Mobile Computing: Data Management Issues

Alfredo Goñi and Arantza Illarramendi

1. Introduction

In the last few years, the use of portable computers and wireless networks has been widespread. The combination of both opens the door to a new technology: **mobile computing**. Although the wireless communication networks were designed for the transport of voice signals, their use for data transport is growing.

Mobile computing allows users to access from anywhere and at anytime the data stored in repositories of their organizations (i.e. the DBs of the company in which they work) and also for available data in a global information system through Internet. Many professionals use mobile computers for their work: sales personnel, emergency services etc. in order to obtain and send information in the place where and at the moment when they actually need it. Moreover, there exist applications in this new framework where the location is an important aspect, such as those that provide information about the nearest hotels, restaurants, etc. That is, mobile computing adds a new dimension to distributed data computation and that dimension enables a new class of applications.

So far distributed data management has been mainly considered for fixed computers. Distributed Data Base Management Systems, Federated Databases, Interoperable Databases and Global Information Systems are topics in which a great research effort is being made. The new framework of mobile computing can profit from some new proposals on those topics. However, specific problems related to this new framework must be taken into consideration. Some of these problems are intrinsic to portable computers, which generally provide fewer resources than fixed hosts because they must be small, light and consume little energy. Other problems are related to the wireless connection, which presents a poor quality and is influenced by a multitude of factors that cause the wireless networks to have a high rate of errors and a limited bandwidth. It is also necessary to add to the previous problems the continuous disconnections that occur. One could say that mobile computing is the worst case of distributed computation since fundamental assumptions about connectivity, immobility and scale have lost their validity.

With respect to related works, there are different mobile software systems that can be grouped in the following way:

- Systems that allow a disconnected and/or weakly connected access to *file systems*. Among the different issues that have to be dealt with are: the possibility of pre-fetching files for later access, management of cached data, model of consistency used, how to propagate changes, transparency of support to client applications, transparency of mechanisms to users, etc. In the related literature, there can be found many works that deal with some of these issues among which the following can be cited: Coda (probably the most known) [1], Mazer/Tardo [2], Ficus [3], AirAccess2.0 [4], Howard [5] and LapLink [6].
- Systems that allow a weakly connected access to *databases* such as the Bayou system that proposes and implements an architecture for mobile-aware databases. In the system architecture there are several Bayou servers containing the full replicated data and several client applications interacting (reading and writing) with those servers and that are aware that they may be working with inconsistent data. Disconnected operation with the system is not allowed [7].

- Systems that allow a disconnected and weakly connected access to the *world wide web*. These systems differ with respect to the systems that access to file systems in the following aspects: a URL is not a file reference because it can contain embedded references, so some content filtering and transformation may need to be made. It can be also possible to have dynamic URLs. In these systems pre-fetching and caching techniques may also be applied and *proxies* can be added to the architecture. In the following works, the proxy is added to the client part of the system: Caubweb [8], Teleweb [9], Weblicator [10]. In other works like TranSend [11] and Digestor [12] the proxy is added to the web server and finally, in WebExpress [13] there are proxies in both parts, client and server.
- Systems that provide an environment for the development of mobile applications. For example the Rover Toolkit that is software that supports the construction of *mobile-transparent* and *mobile-aware* applications based on the idea of *relocatable dynamic objects* and *queued remote procedure call* [14]. Also interesting is the work developed by the DATAMAN group that has defined a set of classes called *Mobjects* that would form the basis of a toolkit to implement applications for mobile computing devices [15].

In the rest of this chapter we present basic concepts and issues related to data management for mobile computing. First we present a motivation. Next, we show the widely accepted architecture. Then, we explain briefly different wireless networks. Later on, we introduce the problems inherent to the framework and specifically, the impact of mobile computing in the area of data management. After that, we explain the main features of the more frequently used communication models and those related to agent technology. We finish with some design characteristics, conclusions and further reading.

2. Motivation

There is no doubt that mobile computing offers new computing opportunities to the users. In the following we describe three different scenarios in which three main features of the new paradigm are reflected: *a) access to distributed information repositories from anywhere at anytime, b) information delivery by broadcasting and c) provision of user tailored information*.

In the first scenario we consider the case of a sales person equipped with a Personal Computer. Let us suppose one day in her life. Our sales person plans her labor day in advance, so before going to bed she switches on her mobile computer and asks if there have been new assignments. Once she revises the messages, the sales person adds the new tasks to her agenda and plans the following day's itinerary. The following morning our sales person leaves home to visit the first client of the day. Before leaving she asks, by using her mobile computer, for the best route from home to the client's office based on up-to-date traffic conditions. Moreover, she also requests the client's record. Our sales person stops in a café before arriving at the client's office, and while she has a coffee she revises the most important data related to the client in order to prepare the interview. Once the interview has been concluded, she enlarges the client's record, registering that day's visit and its results, and sending the new data to the DB of the company. During the rest of the day our sales person goes on in the same way visiting clients and revising the emails and messages that she receives. When she is about to conclude her working day she receives a broadcasted message sent from the client assistant director to all the sales persons. It is a report about the results of the previous month and the expectations for the month in course.

