


maintained without having to contact mobile hosts all the time.

Recovery introduces new problems due to various forms of elective failures - failures which can be anticipated and prepared such as handoff and various forms of disconnection.

6.3 Little or No Impact

Data Modeling

New techniques will be necessary for user’s profile specification. These may involve temporal, spatial and probabilistic rules to describe the mobility pattern of the user, and specifying access rights based on location and identity of both the subject and the object.

We see little impact of mobile computing on the data manipulation languages design (except of special constructs for spatial queries). Active rules and triggers can be based on location and need special syntactical constructs to handle spatial data as well.

7 Conclusions

Management of data in the massively distributed environment of mobile computing offers new challenging research problems. We have identified those challenges and formulated a number of open problems.

Data management issues offer new challenges both at the global, network level as well at the local computing platform of a palmtop computer. The scale of the system and mobility of its parts are unprecedented and the current network infrastructure is simply not capable of providing adequate support.

We have categorized new research problems into mobility, scalability, bandwidth and energy management. In general, we believe that mobility will have a similar impact on data management as distribution had on data management. The fundamental question “centralized or distributed” will now be extended to “static or mobile”.

8 Acknowledgments

Special thanks go to David Goodman for introducing us to the field of wireless information services and for his continued support and discussions. We would like to thank Arup Acharya, Rakesh Agrawal, Daniel Barbara, Phil Bohannon, Alex Borgida, Dan Duchamp, Marcin Imieliński, Witold Litwin, Ashar Mahboob, and Avi Silberschatz for comments and discussions on the subject of the paper and its presentation. Needless to say any shortcomings or flaws in the paper are the sole responsibility of the authors.

References


Being accessible at any location and at any time creates great concern about privacy issues among potential users (see for example [11] for impressions after early developments of active badge technology). To protect privacy it is necessary to develop sophisticated software for specification and enforcement of personal profiles of users. In such profiles users should be able to specify who, when and where is authorized to reach them (receiving e-mail consumes battery power of the local terminal; so one may want to restrict the list of users who are allowed to "wake up" the mobile unit to send a message to it). Management of profiles is a nontrivial issue - where are such personal profiles to be stored? Should the profile migrate with the user or follow the user?

With mobile users, the problem of authentication and security will be a global problem and distributed services to support authentication across administration domains will be necessary. An example of this service is black pages [10]. Black pages is a distributed service for public key data. Black pages will have the bindings of users and keys and can be used to provide authentication and facilitate secure communication.

**Interface Design**

*Data Entry Limitations*

New interfaces which do not rely on keyboard have to be developed in order to deal with the physical limitations of the mobile terminals. Pen based and speech based front ends are leading contenders.

*Data Output Limitations*

Due to the limitations on the size of the screen, one has to look for new, more innovative, data output methods for displaying answers to queries. New possibilities involve approximate answers (subsets of the real answer) and so called intensional answers.

### 6.2 Some impact

**Transaction Processing**

*Form Based Transactions*

Transactions submitted from mobile terminals will have much simpler, *form based* structure than the general transactions which are general purpose programs with embedded database calls. These forms can be broadcasted to a number of users, and the users will then respond by entering specific data values.

*Long Disconnections*

Since the mobile terminals will often be disconnected from the fixed part of the network transactions will often be processed locally on the *cached* data. The degree of "connectivity" of the mobile terminals to the fixed part of the network will vary widely. Mobile users will "check out" portions of the database for long periods of time. Thus, new methods of cache synchronization reflecting different degrees of connectivity will be necessary.

In general, basic concepts such as *locking* and *commit protocols* have to be redefined in the context of mobile hosts carrying shared data.

**Integrity and Recovery**

How should a constraint such as referential integrity which involves data residing on a number of mobile and fixed hosts, be efficiently maintained? Mobile hosts may be frequently disconnected and unreachable. The main question is how to partition data among static and mobile hosts such that the constraint can be
Techniques for distributed data management have been based on the assumption that the location of hosts in the distributed system does not change and the connections among hosts do not change. However, in mobile computing, these assumptions are no longer valid. Hence, distributed data management will be changed due to mobility. Resources will have to be dynamically reconfigured in response to the mobility of the clients. This involves dynamic replication of data as well as services.

**Query Processing**

*Querying wireless, broadcasted, data*

Another new problem involves querying data broadcasted by the server and addressed to a large number of clients. What is the best execution plan for a query which involves data broadcasted on different channels (streams)? What should be the organization of the broadcasted data so that the energy spent on the client’s side is minimized? Which information should be broadcasted and which should be provided “on demand”? How will contiguous\(^\text{12}\) queries [23], be evaluated using broadcasted data?

