching data effect is only perform well in the low traffic arrival rate condition with handover probability $q = 0.5$. Furthermore, the query response time is more sensitive in the small *shape* parameter τ of gamma distribution.

Figure 18 reveals the caching data effect on the query response time. Even though the cache probability at VLR is increasing, the performance of query response time decreases significantly. In addition, the handover probability size effects the query response time obviously. Again, we can find the linear coefficient p on (9) results in the approximate linear effect on the query response time of figure 18.

The results confirm that the impacts of the cache probability p and the handover probability q on the update response time are similar to those on the query response time. Therefore, we only demonstrate the update response time with various traffic arrival rates shown in figure 19. One of the important results indicates that the size of cache probability at VLR improves the update response time very little. Therefore, the lines shown in figure 19 almost exist by pairs. Another result is the update response time almost dominated by the handover probability size compared with the query response time under each traffic arrival rate.

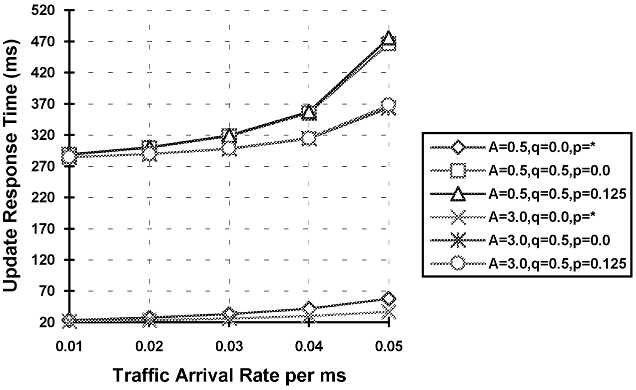

Figure 19. Update response times with various traffic arrival rates and gateway queueing delays. (A denotes τ.)

It is important to notice that our analysis is based on the scenario of transparent signaling connection. However, the time for GTT function is required in a non-transparent signaling connection case discussed in the next section.

4.4. International gateway with GTT function and queueing delay

The Scenario B and Scenario D are integrated STP which have SCCP relay function for GTT. In addition to the processor handling time in the cross-STP delay, an additional transit time of a message is required in the GTT translation function. The provisional values for the cross-STP delay and transit time in a relay point are specified in [4,5].

Let n be the total number of international gateway relay nodes for an international roaming message passing through. Also, let the mean processor handling time and the mean transit time be μ_{P} and μ_{R}, respectively. Furthermore, let p_{T} be the probability of performing GTT in a relay node.

Therefore, the international roaming network set

$$\begin{aligned}\mathrm{IR}_3 = \bigl\{(x,y) \mid (x,y) \in \bigl\{&(\text{Scenario A, Scenario B}),\\ &(\text{Scenario B, Scenario A}),\\ &(\text{Scenario A, Scenario D}),\\ &(\text{Scenario D, Scenario A}),\\ &(\text{Scenario B, Scenario B}),\\ &(\text{Scenario B, Scenario D}),\\ &(\text{Scenario D, Scenario B}),\\ &(\text{Scenario D, Scenario D}),\\ &(\text{Scenario B, Scenario C}),\\ &(\text{Scenario C, Scenario B}),\\ &(\text{Scenario C, Scenario D}),\\ &(\text{Scenario D, Scenario C})\bigr\}\bigr\}\end{aligned}$$

belongs to this case, also (6) and (7) can be modified as follows:

$$
\begin{aligned}
Q''_{\mathrm{rsp}} &= p\big[W_{\mathrm{VLR}} + T^{\mathrm{Query}}_{\mathrm{VLR}}\big] + (1-p) \\
&\quad \times \big[2T_{\mathrm{msg}} + n(\mu_{\mathrm{P}} + p_{\mathrm{T}}\mu_{\mathrm{R}}) + W_{\mathrm{HLR}} + T^{\mathrm{Query}}_{\mathrm{HLR}}\big], \\
U''_{\mathrm{rsp}} &= \big[W_{\mathrm{VLR}} + T^{\mathrm{Update}}_{\mathrm{VLR}}\big] \\
&\quad + q\big[2T_{\mathrm{msg}} + n(\mu_{\mathrm{P}} + p_{\mathrm{T}}\mu_{\mathrm{R}}) + W_{\mathrm{HLR}} + T^{\mathrm{Update}}_{\mathrm{HLR}}\big].
\end{aligned}
$$

Let

$$\Delta = \frac{Q''_{\mathrm{rsp}} - Q_{\mathrm{rsp}}}{Q'_{\mathrm{rsp}} - Q_{\mathrm{rsp}}} \quad \text{and} \quad \delta = \frac{U''_{\mathrm{rsp}} - U_{\mathrm{rsp}}}{U'_{\mathrm{rsp}} - U_{\mathrm{rsp}}},$$

it is easy to find that Δ and δ are linear function of $(\mu_{\mathrm{P}} + p_{\mathrm{T}}\mu_{\mathrm{R}})$. Therefore, the performance of the query and update response time on IR_3 will be dominated by the mean processor handling time and the mean transit time, i.e., $(\mu_{\mathrm{P}} + p_{\mathrm{T}}\mu_{\mathrm{R}})$. It implies that the performance of the update and query response times of IR_2 is better than that of IR_3. Again, it implies that the query and the update performances of the standalone STP scenario is better than that of integrated STP on the basis of the performance metrics proposed in this paper.