The second scenario considers the case of an archeologist, but in general it could be applied to any autonomous worker. To a user like this one, to have a mobile computer can be very interesting because it gives him the possibility of having a small computer, even with limited capacities, during his field work. The archaeologist of our example needs access to Internet, to be able to access the big data repositories of the universities, libraries, etc., which store the data that he needs for his research. Let us suppose one day of this archeologist life. He gets up early; a day of fieldwork awaits him in the current excavation. In the excavation they are working on remains of prehistoric tools. While he has his breakfast he switches on his mobile computer. Then, he sends queries to obtain the information stored in a series of databases about certain types of prehistoric tools, periods to which they correspond, areas in which they were located, information about the people that used them and about those that made them, etc. He will need that information when he arrives at the excavation so he finishes his breakfast and before driving to the excavation he examines the weather report provided by the computer. Later on, when he arrives at the excavation, he goes to the place where he works and once there he switches the mobile computer on again. The requested information has arrived. The archaeologist begins his work with the found samples. He consults the stored data, formulates new queries to the databases, etc., until he identifies the period to which the samples found belong. He also obtains information about that period and their people, uses that they gave to those tools, cases of similar tools found or of similar discoveries of the same period or others.

Finally, in the third scenario is considered the case of a user equipped with a palmtop (an equipment with limited capabilities). Usually she only uses the mobile computer to register data but today, using some specialized keys, she asks for information about the films that are going to be playing tonight in the city. The information that she receives is only textual; all the associated multimedia information has been eliminated taking into account the limitations of the mobile computer.

In summary we can conclude that mobile computing should allow the users to make use of all the advantages that fixed computers provide but in a mobile environment.

3. Architecture

In the widely accepted architecture for mobile computing (see figure 1) [16] the following elements can be distinguished:

- **Mobile Unit** (MU): portable computer equipped with a wireless interface.
- **Base Station** (BS): Fixed host augmented with wireless interface. It is also called Mobile Support Station (MSS). The geographical area covered by each Base Station is called a *cell*. Mobile Units communicate with other units through those Base Stations of the cells in which they reside.
- **Fixed Host** (FH): computer without a wireless interface.

Figure 1: Mobile Computing environment.

Cell sizes vary widely, from 400 miles in diameter (covered by satellites), to a few miles (covered by cellular transceivers) or to a building (covered by a wireless LAN). When mobile units move, they can cross the boundary of a cell, and enter in an area covered by a distinct Base Station. This process is called *handoff*. Taking into account that mobile units can be disconnected, a mobile unit can abandon a cell and can appear in another one faraway. That is, movements among cells are not necessarily among adjacent cells.

Previous architecture must support different kinds of mobile units such as *palmtops* and *laptops*. Because the former ones provide less functionalities than the latter ones the features of mobile units must be taken into consideration when implementing the architecture.

With respect to the wireless part of the architecture, there exists a lack of standards and limited performance features with today's second digital generation of mobile communications systems. However, the third generation of mobile communication systems is emerging. It is formed by systems like the European Universal Mobile Telecommunication Systems (UMTS) and the international Future Public Land Mobile Telecommunications in the year 2000 (FPLMTS/IMT 2000). These systems have the goal of providing services and capabilities on the same level as fixed networks, making those services globally available independent of the user's current location, so there will be a strict distinction between network and service providers.

4. Technological Aspects: Wireless networks

Nowadays, there are different wireless access technologies that may connect mobile users to wired networks: Analog Cellular Networks, Digital Cellular Networks, Wireless Wide-Area Networks, Wireless Local-Area Networks and Satellite Systems. With these wireless access technologies it can be possible to use circuit switch or packet switch communications with important implications in money spent and speed obtained.

Analog Cellular Networks

The first generation of analog cellular systems is called AMPS (Advanced Mobile Phone Service). It is still used for cellular telephone technology and utilizes analog frequency modulation (FM) for speech transmission. The technique Frequency Division / Multiple Access (FDMA) is used to make individual calls. The bandwidth is divided into different channels and neighboring cells use different channels controlled by Mobile Support Stations.

Digital Cellular Networks

The second generation of cellular systems uses digital modulation instead of analog techniques and, apart from voice services, it can offer ISDN (Integrated Services Digital Network) services. Although there are several advantages of using digital cellular communications such as error corrections, intelligence of the digital network, integration with wired digital networks, encrypted communications, etc., the effective date rate is low (ranging from 9 to 14 Kbps). There are two basic techniques for sharing the digital cellular network: TDMA (Time Division Multiple Access) and CDMA (Code Division Multiple Access). There are several basic standards deployed in Europe and USA such as: the Global System for Mobile Communications (GSM) in Europe, based on TDMA; the IS-54 standard in the USA, based on TDMA, and the IS-95 standard in the USA, based on CDMA. There are also other cordless telephony technologies, that are limited to short ranges, like the British Second Generation Cordless Telephone (CT2), based on FDMA, and the Digital European Cordless Telephone (DECT), based on TDMA.

Wireless Wide-Area Networks

As the cellular products previously mentioned are relatively expensive and slow, some other technologies are being deployed that are based on packet switching instead of circuit switching. For example, the ARDIS (Advanced Radio Data Information System) that provides a data transmission rate from 8 Kbs to 19.2 Kbs, the RAM Mobile Data providing a data rate of 8 Kbs, the CDPD (Cellular Digital Packet Data) that provides data services on top of the AMPS analog system with a maximum data transmission rate of 19.2 Kbs but with an effective rate of 9.6 Kbs, and the GPRS that is a product that is being developed by the GSM consortium to include packet-switching with higher expected data rates than the previous ones. In general, with all these technologies the transmission rates are not very high.

Wireless Local-Area Networks

Wireless LANs provide higher data rates (more than 1 Mbs) to mobile users that have less possibilities of mobility (inside a building, campus, etc.). Some of these products try to provide wireless Ethernet connections and use different link technologies like radio frequencies, infrareds, microwaves, etc. Examples of products are FreePort, WaveLAN and Altair, and there are also standards, like the IEEE 802.11 and the HiperLAN, that are being developed.