*Querying Fast Changing, Update Intensive Data*

How do we represent and query fast changing, update intensive data such as location? Precise tracking of all changes may be impossible or simply unnecessary. Instead, we may have to store incompletely specified information. Consequently, queries will be answered in an approximate way. Location management viewed as establishing locations of individual mobile users is only a special case of this problem. More complex queries involving locations include finding a number mobile objects (such as taxi cabs) in a given area or locating the nearest object (such as an ambulance).

*Scale* is a very important factor when dealing with the above two categories of queries. The benefits of data broadcast are proportional to the number of users serviced and the cost of broadcasting does not depend on the number of users. Therefore broadcasting seems to be a method particularly suitable for large number of clients. In location management, scalability of solutions is extremely important: what may work for 100 users in the local campus may not work for 10,000 users spread across the wide area network. For example, as a location management technique, informing the location server about each user’s move will not scale up beyond local area networks.

*Tariff-driven Query Optimization*

New methods for dynamic and distributed query optimization will have to be developed in order to handle different access costs (also in terms of dollars) from different locations. For example the same query may have different cost when formulated within a local area wireless network or wide area environment. Therefore, the cost of query evaluation may depend on location of the querying party. With competitive service providers, the same information may be available from different servers at different costs. Query optimizers will have to be flexible enough to deal with such varying costs.

**Security and Integrity**

Security and privacy is a major controversial issue in mobile computing. Presently, it is largely ignored which is typical for a discipline in its infancy. Here we will concentrate on the software issues and will not discuss the very important and difficult problem of security of the physical wireless channel.

*Profiles*

\(^{12}\) Queries are contiguous if we want to keep track of the value of the query in a changing environment all the time.
Table 1: Channel Assignment

<table>
<thead>
<tr>
<th>Stock data</th>
<th>Weather</th>
<th>Traffic</th>
<th>Sales</th>
<th>Advertising</th>
<th>Index Channel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ch 2</td>
<td>Ch 1</td>
<td>Ch 1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2: Wireless Broadcast

There may be various degrees of disconnection ranging from total disconnection to weak disconnection. Weak disconnection or narrow connection occurs when a terminal is connected to the rest of the network via low bandwidth (may be intermittent) wireless channel. CODA [16] is an excellent example of a file system that allows disconnected operation. A user can cache all the required files prior to disconnection, work on the locally cached files while being disconnected and later reintegrate. During reintegration, any conflicts due to updates are handled based on an optimistic concurrency control scheme [16].

Caching algorithms in the presence of frequent disconnections are discussed in [6]. It is demonstrated there that the disconnection time (sleep time) has a profound effect on cache synchronization strategies.

6 What is affected

A natural question to ask after reading the above sections is “how will all of these new issues affect my work?” So far we have specified new issues brought about by the mobile wireless computing environment. How do they affect traditional data management research areas? Below, we present briefly how specific areas will be affected. In general, we believe that mobility and portability may have as wide ranging an impact on systems design as distribution had in the past. We believe that moving from static to mobile environment will have a similar scope of influence as migrating from centralized to distributed systems. Thus, instead of asking “centralized or distributed?” we will simply ask “static or mobile?”

We have created three categories to reflect the impact of mobility on various aspects of data management: large impact, some impact, and little or no impact.

6.1 Large Impact

Distributed Data Management
is downloaded without need for prior requests. Minimizing wake up time is possible due to selective tuning
provided by the self-addressable channel.

Paraphrasing [14] we say that air holds data too and treat ether as a specific memory medium with
its own access characteristics. The point of view in which a network is viewed as memory medium was first
proposed in the DataCycle work [13].

In [8], we discuss basic methods for data organization on the broadcast channel. In order to provide
selective tuning capabilities for the clients, index information has to be broadcasted together with the data.
We organize the broadcasted data in such a way that the following two parameters are minimized:

- **Access Time**: Time elapsed from the moment a client issues a query to the moment the answer is
  received by the client.

- **Tuning Time**: Amount of time spent by the client, listening to the channel.

Access time is proportional to the overall size of the broadcasted data. Therefore, somehow paradoxically,
the presence of index increases the access time since the presence of the index increases the overall broadcast
size. However, the presence of the index dramatically reduces the tuning time (which corresponds to the
access time for disk based files). This in turn, as demonstrated in [8] significantly reduces the energy
consumption.

Data broadcasting has been used in the past as a method of information dissemination over wireless as
well as fixed networks. Gifford in [12] describes design of Boston Community Information System (BCIS)
which uses FM Subsidiary Communications Authority (SCA) channel. The DataCycle project at Bellcore
[13] utilizes broadcasting over a very fast fixed network in order to disseminate a large database to a large
number of users. Bandwidth, in this case, is a much less significant problem than that of wireless broadcasting
and energy is not a problem at all. Thus, filtering the wireless broadcast is a significantly different problem.

