Replication in Mobile Information Systems

Hagen Höpfner* *hoepfner@iti.cs.uni-magdeburg.de*

Abstract: Replication techniques are already used in distributed database systems for increasing performance and availability. But also mobile information systems (MIS) need replication for making required data available at any time at every place. In this paper we discuss what replication in MIS means and which additional requirements are caused by mobility.

1 Introduction and Motivation

Due to the increasing usage of mobile equipment new application areas for information systems arise. Today the availability of information at any time at every place becomes more important and requires new techniques which consider the special properties of the underlying hardware. The central problems are the limitation of availability and bandwidth as well as the latency of network access in mobile environments. Current wireless technologies like GSM, GPRS or HSCSD guarantee a nearly area-wide disposability but are quite expensive and slow. Also a lot of scenarios can be mentioned, that prohibit a permanent wireless connection (e.g. caused by shielding). Replication is already used in distributed DBMSs. Here performance and availability enhancements are the deciding reasons for managing redundant data. In mobile environments replication is essential because a permanent connection can not be guaranteed but a permanent availability is claimed.

In this paper we discuss what replication in mobile information systems means. In Section 2 we explain the chosen network architecture. Furthermore, we present in Section 3 a generalised replication process and show aspects involved in the particular steps. Section 4 concludes the paper and gives an outlook on future works.

2 Basic Network Architecture

The pervasiveness of cellular mobile phone networks causes the usage of a basic network architecture which is based on the architecture introduced in [7]. *Fixed hosts* are connected to a *fixed network*. Some of these fixed hosts have wireless interfaces and act as *base stations* (BS). These hosts permit the connection of *mobile units* (MU) to the information system. Furthermore, the mobile units build a *wireless radio cell* if they are located in the same geographical area and if they are connected to the same base station. We supplement this basic architectureby some logical concepts and devices.

Database servers (DS): The sources of data to be used by mobile clients are fixed hosts, on which a database is located. These servers may include distributed or integrated systems.

Mobile (database) clients (MC): The devices actually using the replicated data are mobile units applying a replication mechanism for managing local replicas of fragments of the database of the DS they are connected to via an unstable network, that most likely is limited regarding bandwidth.

Because of the given space limitation we can not discuss other network architecture like ad-hoc- or peer-to-peer-nets, which are also of interest for mobile information systems.

* Research supported by the DFG under grant SA 782/3-1.

3 The Replication Process (RP)

One of the central objectives of the usage of DBMS , on which most information systems are based, is to avoid unnecessary redundant data. Because the databases on mobile clients and/or database servers may become updated over time, replication must be a continuous process. In case of considering replication as a non-continuous process, the replica is a backup or an "useless" redundant copy and will not become updated but soon outdated.

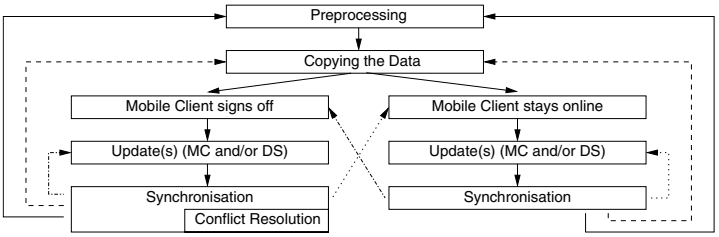

Figure 1: The Replication Process

As shown in Figure 1, the process is splitted into two parts. The first one describes the replication with mobile clints, which disconnect after getting the data. In the second part mobile units permanently stay online[1]. In both cases the database on the DS must be preprocessed. The resulting replica set is subsequently copied to the mobile unit. After that, a modification of the data is possible on the database server as well as on the mobile client. Especially with off-line working mobile clients, updating causes inconsistent data. This means that different versions of data may occur. Hence, mobile clients and database servers must synchronise with each other. Because of the high probability that off-line modified data on a mobile client conflicts with the version on the database server, advanced concepts for conflict resolution are necessary. After synchronisation is finished, three ways are possible. First, a complete new replication can be started. In the second case, already existing preprocessing information[2] are reused. In the third case, no new data is replicated. After synchronising the updates, the mobile client signs off or stays online in the face of new updates.

3.1 Preprocessing

The preprocessing step of the RP is the definition of the data which shall be copied. This is necessary, because a complete copy of the data of the base station to the mobile host is not feasible in information systems with a huge amount of data. The replica must be only a part of the entire database.

[3] presents an overview about data reduction techniques.Non-content-based data reduction approaches allow the reduction of the data without an user interaction. In commercial mobile database systems replica sets[3] of data stored in the database are defined in such a manner. In contrast to non-content-based data reduction, content based data reduction allows to reduce the data in a more user dependent manner. Here the context of a user and/or of the users mobile equipment defines the replica set dynamically. Therefore the number of objects or the object size can be adapted by transforming queries or by transforming query results. For example, level of detail techniques (e.g. fisheye views) allow a format

[1] similar to classic replication in distributed databases
[2] e.g. semantically cached data
[3] may consist of tuples of one ore more tables in the relational case

specific adaption of query results to the context of the inquiring hosts. The combined use of several data reduction techniques should be possible.

Another aspect of the preprocessing step is to allow or deny updates on the replicas. This essential question corresponds directly to the definition of replica sets because the necessary conflict resolution in mobile environments must take the dis-connectivity of mobile units as well as the view update problem into account. Updateable replicas require a restrictive definition of replica sets.