Satellite Networks

Mobile satellite networks allow global coverage for two-way voice communications but limited data capabilities. Data rates and propagation times depend on the type of satellites used. Geostationary satellites (GEOS) provide global coverage with few but expensive stations and with great delays when establishing communications that require a high power cost. Low earth orbit satellites (LEOS) are smaller and less expensive and communications have low cost but also low data rate. The cells are much smaller and allow for more frequency reuse but imply more handoffs. Medium earth orbit satellites (MEOS) can represent a trade-off among GEOS and LEOS.

The future

There are some trends known as Personal Communication Service (PCS) in the USA and the Universal Mobile Telecommunication Systems (UMTS) in Europe that may lead to a new generation of mobile communications. They try to define and develop personal communication systems with global coverage and integration with broadband public networks.

5. Specific issues that must be taken into consideration

Three main features of the new context: mobility, wireless medium and portability of mobile elements, require special consideration [17-18]

a) Mobility

The location of mobile units is an important parameter when locating a mobile station that may hold the required data and when selecting information especially for location dependent information services. But the search cost to locate mobile units is added to the cost of each communication involving them. Two solutions have been discussed in the literature [16] for the first problem. In one solution, each mobile unit has a *home* base station that keeps track of its location by receiving notifications of its movements. The second solution is based on restricted broadcast within the area of the mobile unit that wants to access.

b) Wireless Medium

The following are some specific features of the wireless medium: the scarce bandwidth, asymmetry in the communications and the high frequency of disconnections.

The scarce bandwidth of wireless networks and asymmetry in the communications

On the one hand, wireless networks offer a smaller bandwidth that wired networks; the first ones offer a bandwidth that varies between 9 and 14 Kbps. while any Ethernet offers a bandwidth of 10 Mbps. Moreover, the oscillation in the bandwidth is more noticeable that in the traditional wired networks. On the other hand, there is an asymmetry in the communication because the bandwidth for the downlink communication (from servers to clients) is much greater than the bandwidth for the uplink communication (from clients to servers). In order to resolve the previous constraints one approach consists of the use of broadcasting to disseminate information of general interest. The use of broadcast, on the one hand, saves bandwidth by reducing the need of point-to-point communications among the mobile units and the Base Stations, and on the other hand, saves battery power of mobile units because it reduces uplink communications, which are more expensive in terms of energy than downlink communications.

Disconnections

The wireless communications are very susceptible to suffering disconnections, so this is a very important aspect to keep in mind when an architecture to support mobile computing is designed. The disconnections can be classified in two types: a) Forced disconnections: these disconnections are usually accidental and unavoidable, like for example those disconnections that take place when entering an out-of-coverage area. b) Voluntary disconnections: these take place when the user decides to disconnect his unit with the goal of saving energy. Other types of voluntary disconnections can take place when abrupt changes in the signal are detected or when the power level in the batteries is low, etc. In these cases the unit can take the necessary measures to change to disconnected mode in a stable way and without risks of loosing data. The undesirable effects of the disconnections can be mitigated using caching techniques [19-20]. Using these techniques the user will be able to continue working even in disconnection states. This will also contribute to avoiding the unnecessary use of the wireless communication.

c) Portability of Mobile Elements

Although mobile computers exist that present different capabilities, in general their limitations are mostly related to their size and battery life.

Limitations on the size and capabilities of mobile computers

The design of portable computers implies that they must be small, light, consume little energy, etc. This causes these computers to have generally more limited functionalities than fixed hosts, mainly in aspects such as computation power, storage capacity, screen size and graphic resolution, autonomy, etc. Among the solutions that try to overcome these limitations can be highlighted those that adapt the images for their visualization by reducing their size, definition or colors, by using filters. Filters are programs used for processing every message coming or going to the mobile units. They can abolish, delay, reorganize into segments or transform the message [21].

Battery power limitation

Because of the limited autonomy of the batteries, to optimize the energy consumption is generally a critical aspect in mobile computing. Even with the new advances in battery technology, the typical lifetime of a battery is only a few hours. This problem is not likely to disappear in the near future. The use of asynchronous models allows the disconnection of the portable computer of the network while their requests are heeded in the server of the fixed network, and so these units can be in doze mode, which is an energy saving mode.

6. Impact of mobile computing in the area of data management

This section explains the impact that the previous issues of mobility, scarce bandwidth, disconnections, limitations on the size and battery power limitation, have in data management. In particular in subjects like transactions, data dissemination, query processing, caching and database interfaces.

6.1. Transactions

As it is accepted in the database community, all transactions must satisfy the ACID properties. So they must be *Atomic* (all actions performed or none of them), *Consistent* (database must be left in a consistent state), *Isolated* (a transaction does not read intermediate results made by other transactions) and *Durable* (results must remain after the transactions commit). Moreover, the schedule in which different concurrent transactions are performed has to be *serializable*. This is enforced by the concurrency control methods, like *two-phase locking* or *timestamping* methods, implemented by the DBMS. However, the problem is more complicated in a distributed context where different protocols like the *two-phase* and *three-phase commit* have been defined in order to ensure ACID properties for transactions performed in different computers (see [22]).

Mobile transactions are, in general, distributed transactions where some actions are performed in mobile computers and others in fixed hosts. In this case, the previous properties: atomicity, consistency, isolation and durability are hard to enforce, especially when the mobile computers are disconnected. In this case, techniques like *two-phase locking* and *two-phase commit* may seriously affect the availability of the database system. For example, when a disconnected mobile computer owns a *lock* over database items or if other computers are waiting for the mobile computer to know if it is ready to perform a *commit* or not. Therefore, it is necessary to provide transaction support for mobile environments and/or to define some notions of *different kinds of transactions*.

