# 行政院國家科學委員會專題研究計劃成果報告

## 行動計算環境中之交易處理(2/2)
## Transaction Processing in Mobile Computing Environment (2/2)

計畫編號: NSC89-2218-E009-029
執行期限: 89 年 8 月 1 日至 90 年 7 月 31 日

主持人:李素瑛　國立交通大學資訊工程系
sylee@csie.nctu.edu.tw

**中文摘要**

在行動計算環境中，交易可分為兩類：一類是為了記錄使用者位置(location management)而產生的交易，另一類交易則是由 mobile unit 所發出，目的是存取位於網路上之資料庫，這類型的交易我們又稱為行動交易(mobile transaction)。

在使用者位置管理方面，主要解決的是是 HLR 為效率瓶頸(performance bottleneck)的問題。針對此問題，我們提出了利用複製(replicating)及叢聚(clustering)的技巧，將 HLR 複製並分為幾個叢聚。複製 HLR 可以提高系統可靠度並減少因為查詢所產生的通話建立延遲，然而卻也同時增加了 location update 的成本，因此我們利用叢聚的技巧，使得 location update 的成本降低。除此之外，我們也根據使用者位置管理交易(location-management-related transaction)的特性，提出了一套名之為 preemptive read-one-write-all 的並行控制協定。

在行動交易方面，我們主要解決的是當使用者在進入新的基地台通訊區域時，由於無法取得可供通訊的頻道而被迫終止(forced termination)的問題。目前有關 handoff 的研究都只考慮到電話的聲音服務，我們考慮了交易執行的特性，分別由個人通訊系統及並行控制協定的角度，提出幾個降低交易因通訊中斷而被迫終止之機率的方法。

關鍵詞: 行動計算，位置管理，行動交易，並行控制

**ABSTRACT**

More and more information services are provided on the wireless networks. The characteristics of the wireless medium and the mobility of the mobile users raise new challenging problems. One of the major problems about location management is that the HLR tends to be the performance bottleneck. In our project, we try to replicate the HLR and distribute them to the network. By this way, we can prevent the HLR from bring the performance bottleneck. However, replicating the HLR will cause extra cost since an update operation should update all the replicas. We divide the replicas of HLR into clusters to reduce the update cost. In addition, we propose a new concurrency control protocol called preemptive Read-One-Write-All protocol for location-management-related transactions.

Due to long network delay of wireless links, transactions will be long-lived transactions. In such a situation, the occurrence of handoff is inevitable, and thus a wireless link held by a mobile station crossing cell boundaries might be forced to terminate. It is undesirable that an active transaction is forced to terminate. In our project, we first analyze the performance of conventional concurrency control protocols (two-phase locking, timestamp ordering and optimistic concurrency control) in mobile computing environments. We then propose two handoff schemes — Queue-Limit scheme and Guard Channel scheme to reduce the probability of forced termination of transactions. Furthermore, we propose a new concurrency control protocol called Early-Release of Lock (ERL) protocol to alleviate the blocking effect of locking-based protocols in mobile computing environments. The experimental results reveal that the proposed handoff schemes can reduce the probability of forced termination of transactions and ERL protocol can reduce the waiting time of an active transaction and thus improve the transaction throughput.

## 1. Introduction:

Recent technology advances in portable computers (notebooks) and in wireless communication have made the mobile computing environment a reality. The characteristics of the wireless medium and the mobility of the data consumer raise new challenging problems. There are two types of transactions in mobile computing environments. One is to support the location management. That is, the queries and updates in home location registers and visitor location registers. The other is the transaction submitted from a mobile user to access the databases located in the wired network.

To support user mobility, the PCS networks have to store and maintain location information of mobile units (MUs) so that an incoming call can be delivered to the target MU. The operations on location information consist of location updates and location queries. An update occurs when an MS changes location. A query occurs when an MU needs to be located, e.g., to deliver an incoming call to this MU. The widespread deployment of PCS will lead to a tremendous increase in the number of updates and queries to the location database. Thus, a key challenge for location management is to develop an efficient database architecture so that the location data can be readily available for signaling such as call setup and routing.