5. Conclusions

Four network interconnection scenarios and their related signaling aspects for the international roaming in mobile and personal communications are proposed in this paper. With these scenarios, three international roaming sets $\{\mathrm{IR}_1, \mathrm{IR}_2, \mathrm{IR}_3\}$ are derived in order to observe the signaling traffic performance at two-level databases. We also show that under some reasonable assumptions the signaling traffic of international roaming significantly impacts the databases' waiting times, query and update response times, and total transaction messages between each local database VLR and the centralized database HLR.

Furthermore, analysis results show that:

(i) caching data at VLR can not completely reduce the waiting time at HLR, but it can smooth the increasing rate of waiting time at HLR even in the worst case of the high handover probability and traffic arrival rate;

(ii) the cache probability and the handover probability simultaneously dominate the performance of the query and update response times;

(iii) the larger size of cache probability at VLR improves the update response time very little because all of update operations have to be performed at HLR in the existing mobile and personal communications networks;

(iv) the larger size of cache probability positively reduces the total numbers of message required for the international roaming traffic; and

(v) the query and update performances of standalone STP is better than those of the integrated STP because of the sum effect of mean processor handling time and the mean transit time in the integrated STP.

We believe that our proposed scenarios, international roaming network sets, HLR and VLR waiting times, and signaling response time formulation contribute a good basis to further study the signaling traffic issues on the international network. In addition, more detailed analysis about the combination of signaling nodes and our proposed interconnection scenarios are desired for future work. Finally, a new efficient data management strategy for distributed hierarchical databases is required to simultaneously improve the performance of the query and update response time for the future mobile and personal communications.

References

[1] I.F. Akyildiz and J.S.M. Ho, On location management for personal communications networks, IEEE Commun. Mag. 34 (1996) 138–145.

[2] R. Alonso, D. Barbara and H. Garcia-Molina, Data caching issues in an information retrieval system, ACM Trans. Database Syst. 15 (1990) 359–384.

[3] *CCITT Blue Book*, Recommendation Q, 706.

[4] *CCITT Blue Book*, Recommendation Q, 701–716.

[5] *CCITT Blue Book*, Recommendation Q, 766.

[6] *GSM 03.01*, European digital cellular telecommunications system (phase 2): Network functions, ETSI, Sophia Antipolis (1994).

[7] *GSM 03.03*, European digital cellular telecommunications system (phase 2): Numbering, addressing and identification, ETSI, Sophia Antipolis (1994).

[8] *GSM 03.05*, European digital cellular telecommunications system (phase 2): Technical performance objectives, ETSI, Sophia Antipolis (1992).

[9] *GSM 09.02*, European digital cellular telecommunications system (phase 2): Mobile Application Part (MAP), ETSI, Sophia Antipolis (1994).

[10] *GSM MoU Permanent Reference Document: IR. 21*, MoU roaming database, structure, and updating procedure (1993).

[11] *GSM MoU Permanent Reference Document: IR. 22*, SCCP signaling aspects for roaming (1993).

[12] N.A.J. Hasting and J.B. Peacock, *Statistical Distributions* (Wiley, New York, 1975).

[13] R.V. Hogg and A.T. Craig, *Introduction to Mathematical Statistics* (Macmillan, 1970).

[14] B. Jabbari, Intelligent Network concepts in mobile communications, IEEE Commun. Mag. 30 (1992) 64–69.

[15] B. Jabbari, Teletraffic aspects of evolving and next-generation wireless communication networks, IEEE Personal Commun. Mag. 3 (1996) 4–9.

[16] B. Jabbari, G. Colombo, A. Nakajima and J. Kulkarni, Network issues for wireless communications, IEEE Commun. Mag. 33 (1995) 88–99.

[17] R. Jain, *The Art of Computer Systems Performance Analysis* (Wiley, New York, 1991).

[18] R. Jain, Y.B. Lin, C. Lo and S. Mohan, A caching strategy to reduce network impacts of PCS, IEEE J. Selected Areas Commun. 12 (1994) 1434–1444.

[19] P.J. Kühn, C.D. Pack and R.A. Skoog, Common channel signaling networks: past, present, future, IEEE J. Selected Areas Commun. 12 (1994) 383–393.

[20] Y.B. Lin, Determining the user locations for personal communications services networks, IEEE Trans. Vehicular Technol. 43 (1994) 466–473.