Some proposals work with the notion of *weak transactions* that are those that read or write local and probably inconsistent data [23]. Other proposals present mechanisms so that applications have views of the databases consistent with their own actions, what is known as *session guarantees* (see [24]). Others present notions of transactions based on *escrow* methods, which are especially interesting in sales transactions. The total number of available items is distributed in the different sites and local transactions can commit if the demand does not exceed the quantity in the local site (see [25]). In [26] *isolation only transactions* are proposed where the rest of ACID properties not. In [27] a technique is explained where the broadcast channel is used so that the mobile clients know if they have to abort the transactions that are running. In order to do that, *Certification Reports* are sent through the broadcast channel. These reports contain items over which commits are going to be made. Another related idea called *transaction proxies* is presented in [28] that consists of defining dual transactions (one for each transaction performed in a mobile host) that will be executed in a fixed host that acts as the host support of the mobile one. These dual transactions contain only the updates made by the mobile transaction in the case that recovery were needed.

6.2. Data dissemination by using broadcasting

The feature of asymmetry in the communications along with the feature of power limitation that are present in the mobile computing framework make the model of *broadcasting* data to the clients an interesting alternative. Broadcasting consists of the delivery of data from a server to a large set of clients (sometimes it is also called *push-based*). By pushing data, the server avoids interruptions caused by requests of clients and so optimizes the use of the bandwidth in the upstream direction.

The main aspects that the *broadcasting* system must decide are the following ones: the clients' needs and whether to send the data periodically or aperiodically. Periodic push has the advantage of allowing clients to disconnect for certain periods and still not miss out items. In [29,30] there appears the use of a periodic dissemination architecture in the context of mobile systems. Aperiodic dissemination, on the other hand, is a more effective way of using the bandwidth available. In [31,32] they work with the concept of *"Indexing on Air"* i.e., transmitting an index along with the data, so clients can tune only during the times they need to. One issue that arises in the former approach is how the index is multiplexed with the data to make the latency and tuning time minimal.

Since no broadcast program can perfectly match the needs of an individual client, mechanisms have been defined to compensate the existing mismatches. One of them consists of doing intelligent caching and prefetching at the client side. These have been studied in [29,30]. Furthermore, in [19] the authors present a way of sending invalidation messages over a limited bandwidth network. With these messages the server can notify the clients about changes in the items that are being cached by them. The issue of relaxing consistency of the caches is also addressed.

Alternatively, broadcasting can be achieved by using multicast addresses. The server sends data to a group of clients using the same address. Hashing can be efficiently used in combination with multicast addresses.

Moreover, there also exist works that integrate the pull-based and the push-based approaches. In a pull-based operation, clients explicitly request items by sending messages to the server which in turn send the information back to the clients. When dealing with both approaches two independent channels – front and backchannels – are used. The frontchannel is used for push-based operation, while the backchannel serves as the media for the pull-based operations. The available bandwidth is shared between these two channels. Finally, some recent applications of data dissemination include information dissemination in the Internet [33] and in private networks [34], and dissemination using satellite networks [35].

6.3. Query Processing

Query processing is affected when mobility is considered and it is possible to formulate *location dependent queries*. For example: “where is the nearest gas station?” or “which are the cinemas that project some film at 8:00pm in this city?” return different values depending on the location of the mobile computer.

Query optimization methods try in general to obtain execution plans which minimize CPU, input/output and communication costs. In centralized environments the cost that affects most is the input/output whereas in distributed environments, communication cost is the most important one but the two others may also be important if communication costs are not very high (for example in local area networks) [36]. In a mobile distributed environment, the communication costs are much more difficult to estimate because the mobile host may be situated in different locations. The best site from which to access data depends on where the mobile computer is located. In general, it is not worth calculating plans and their associated costs statically, but rather, dynamic optimization strategies are required in this mobile distributed context.

Among the works related to query processing in mobile computing we can mention the following ones [37,38,39,40]. In [37] they present how to deal with queries with location constraints, that is, queries that involve the individual location of users. Because the location of users is not exact, it is expensive, in terms of communication costs, to find out the missing information necessary to answer queries with location constraints. The rest of the mentioned works try to provide solutions to more specific problems. In [38,39] they try to facilitate travelling by providing updated information on traffic conditions, weather, available parking, shortest distances, emergency services, etc. Needed data can be obtained by making specific requests (*pull* based) and by data dissemination or broadcasting (*pushed* based). In [40] a Web information system for a mobile wireless computing environment is presented. The Web is extended in order to allow documents to refer and react to changing contextual information, like current location in a wireless network. It introduces the concept of *dynamic URLs* (which allow to return different documents or execute different commands depending on dynamic environment variables), and also the concept of active documents that automatically update their contents in response to changes in the user's mobile context.

6.4. Caching

As said in the previous subsection, *query optimization* methods try to minimize CPU, Input/Output and communication costs, and in particular in the mobile distributed context, the communication costs may be important. It is accepted that applying *caching* techniques for query processing can reduce communication costs dramatically. However, it is more difficult to apply caching techniques in this mobile context because cache contents may change rapidly or get out-of-date due to mobility; in addition, updates to the cache memory may not be sent due to disconnections of the mobile unit.

In [18] they present several techniques related to caching such as *prefetching*, *replacement* strategies and *consistency* of the cache memory used in combination with *broadcasting* techniques. The idea is to send by broadcast channel some data that may be needed in the future (*prefetching*) and/or data that have become *inconsistent*.

6.5. Database Interfaces

The limited screen sizes of many mobile computers has motivated the development of new interfaces for them and, in particular, the design of new database interfaces for mobile computers. In [41] there appears a query processing interface called Query By Icons that addresses the features of screen size along with the limitations in memory and battery power and the restricted communication bandwidth. In [42] the issue of how the pen and voice can be used as substitutes for the mouse and keyboard is addressed. Moreover, in [43] there appears an implementation of a pen-based graphical database interface on a pen computer.