The current approach to support user mobility requires a two-level database [1]. This architecture consists of a home location database (or Home Location Register, HLR) and a visitor location database (or Visitor Location Register, VLR). This HLR-VLR architecture has been established as the de facto, and is widely used in the industrial field such as Global System for Mobile Telecommunication (GSM) for Europe and the IS-41 recommendations for North America. In such systems, HLR is centralized and stand-alone in the network. Some queries and updates have to travel long distance when the incoming calls or the MUs are far away from the HLR. Another drawback of conventional HLR-VRL architecture is that HLR tends to be the bottleneck. A typical GSM mobile customer traffic profile is as follows [2].

- 9 – 12 min/day usage
- 60/40 split (60% of traffic is from wireless network to fixed network)
- average call duration approximately 50 sec.
- switching the phone on/off approximately 4 times per day

HLR implementations are commercially available to support approximately 300,000 customers. Each user with the above mentioned traffic profile would submit approximately 20 HLR operations per day (location updating, routing, authentication, network attachment). For a GSM with 300,000 subscribers, the load in HLR will thus be approximately 6,000,000 operations per day. From SONOFON GSM experience, 12-13% of the operations are during the busy hour, i.e. about 800,000 queries per hour (or 220 transaction per second). Peak value may be 50% higher [3]. Such heavy load cannot be supported by the standard relational databases even on the most powerful processing equipment available today. As a consequence, the HLR can be potential bottleneck and cannot guarantee the quality of service (e.g., call setup time). Therefore, we will try to propose a new database architecture to solve the problem that HLR tends to be performance bottleneck.

Now we describe the problems arisen with mobile transactions. In cellular network architecture, an MU, which is a notebook or laptop computer with a wireless network card, communicates with servers on the wire network through a base station (BS). The communication area covered by a base station is called a cell. When a mobile unit is within the cell of a base station, the base station will provide a communication channel to the mobile unit if there are some idle channels available in this cell. A general mobile information system model, as shown in Figure 1, has been proposed [4,5,6].

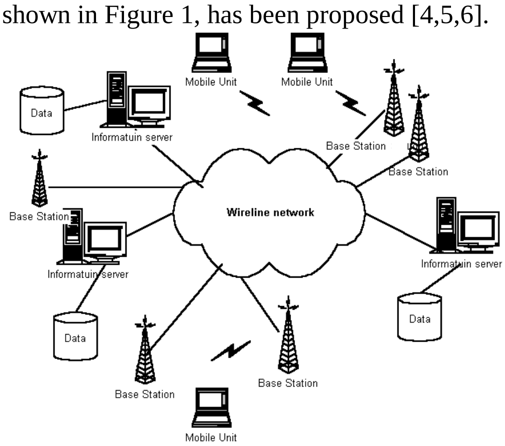


Figure 1. Mobile information system model

This model consists of stationary and mobile units. Stationary units are classified as either fixed hosts or base stations. Fixed hosts, which are information servers with associated databases, are connected to the existing wire network, and are not capable of connecting to a mobile unit. A base station is equipped with a wireless interface and is capable of connecting to mobile units. In other words, the base station plays a role of coordinator and communication interface between mobile units and stationary units.

The mobile user accesses the database by submitting transactions. A transaction submitted from a mobile unit is sent to the base station through a wireless link and then sent to the information server via the existing wired network. We call the transactions submitted from a mobile unit as mobile transactions.

During the executing time, a transaction may need the user's participation to input data. Many studies have pointed out that the cost of a call setup is very expensive [7]. In order to avoid reestablishing the communication each time when the transaction needs the user's interaction, we assume that the communication link must be kept while the transaction is executing.

When a mobile unit is within the cell of a base station, the base station will provide a communication channel to the mobile unit if there are some idle channels available in this cell. Due to the movement of a mobile user, when a mobile user enters a new cell, the new base station should provide an idle channel to the mobile unit to continue its communication. This process is called handoff. If there is no idle channel in the new cell to provide for the user, then the connection will be forced to drop or terminate. The forced termination of an active transaction not only interferes the users but also wastes the system resources since the database should be rolled back and the transaction should be restarted later. As we know, handoff will occur when a mobile unit enters a new cell. With the growth of the number of mobile users and for providing good quality of services, it is a trend to set up more base stations. As a consequence, the handoff frequency will become higher in the future. Therefore, the probability of forced termination of transactions will also become higher.