Furthermore, some more mobility dependent aspects must be considered in the preprocessing step. The usage of caches, which is well known from distributed database systems, improves the performance of information systems. Especially in mobile information systems caching strategies are used in a semantic manner. That means that the results of former queries are reused for faster answering new, semantical equivalent queries [4]. The usability of those approaches for moving MCs is presented e.g. in [8]. In contrast to semantic caching, profiling techniques use explicit collected hints about users mobility and access patterns to predefine user specific replica sets. For example, in [9] the location information are extracted from the cell where the MC is included in. If a MC tries to replicate data later on from the same DS, the history is used to pre-calculate possible relevant data.

3.2 Copying the Data

As described above, the second step in the RP is the delivery of the data. Here, the central problem is the connectivity between the mobile client and the base station.

The kind of connection between base station and mobile client influences the choice of a transport media. If no network connection can be established, a *disk-file-based* transport has to be used. Because this method requires additional hardware like floppy drives, this *file-based* approach is not suitable for small mobile clients like PDAs. Therefore, a network connection must be available. For this purpose there are three approaches proposed in the literature: (1) *Agent based replication*, (2) *Network protocol based replication* and (3) *Broadcast based replication*. In the agent based replication approach [6] application and database agents are used for the data transfer. Database agents are the interface to the database. They communicate with application agents using *primitive application methods*. These are predefined methods which allow the querying of a database and of agent data. The application agents transport the data to the target database agent, which then coordinates the necessary database operations on its site.

The network protocol based replication allows the use of standard network protocols for the data transfer. Most commercial database management systems support the HTTP protocol.But also other protocol types like FTP or SMTP/POP are usable. This approach may also provide data in form of files.

In systems with data, which is probably interesting for a great number of mobile clients, a broadcast approach can be used. In such scenarios the transport of the data can be done using a broadcast medium like radio waves. There are several publications which discuss the usability of broadcast techniques for the replication in mobile information systems (e.g. [4]). But, the usage of broadcast techniques require either an additional up-link-chanel or a lot of additional functionality [5] on the client site.

3.3 Updating the data and synchronisation

One aim of the use of mobile equipment is the possibility of modifying the data at any place at any time. In this section we discuss the problems of off-line updates. We do not

[4] at best without querying the DS

[5] for separating needed information from the broadcast stream

refer to the online case, because it is the classic and well known problem of updating data in distributed database systems. Here, normally quorum or primary copy approaches are used for keeping the database in a consistent state.

If the updates were done off-line, it is not as easy to ensure that the consistency is given as in the online case. In fact, off-line modifications of the data cause differences between the database on the DS an the replicated data on the MC. This problem must be resolved in the synchronisation step. There are two popular approaches for handling updates in mobile database systems. The first one is the use of tentative transactions [2]. Here, all updates of mobile data are temporary updates and may be revoked in the synchronisation step, if a given scope rule is violated. The second approach uses different versions of the snapshot for storing the data.As shown in [5] this multi-version reconciliation approach works automatically for simple but not for complex conflicts. For real live problems both approaches require additional application logic. In the Bayou project [1] the transaction based approach is complemented by *per-write dependency checks* and *per-write merge procedures*. These extensions allow the recognition and the automatic adjustment of conflicts, provided that the possibility of the appearance of these conflicts is known at the implementation time of the information system. Also, for advanced conflict handling in the multi-version reconciliation approach knowledge of possible conflicts is needed.

As shown, the two steps ”updating the data” and ”synchronisation” belong together and depend on the connectivity of the mobile clients. If data is modified online, the additional conflict resolution functionality is not necessary. The RP must react to this fact by adaptively choosing a conflict resolution level.

4 Conclusions and Outlook

In this paper we discussed issues of replication in mobile information systems. Therefore we defined a generalised replication process and looked at the three main steps of this process. Because of the given space limitation we only discussed cellular nets. In further works we will look at replication in ad-hoc- and other peer-to-peer-nets. Furthermore we will make the replication process more adaptive considering the requirements of its users. For the context-based preprocessing we already worked out first ideas. Currently, we work on the adaptive synchronisation of modified data in mobile information systems.

References

[1] A. Demers, K. Petersen, M. Spreitzer, D. Terry, M. Theimer, and B. Welch. The bayou architecture: Support for data sharing among mobile users. In *Proc. of WMCSA'98*, pages 2–7, Santa Cruz, USA, December 1994. IEEE.

[2] J. Gray, P. Helland, P. O'Neil, and D. Shasha. The dangers of replication and a solution. *SIGMOD Record*, 25(2):173–182, 1996.

[3] A. Heuer and A. Lubinski. Data Reduction - an Adaptation Technique for Mobile Environments. In T.Kirste and B. Urban, editors, *Proc. of the IMC'98*, pages 139–144, Rostock, Germany, November 1998. FHG IGD Rostock.

[4] F. Pedone, R. Guerraoui, and A. Schiper. Exploiting atomic broadcast in replicated databases. In *Proc. of EuroPar (EuroPar'98)*, September 1998.

[5] Shirish H. Phatak and B.R. Badrinath. Transaction-centric reconciliation in disconnected databases. *ACM Monet Journal*, 53, 1999.

[6] E. Pitoura and B. Bhargava. A framework for providing conistent and recoverable agent-based access to heterogeneous mobile databases. *Sigmod Record*, 24(3):44–49, September 1995.

[7] E. Pitoura and G. Samaras. *Data Management for Mobile Computing*. Kluwer Academic Publishers, Dordrecht, 1998.

[8] Q. Ren and M. H. Dunham. Using semantic caching to manage location dependent data in mobile computing. In *Proc. of the 6th MobiCom*, pages 210–221. ACM Press, 2000.

[9] S.-y. Wu and Y.-t. Chang. An active replication scheme for mobile data management. In A. Chen and F. Lochovsky, editors, *Database Systems for Advanced Applications, Proc. of the 6th DASFAA*, pages 143–150. IEEE, 1999.