7. Communication Models and Agents

In this section we present some other issues that also have to be considered in order to build systems that allow accessing services from mobile computers.

7.1. Communication Models

Two main types of models are being used in the mobile computing environment [18]: the client/server model (in its different versions) and the peer-to-peer model. An important difference between them resides in the role that each element of the environment plays. In the former case, the mobile unit – client – requests a service from another computing system – server – located at the fixed network, while in the latter case, there is no distinction between clients and servers. Each site (ideally) has the full functionality of both a client and a server. Although there are certain applications for which the peer-to-peer model is adequate (e.g. two partners performing cooperative work on the same data using their portable computers) the client/server model is more broadly used.

The traditional client/server model presents some shortcomings when dealing with wireless networks because these kinds of networks present a high rate of errors, limited and variable bandwidth and continuous disconnections. For this reason the following client/server extensions have been proposed: the *client/agent/server* model and the *client/intercept/server* (both can be grouped under what is called the indirect model [44]). The basic idea of the indirect model is: whenever the interaction between two computers takes place over two radically different media, like wire and wireless, their interaction is broken down into two phases, one for each kind of media. An intermediary element is placed in one point between the two computers. That element manages the interaction between the computers, taking into account the different nature of the two media involved, so it tries to relieve the more limited extreme of the communication of some tasks; but its existence can even remain unnoticed by the two computers. More particularly, the client/agent/server model alleviates the impact of the limited bandwidth and the poor reliability of the wireless link by continuously maintaining clients presence on the fixed network via an agent. The agent splits the interaction between mobile clients and fixed servers in two parts, one between the client and the agent, and one between the agent and the server. This model moves responsibilities from the client to the agent. Moreover, agents are used in a variety of forms and roles. At one extreme, an agent acts as the complete surrogate of a mobile host on the fixed network. At the other extreme, the agent is attached to a specific service or application.

Furthermore, used to address the shortcomings of the client/agent/server model is the client/intercept/server model [45]. In the client/agent/server model the mobile client cannot continue to operate uninterrupted when an event such a disconnection happens. That model requires changes to the client code and the agent can optimize only data transmission over the wireless link from the fixed network to the mobile client and not vice versa. The client/intercept/server model proposes the use of two agents, the client-side agent who is co-resident with the client and the server-side agent who resides on the fixed networks. This model is transparent to both, the client and the server, offers flexibility in handling disconnections and optimizes data transmissions between fixed network to the mobile client and vice versa.

7.2. Agents

Agent technology is not new in Computer Science [46]. It has been used, for example, on Artificial Intelligence. In general, an agent is a computer program that acts autonomously on behalf of a person or organization. Each agent has its own thread of execution so that it can perform tasks on its own initiative. An agent system is a platform that can create, interpret, execute, transfer and terminate agents. When an agent moves, it travels between execution environments called places. A place is a context, within an agent system, where an agent can execute. The source place and the destination place can reside in the same agent system, or in different agent systems that support the same agent profile [47].

The use of the agent technology when implementing mobile systems provides one with the following advantages [48]:

- *Asynchronous communications*. The elements involved in the communication do not have to be connected all the time. That means that mobile units may decide to disconnect while the agents that represent them, are working in other computers. This may be interesting, for example, when accessing databases where transactions can take a long time.

- *Autonomous communications*. This means that agents may take some decisions on behalf of the user when representing the mobile units. This may be interesting, for example, when accessing databases and some transactions fail; in such case, the agent can take the decision of retrying the transaction or not, trying another one and so on, considering the knowledge that it has about the mobile unit.
- *Remote communications*. This means that agents can make use of remote facilities or resources such as memory, CPU, etc. This may be interesting, for example, when a mobile unit has not enough capacities to develop a task; in such case an agent can realize the task in a remote computer and once finished returns the results to the mobile computer.

Recently there exists a great research effort with respect to the mobility feature of agents. A mobile agent is not bound to the system where it begins execution; it has the unique ability to move from one system in a network to another. The ability to travel lets a mobile agent move to a system that contains an object with which the agents wants to interact and take advantage of being in the same host or network as the object. When an agent travels, it transports its state and the code with it. Mobile agent technology, apart from the previous advantages that we have mentioned for implementing mobile systems using agents, allows one to migrate processes among different machines.

8. Design Features for Systems that allow Mobile Computers to access Data Services

The number of people who use or work with mobile computers is continuously increasing. Although the performance features of those kinds of computers when working disconnected (for example in the case of laptop) are equivalent to those offered by fixed computers, when working connected to a wireless network the same does not occur. The intrinsic features of wireless communications – poor quality, limited bandwidth, continuous disconnections – make working connected to a wireless network more difficult. However, one important wish of mobile users is to have the possibility of working connected to a wireless network in the same way as working connected to a fixed network or with at least better quality of service than that offered by existing networks.

Different research projects consider that wish and try to build mobile systems that overcome the existing limitations [49-54]. All the previous works consider different aspects of mobile computing by using agent technology. In the same line, we present in [55] a system based on the use of the *client/intercept/server* model which incorporates some modules and agents in the mobile computer as well as in an intermediary element situated in the fixed network. That element, called *Gateway Support Node (GSN),* (see figure 1) is the intermediary element in the communication between the mobile computers under its coverage and all other hosts of the network (mobile or fixed). Its aim is to relieve mobile computers from many tasks and increase their capabilities, respecting, at the same, time their natural limitations and taking into consideration the problems of the mobile computing framework and trying to solve them. The pair formed by the GSN and the Mobile Unit (MU) allows the MU to behave like a fixed computer for the rest of the network. The GSN lends its identity to the set of mobile computers that it monitors, so when the GSN receives messages and data sent to the mobile computers, it distributes them to the suitable MU.