Many schemes have been proposed to reduce the probability of forced termination for voice calls (VC) [8,9,10]. However, little attention is paid to transaction calls (TC). In our project, we will propose some handoff schemes and concurrency control protocol to reduce the probability of forced termination of transactions.

# 2. Result

## 2.1 Location Management in PCS System

In this project, we present two observations in mobility management of PCS network. The first observation is that the frequency of an HLR being queried is more than that of it being updated. Based on this observation, we propose to replicate the HLR. The system engineer can determine the number of replicas according to what quality-of-service the system would provide and how much cost it would pay. The optimal placement of the replicas is developed using the technique of linear programming in this project. Another observation is that the update procedure is executed after the movement of an MS. According to this observation, we propose a new concurrency control protocol based on the ROWA (read-one-write-all) protocol and referred to as Preemptive ROWA (PROWA).

Replicating an HLR and distributing them in the network can increase the reliability of HLR and reduce the communication cost as well as query completion time as compared with GSM and IS-41. On the other hand, replicating an HLR would increase the cost to set up and to maintain database copies. Besides, the database load will increase due to the extra overhead of update operations making the contents of databases consistent. In CHLR (clustering HLR), the replicas are divided into several clusters. An update is executed in one of the replica in each cluster and a query is executed in all the replicas in the cluster. By this way, we can reduce the cost of making the contents in the replicated HLR consistent by causing little overhead in location query.

From the experimental results, we can find that if the value of query-to-update ratio is small, although the average query completion time and misrouting probability in CHLR are a little longer (higher) than those in RHLR, the average cost per query in CHLR is far less than that in RHLR. In other words, CHLR performs better than RHLR if the call-to-mobility of users is low. On the other hand, when the value of query-to-update ratio is high, CHLR may decay to be the same as RHLR.

## 2.2 Mobile Transaction Processing

For mobile transaction processing, we build an analytic model and simulation environment to analysis the performance of traditional concurrency control protocols. We also propose two handoff schemes – Queue-Limit and Guard Channel schemes. Queue-Limit scheme works well when timestamp ordering or optimistic concurrency control is employed. However, it does not improve the system performance when two-phase locking is employed. Both the static GCS and dynamic GCS dramatically decrease the probability of forced termination of transactions while causing little increase in blocking probability of voice calls (new or handoff) and new transaction calls. From the experimental results, we can find that static GCS can fail in full utilization of the limited channels. Since the handoff traffics which depend on the mobility of users are not always stationary, static GCS can lead to a waste of or lack of the number of guard channels. On the other hand, dynamic GCS can achieve a better utilization of channels since the number of guard channels is dynamically adjusted.

In GCS, since some channels are reserved, the precious wireless channels may not be fully utilized. Although Queue-Limit scheme can achieve a better channel utilization, when the degree of data contention becomes higher, the performance of Queue-Limit scheme will degrade. This is because 2PL performs poorly in the presence of increasing data contention. In order to alleviate the blocking effect of 2PL, we propose a new locking-based protocol for Queue-Limit scheme to increase the concurrency of transaction execution. This protocol distinguishes the modes of locks as strict mode and soft mode. Locks in strict mode are just like the locks in conventional 2PL. When a transaction is suspended, all the locks held have to be converted to soft mode. A lock in soft mode can be preempted by an active transaction. In other words, the active transactions do not need to wait for the lock release of the suspended transactions. Therefore, the concurrency of transaction execution can be enhanced by this way.

From the experimental results, we can find that both Queue-Limit protocol and ERL protocol perform better than NPS in reducing the probability of forced termination of transactions and transaction throughput. However, when the degree of data contention increases, since Queue-Limit protocol will elongate the lock holding time of a transaction and thus degrade the system performance. In ERL protocol, the locks held by a suspended transaction will be converted to soft mode and will not affect the transaction execution of active transactions. Therefore, the concurrency of transaction execution can be enhanced.

## 3. Conclusion and Future Works

### 3.1 Conclusion

In this project, two issues are addressed. The first is that HLR tends to be the performance bottleneck. The second is the forced termination of transactions. We propose to replicate the HLR to solve the problem of HLR being performance bottleneck. The system engineer can determine the number of replicas according to what quality-of-service the system would provide and how much cost it would pay. The optimal placement of the replicas is developed using the technique of linear programming in this project.