[21] Y.B. Lin and S.K. DeVries, PCS networking signaling using SS7, IEEE Personal Commun. Mag. 2 (1995) 44–55.

[22] Y.B. Lin and S.Y. Hwang, Comparing the PCS location tracking strategies, IEEE Trans. Vehicular Technol. 45 (1996) 114–121.

[23] E.H. Lipper and M.P. Rumsewicz, Teletraffic considerations for widespread deployment of PCS, IEEE Network (1994) 40–49.

[24] A.R. Modaressi and R.A. Skoog, Signaling system No. 7: A tutorial, IEEE Commun. Mag. 28 (1990) 19–35.

[25] S. Mohan and R. Jain, Two user location strategies for personal communications services, IEEE Personal Commun. Mag. 1 (1994) 42–50.

[26] G.P. Pollini and D.J. Goodman, Signaling system performance evaluation for personal communications, IEEE Trans. Vehicular Technol. 45 (1996) 131–138.

[27] G.P. Pollini, K.S. Meier-Hellstern and D.J. Goodman, Signaling traffic volume generated by mobile and personal communications, IEEE Commun. Mag. 33 (1995) 60–65.

[28] R.B. Robrock II, The Intelligent Network – changing the face of telecommunications, Proc. IEEE 79 (1991) 7–20.

[29] *Specifications for GSM*, Directorate General of Telecommunications, Taiwan (1994).

[30] P.E. Wirth, Teletraffic implications of database architectures in mobile and personal communications, IEEE Commun. Mag. 33 (1995) 54–59.

Jyhi-Kong Wey is a researcher of Telecommunication Labs. (TL) of Chunghwa Telecom Co., Ltd., in Taiwan, R.O.C. He received his B.S. and M.S. degrees in statistics from the National Chen Chi University and the National Cheng Kung University in 1981 and 1983, respectively, and Ph.D. degree in computer and information science from the National Chiao Tung University in 1996. From July 1985 to June 1996, he was with the Network Planning Lab. of TL for the study and implementation of SS7 protocol, Intelligent Network technologies, and Personal Communications Services and mobile communications systems. His research interests include telecommunication network, communications protocol, nomadic communications, object modeling, and distributed database systems. He is a member of ACM, IEEE, and IICM.
E-mail: jkwey@ms6.hinet.net

Wei-Pang Yang is a professor in the Department of Computer and Information Science of National Chiao Tung University (NCTU). He received the B.S. degree in mathematics from National Taiwan Normal University in 1974, and the M.S. and Ph.D. degrees from NCTU in 1979 and 1984, respectively, both in computer engineering. Since August 1979, he has been on the faculty of the Dept. of Computer Engineering at NCTU. In the academic year 1985–1986, he was awarded the National Postdoctoral Research Fellowship and was a visiting scholar at Harvard University. From 1986 to 1987, he was the Director of the Computer Center of NCTU. In August 1988, he joined the Dept. of Computer and Information Science at NCTU, and acted as the Head of the Dept. for one year. Then he went to IBM Almaden Research Center in San Jose, California for another one year as visiting scientist. From 1990 to 1992, he was the Head of the Dept. of Computer and Information Science again. He was assigned to be the Deputy Director of Electrical & Information Center from 1994 to 1996. His research interests include database theory, database security, mobile communications databases, object-oriented database, image database and Chinese database systems. He was the winner of the 1988 and 1992 Acer Long Term Award for Outstanding M.S. Thesis Supervision, and the winner of 1990 Outstanding Paper Award of the Computer Society of the R.O.C. He also obtained the 1991–1993 Outstanding Research Award of National Science Council of the R.O.C. Dr. Yang is a member of ACM, IEEE, and the Phi Tau Phi Society.

Lir-Fang Sun is the Managing Director of the Planning and Technology Promotion Department, Telecommunication Labs. (TL) of Chunghwa Telecom Co., Ltd. She received the B.S. and M.S. degrees in computer science from Tamkiang University and National Chiao Tung University in 1975 and 1981, respectively, and the Ph.D. degree in electrical engineering from National Taiwan University in 1988. From December 1977 to June 1982, she was with Switching Systems Lab., TL/MOTC for implementation of PABX prototype and PEX-250. From 1982–1984, she was a senior software designer in C400 auto-dial project. From 1984 to 1989, she was the project leader of new telephone services project in Network Planning Lab. From 1989 to 1991, she visited AT&T Bell Labs. in Chicago as a technical staff for interworking of SS7 with 5ESS. Since 1991, she has been in charge of the IN project in Network Planning Lab. She was assigned to be the director of Labor Safety and Health Dept., since November 1992, and was voted to be the chairman of TP SIG, OSI Working Group/ ROC since December 1992 to July 1996. Her research interests include switching systems, telecommunication databases, distributed systems, software engineering, project management for large scale software, Intelligent Network, Signaling System No. 7, Transaction Processing, and protocol design and testing.