Concerning the use of agents, in the mentioned works the process consists of creating an agent for each task to be carried out, giving it the data necessary to access a certain source of information and sending it from the mobile computer to the network. Once the results are obtained, the agent returns to the mobile computer. In our proposal the underlying philosophy is different. We advocate using a majordomo agent *Alfred* in order to avoid the continuous transferences of agents through the wireless link and, therefore, the high cost that it represents. Alfred is an efficient majordomo for mobile computers. Each mobile computer will have its own version of Alfred with the aim of giving adequate services to its owner. From the implementation point of view, Alfred is the union of two agents: SAlfred (Static Alfred) – a static agent situated in the mobile computer –, and MAlfred (Mobile Alfred) – a mobile agent situated in the intermediary element –. MAlfred is created in the mobile computer, but it travels to the intermediary element, where it works on behalf of the mobile user, representing him in the network, becoming the common point to all the communications in which the mobile unit is involved, even when the mobile computer is disconnected. When a task must be carried out, SAlfred sends a message to MAlfred with the necessary data. Then MAlfred carries out the task or creates a new agent, a specialist (the specialist mobile agents are situated in the GSN), and orders it to carry it out. Once the results are obtained, MAlfred sends them with a message to SAlfred.

The system that we propose can be presented from two different points of view, depending on where the GSNs are situated and who owns them. On the one hand, the case of a GSN owned by the company that is offering the wireless communication infrastructure, for example a cellular phone company can be considered. On this model the GSN can offer some services for the general use of mobile users that contract them to the company. Among those services can be found (see figure 2): *Broadcast transmitter* for disseminating general interest information such as local traffic condition, weather forecast; *Yellow pages* for providing access to different data repositories that contain general interest information such as local restaurants; *Access to Internet a la Carte* for facilitating the use of the *push* technology to the mobile users; *Available software* such as freeware software, that can be used on the mobile computer or on the GSN on behalf of the mobile user; *Rent of lockers* that allows the mobile user to have services such as confidential access protected with a secret key to a locker.

On the other hand, the GSN can be situated in a computer that is in charge of monitoring the access to a private corporate network. That is, the GSN is part of the Intranet of the company and the offered services can be customized according to the needs and characteristics of the particular company and of its mobile workers. For example, the GSN could offer the following services: *Access to data repositories* for allowing the users to perform queries and updates in any database server of their organization; *Access to fixed hosts* for providing access to information stored in fixed hosts and allowing one to get and store files in any host of the fixed network where the user has the right access privileges; *Access to WWW* in order to obtain WWW pages, caching them in the GSN and sending them to the mobile users; *E-mail* for allowing mobile users to get and send mails to any mail server; *Blackboard* for allowing storage of messages of general interest that the users can obtain in several ways.

Figure 2: Mobile Computing environment with GSN elements and mobile agents.

9. Conclusions

In this chapter we have made a brief revision of the main issues concerning data management in mobile computing. We have illustrated the possibilities that this new paradigm offers and the widely accepted architecture, followed by the technologies that are considered. Moreover, we have introduced the main features of the context, that in fact have a great influence on the performance of the mobile systems, focusing on the data management aspects. In summary we can conclude that mobile computing opens new expectations for data applications. However, mobile computing is not mature yet and many problems must be solved, so it is expected that new proposals will appear in the future.

10. Further Reading

The concepts presented in this chapter can be further studied by reading the following two books and article. The books are: *Mobile Computing* edited by Tomasz Imielinski and Henry F. Korth [17] and *Data Management for Mobile Computing* by Evaggelia Pitoura and George Samaras [18], and the article is *Mobile Computing and Databases – A Survey* by Daniel Barbará [56]. The first one starts with a good introduction to Mobile Computing and then it presents a set of projects and systems that study the new problems related to mobile computing in different computing areas: networking, operating systems and information systems. In relation to networking it studies the mobility management, ad-hoc networking protocol and transport-layer issues. With respect to operating systems it studies operating system support for mobile computing, energy-efficient CPU scheduling, storage alternatives and disconnected operations in file systems. And in relation to information systems, it presents the problems of wireless access to information, application design, etc. The second book is more centrated on the data management issues in mobile computing, although it also has a chapter about system-level support. After another good introduction to Mobile Computing it presents different software architectures and techniques for information and location management (broadcasting, caching, replication, etc.). It finishes with some interesting case studies. The survey is focused on research in the area of data management in mobile computing and different techniques such as data dissemination over limited bandwidth channels, location-dependent querying of data, advanced interfaces for mobile computers and techniques to maintain data consistency.

References

[1] J.J. Kistler and M. Satyanarayanan. "Disconnected operation in the Coda file system." *ACM Transactions on Computer Systems*. 10:213–225, 1992.

[2] M.S. Mazer and J.J. Tardo. "A client-side-only approach to disconnected file access." *Proceedings of the IEEE Workshop on Mobile Computing Systems and Applications*. December 1994.

[3] J. Heidemann et al. "Primarily disconnected operation: Experience with Ficus." *Second Workshop on Management of Replicated Data*, November 1992.

[4] AirAccess. *http://www.airsoft.com/*

[5] J.H. Howard. "Using reconciliation to share files between occasionally connected computers." *Fourth Workshop on Workstation Operating Systems*, October 1994.

[6] Traveling~Software Laplink. *http://www.travsoft.com/*

[7] A. Demers et al. "The Bayou architecture: Support for data sharing among mobile users." *Proceedings of the IEEE Workshop on Mobile Computing Systems and Applications*, December 1995.