Replicating an HLR and distributing them in the network can increase the reliability of HLR and reduce the communication cost as well as query completion time as compared with GSM and IS-41. On the other hand, replicating an HLR would increase the cost to set up and to maintain database copies. In order to reduce the cost of making the contents in the replicated HLR consistent, we propose a new database architecture called CHLR (clustering HLR).

For mobile transaction processing, we propose two handoff schemes – Queue-Limit and Guard Channel schemes. Queue-Limit scheme works well when timestamp ordering or optimistic concurrency control is employed. However, it does not improve the system performance when two-phase locking is employed. In GCS, since some channels are reserved, the precious wireless channels may not be fully utilized. Although Queue-Limit scheme can achieve a better channel utilization, when the degree of data contention becomes higher, the performance of Queue-Limit scheme will degrade. This is because 2PL performs poorly in the presence of increasing data contention. In order to alleviate the blocking effect of 2PL, we propose a new locking-based protocol for Queue-Limit scheme to increase the concurrency of transaction execution.

### 3.2 Future Works

In CHLR, the replicas are uniformly partitioned to each cluster. However, The load in each cluster may be different. To assign each cluster the same number of replicas will cause the geographic sizes of some clusters become very large. As a consequence, the completion time of queries in these clusters will become longer. In the future, we will try to develop a new clustering mechanism which allows the clusters to be of different size. In addition, the content in each replica is assumed identical in the current proposed scheme. In the future, the technique of data mining can be used to obtain the user profiles. And then the content of each replica can be decided according to the user profiles.

For mobile transaction processing, we will continue our research on developing new concurrency control protocols for different handoff schemes. Since sometimes the disconnection of a mobile station is predictable (by the strength of signal), a mobile station can cache some data before the disconnection and then executes the transaction locally when the mobile station is disconnected. However, the result of transaction execution has to be sent to the database after the connection is re-established. To develop a new concurrency control protocol to cope with disconnected operations will also be our future research direction.

## Reference

[1] Y.-B. Lin, "Introduction to Mobile Network Management," Baltzer Publisher, 1997.

[2] Gabelgaard, "The (GSM) HLR – Advantages and Challenges," Proc. of third IEEE Int. Conf. On Universal Personal Communications, pp.335-339, Sept. 1994.

[3] S. Mohan, and R. Jain, "Two User Location Strategies for Personal Communication Services," IEEE Personal Communications, pp. 42-50, First Quarter 1994.

[4] T. Imerielinski and B.R. Badrinath, "Wireless Mobile Computing: Challenges in Data Management," *Communication of ACM*, Vol.37, No.10, pp. 19-28, 1994.

[5] P. Evaggelia and B. Bharat, “Maintaining Consistency of Data in Mobile Distributed Environments,” *Proceeding of 15th Internal Conference on Distributed Computing Systems*, pp. 404-413, 1995.

[6] G. D. Walborn and P. K. Chrysanthis, “Supporting Semantics-Based Transaction Processing in Mobile Database Applications,” *Proceeding of the 14th IEEE Symp. on Reliable Distributed Systems*, pp.215-224, Sep. 1995.

[7] H. Dunhan and A. S. Helal, “Mobile Computing and Databases: Anything New?” *ACM SIGMOD Record*, Vol. 24, No. 4, pp. 5-9, Dec. 1995.

[8] S. Boumerdassi, "An Efficient Reservation-based Dynamic Channel Assignment Strategy," *Proc. of 1st Intl. Conf. on 3G Mobile Communication Technologies*, pp. 352 - 355, 2000.

[9] D. Hong and S. S. Rappaport, "Traffic Model and Performance Analysis for Cellular Radio Telephone System with Prioritized and Non-prioritized Handoff Procedures," *IEEE Trans. On Vehicular Technology*, vol. 49, no. 5, pp. 2037 - 2039, 2000.

[10] Y. -B. Lin, A. Noerpel and D. Harasty, "Sub-rating Channel Assignment Strategy for Hand-offs," *IEEE Transactions on Vehicular Technology*. pp. 122-130, Feb 1996.