[8] J.R. LoVerso and M.S. Mazer. "Caubweb: Detaching the web with tcl." *Proceedings of the Fifth Annual USENIX Tcl/Tk Workshop*, July 1997.

[9] B.N. Schilit et~al. "TeleWeb: Loosely connected access to the World Wide Web." *Proceedings of the Fifth International World Wide Web Conference*, May 1996.

[10] Lotus Weblicator. *http://www2.lotus.com/weblicator.nsf.*

[11] A. Fox and E.A. Brewer. "Reducing WWW latency and bandwidth requirements by real-time distillation." *Proceedings of the Fifth International World Wide Web Conference*, May 1996.

[12] T.W. Bickmore and B.N. Schilit. "Digestor: Device-independent access to the World-Wide-Web." *Proceedings of the Sixth International World Wide Web Conference*, April 1997.

[13] B.C. Housel and D.B. Lindquist. "WebExpress: A system for optimizing web browsing in a wireless environment." *Proceedings of the ACM/IEEE MOBICOM '96 Conference*, October 1996.

[14] A.D. Joseph, J.A. Tauber, and M.F. Kaashoek. "Mobile computing with the Rover toolkit." *IEEE Transactions on Computers*, 46(3), November 1997.

[15] G. Welling and B.R. Badrinath. "Mobjects: Programming support for environment directed application policies in mobile computing." *Proceedings of ECOOP'95 Workshop on Mobility and Replication*, 1995.

[16] T. Imielinski and B.R. Badrinath. "Mobile wireless computing: Challenges in data management." *Communications of the ACM*, pages 19–27, October 1994.

[17] T. Imielinski and H.F. Korth. *Mobile Computing*. Kluwer Academic Publishers, 1996.

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

[19] D. Barbará and T. Imielinski. "Sleepers and workaholics: Caching strategies in mobile environments (extended version)." *The VLDB Journal*, 4(4), 1995.

[20] A. Si and HV. Leong. "Adaptive caching and refreshing in mobile databases." *Personal Technologies*, September 1997.

[21] B. Zenel and D. Duchamp. "A general purpose proxy filtering mechanism applied to the mobile environment." *Proceedings of the Third Annual ACM/IEEE International Conference on Mobile Computing and Networking*, September 1997.

[22] R. Elmasri and S.B. Navathe. *Fundamentals of Database Systems*. Addison-Wesley, 1999. 3rd Edition.

[23]. E. Pitoura and B. Barghava. "Building Information Systems for Mobile Environments." *Proceedings of the Third International Conference on Information and Knowledge Management*, Washington DC, November 1994.

[24]. D. Terry, A.J. Demers, K. Petersen, M.J. Spreitzer, M.M. Theimer and B.B. Welch. "Session Guarantees for Weakly Consistent Replicated Data." *Proceedings of the Conference Parallel and Distributed Computing*, Austin Texas, Octuber 1994.

[25] D. Barbará and H. García-Molina. "The Demarcation Protocol: A Tecnique for Maintaining Constraints in Distributed Database Systems." *Proceedings of ACM SIGMOD International Conference on Management of Data*, S. José, California, May 1995.

[26] Q. Lu and M. Satyanarayanan. "Improving Data Consistency in Mobile Computing Using Isolation-Only Transactions." *Proceedings of the Fifth IEEE HotOS Topics Workshop*, Orcas Island, May 1995.

[27] D. Barbará. "Certification Reports: Supporting Transactions in Wireless Systems." *Proceedings of the 17th International Conference on Distributed Computing Systems*, Baltimore, May 1997.

[28]. E. Pitoura and B. Barghava. "Revising Transaction Concepts for Mobile Computing." *Proceedings of the First IEEE Workshop on Mobile Computing Systems and Applications*, pages 164--168, December 1994.

[29] S. Acharya, M. Franklin, S. Zdonik and R. Alonso. "Broadcast Disks: Data Management for Asymmetric Communications Environments." *Proceedings of ACM SIGMOD International Conference on Management of Data*, S. José, California, May 1995.

[30] S. Acharya, M. Franklin and S. Zdonik. "Dissemination-Based Data Delivery Using Broadcast Disks." *Personal Communications, 2(6),* December 1995.

[31] T. Imielinski, S. Viswanathan, and B.R. Badrinath. "Data on air: Organization and access." *IEEE Transactions on Knowledge and Data Engineering*, 9(3), 1997.

[32] H.V. Leong and A. Si. "Database caching over the air-storage." *The Computer Journal,* 40(7) 401-415, 1997.

[33] Hughes Network Systems DirectPC. *http://www.direcpc.com*

[34] D. Glance. "Multicast Support for Data Dissemination in Orbix-Talk." *IEEE Data Engineering Bulletin, 19(3).* September 1996.

[35] S. Dao and B. Perry. "Information Dissemination in Hybrid Satellite/Terrestrial Networks." *IEEE Data Engineering Bulletin, 19(3).* September 1996.

[36]. M. T. Özsu and P. Valduriez. "*Principles of Distributed Database Systems*". 2nd Edition. Prentice-Hall, 1999.

[37] T. Imielinski and B.R. Badrinath. "Querying in Highly Mobile and Distributed Environments." *Proceedings of the 18th International Conference on Very Large Data Bases,* Vancouver, B.C., Canada, August 1992.

[38] S. Shekhar, A. Fetterer and D. Lui. "Genesis: An Approach to Data Dissemination in Advanced Traveler Information Systems." *IEEE Data Engineering Bulletin, 19(3).* September 1996.

[39] S. Shekhar and D. Lui. "Genesis and Advanced Traveler Information Systems." In *Mobile Computing, T. Imielinski and H. Korth, eds. Kluwer, pages 699-723, 1996*

[40] G.M. Voelker and B.N. Bershad. "MOBISAIC: An Information System for A Mobile Wireless Computing Environment." In *Mobile Computing, T. Imielinski and H. Korth, eds. Kluwer, pages 375-395, 1996*

[41] A. Massari, S. Weissman and P.K. Chrysanthis. "Supporting Mobile Database Access Through Query By Icons." *Distributed and Parallel Databases, 4(3), pages 47-68,* July 1996

[42] M.T. Le, S. Seshan, F. Burghardt and J. Rabaey. "Software Architecture of the InfoPad System." *Proceedings of MOBIDATA Conference.* Rutgers University, October 1994.

[43] R. Alonso and V.S. Mani. “A Pen-Based Database Interface for Mobile Computers.” *Proceedings of the First IEEE Workshop on Mobile Computing Systems and Applications.* Santa Cruz, California, *4(3),* September 1994

[44] B.R. Badrinath et al. "Handling mobile clients: A case for indirect interaction." *4th Workshop on Workstations Operating Systems*, 1993.

[45] G. Samaras and A. Pitsillides. "Client/intercept: a computational model for wireless environments." *Proceedings of the 4th International Conference on Telecommunications (ICT'97)*, April 1997.

[46] Pattie Maes. "On software agents: Humanizing the global computer." *IEEE Internet Computing*, July-August 1997.

[47] Object Management Group. Mobile Agent System Interoperability Facilities Specification (MASIF), November 1997. *http://www.camb.opengroup.org/RI/MAF/.*

[48] C. Harrison, D. Chess, and A. Kershenbaum. "Mobile agents: are they a good idea?" *Mobile Object Systems: Towards the Programmable Internet*, pages 46–48, 1997.

[49] S. Papastavrou, G. Samaras, and E. Pitoura. "Mobile agents for WWW distributed database access." *Proceedings of the International Conference on Data Engineering,* 1999.

[50] E. Kovacs and K. Röhrle and M. Reich. "Mobile Agents OnTheMove – Integrating an Agent System into the Mobile Middleware." *Acts Mobile Summit. Rhodos, Greece*. June, 1998.

[51] L. Hurst and F. Somers. "Mobile agents - smart messages.” *Proceedings of the First International Workshop on Mobile Agents (MA 97).* 1997

[52] A. Sahai and C. Morin. "Mobile Agents for Enabling Mobile User aware applications.” *Proceedings of the Second International Conference ACM Autonomous Agents (Agents 98)".* 1998

[53] R. Gray and D. Rus and D. Kotz. "Agent TCL: Targeting de needs of mobile computers." IEEE Internet Computing. 1997

[54] D. Chess and B. Grosof . "Itinerant Agents for Mobile computing”. *IBM Research Report RC 20010, IBM*. 1995.

[55] Y. Villate, D. Gil, A. Goñi, and A. Illarramendi. "Mobile agents for providing mobile computers with data services." *Proceedings of the Ninth IFIP/IEEE International Workshop on Distributed Systems: Operations and Management (DSOM 98),* 1998.

[56] D. Barbará. "Mobile Computing and Databases – A Survey." *IEEE Transactions on Knowledge and Data Engineering, 11(1), 108-117.* January/February 1999.

Figure 1: Mobile Computing environment.

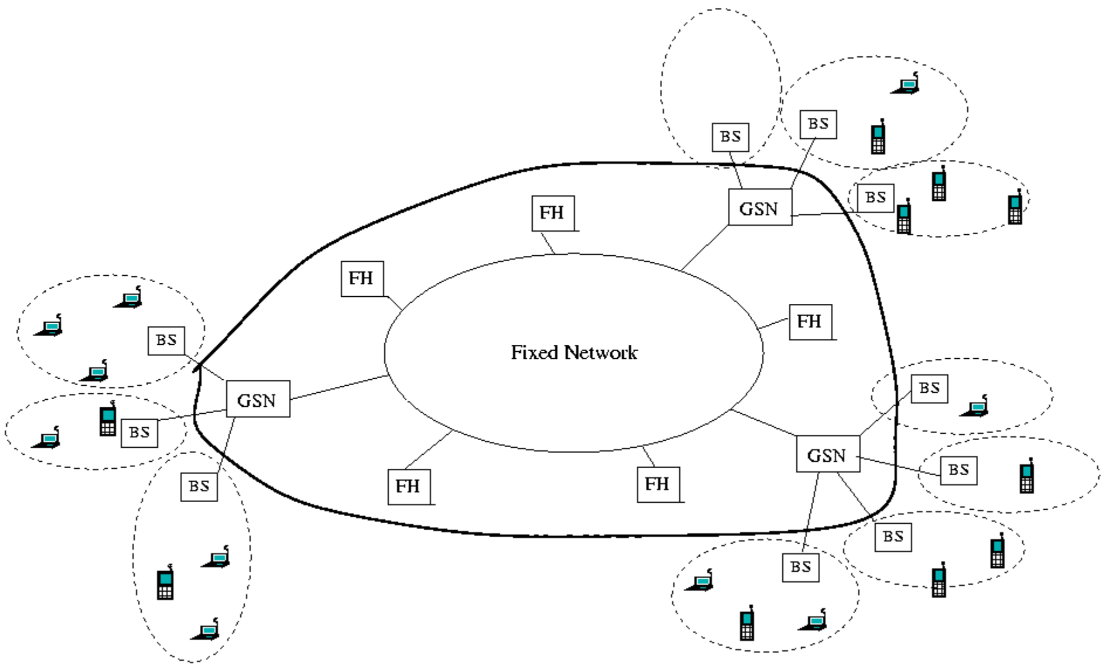

Figure 2: Mobile Computing environment with GSN elements and mobile agents.