**Impact of Tariff Structures**

Mobile hosts will operate in different wireless environments ranging from the wireless LAN through cellu-
lar or special mobile radio (SMR) networks. These networks will differ widely in terms of their tariff
structures. Clients will be interested in minimizing their “data services” bills by using different implementa-
tions of the same data service by tuning them to different tariff structures.

For instance, a standard menu driven information service such as gopher, under the “pay per connection
time” tariff, may have every key stroke sent to the server. In this way, the server can respond as soon as
it recognizes a prefix of a valid command (for example, menu choice). Under “pay per packet” tariff, this
strategy may result in too many small packets (or half-empty packets) to be sent. Therefore, in this situation
we would rather buffer key strokes on the client’s side and send as much information as possible in a full
packet.

Another new factor, arising due to the bandwidth and energy limitations, is frequent disconnection
(switch off) of mobile terminals as a power saving measure\(^\text{11}\). This may involve a partial disconnection,
when the unit is in a doze mode or a total disconnection when network connection is shut down by switching
off. Voluntary disconnections can be treated as planned failures - which can be anticipated and prepared.

\(^{11}\) Keeping the palmtop in a standby (doze) mode all the time is still a considerable power drain. The energy used at this
mode could be more effectively used to increase the active time.
clients with information servers over such narrow bandwidth is one of the key challenges facing the data management research community.

The concepts of bandwidth and energy management play a crucial role in design of the future Wireless Information Services, where mobile clients will query and possibly update remote databases located on the fixed network over a wireless channel. Bandwidth limitations as well as battery power limitations will define new cost measures for accessing and updating data and consequently may favor new solutions.

Consider a scenario of a Wide Area Information service in which a large number of users are querying remote information servers over a narrow band wireless channel. For simplicity, we assume that queries refer to values of individual items stored in the server’s databases. Hence, a query like this will typically refer to the value of individual stock or to the traffic on a particular highway.

We will distinguish between two fundamental ways of providing users with information:

- **Data Broadcasting**: Periodic broadcasting of data on the channel. Accessing broadcasted data does not require uplink transmission and is “listen only.”

- **Interactive/On-Demand**: The client requests a piece of data on the uplink channel and the server responds by sending this piece of data to the client.

Analogous to the call capacity measure defined in telephony[22], we can define a query capacity as the number of queries which can be handled per unit of time by the server over a given amount of bandwidth. We are interested in maximizing the query capacity by better utilization of the wireless bandwidth. At the same time, due to the battery power restrictions we would like to minimize energy per query spent by client. Assuming that the query is processed entirely by the server, in order to minimize energy spent by client we have to minimize the number of transmissions and the total time which the client spends listening to the communication channel.

These requirements can be summarized in the form of postulates for the client/server interactions: The server should try to maximize the overall query capacity. This can be done by **periodically broadcasting most frequently requested data items**. Indeed, usually, (80/20 rule) 80% of queries involve 20% of database items; in fact in many cases the “hot spots” are even smaller\(^\text{10}\). Such hot spot information should rather be periodically broadcasted than provided “on demand” basis. The server should dynamically adjust the content of the broadcasted hot spot depending on the periodically measured demand distribution.

The client **saves energy by avoiding transmissions and waking up from the doze mode only when absolutely necessary**. Avoiding transmissions is possible if the requested data is periodically broadcasted. In such a case no real uplink request is necessary - the client simply waits until the requested information is downloaded by the server. In order to avoid continuous listening to the downlink channel, the directory information should be made available to the client. The directory has to be broadcasted by the server as well so the communication channel is entirely self described.

We may summarize these by saying that the server maintains self addressable **cache on the air** to save bandwidth. The client is **lazy** that is he transmits only when he has to and he dozes off as often as he can. Avoiding transmissions is possible by the spontaneous and periodic server’s broadcast when the data

\[^{10}\text{In case of the menu based information servers such as gophers, top level menu screens are accessed much more often than the lower level ones. In this case it makes sense to broadcast the top level menu screens rather than provide them on demand.}\]
Notice that in this scheme name resolution (directory lookup) is performed only as the last resort and (hopefully) most searches will not involve this process at all (if B is located in A’s local area).

There has been very little work so far on comparing different locating and addressing schemes. The problem is difficult, since it involves several dimensions. Solutions which are optimal in terms of numbers of messages sent may display a very poor performance in terms of latency. Also, it is not clear how detailed the statistical profiles of the users ought to be in order to provide a significant performance advantage.

4.2 Distributed systems

In general, mobility of hosts brings in a new set of issues in distributed systems. Mobility introduces the cost of search to the global cost analysis. In general, the less informed the sender is, the more the search cost incurred. Hence, mobility substantially affects data placement. Mobile hosts have severe resource constraints in terms of limited battery life and limited size of non volatile storage. Thus, the burden of computation and communication load cannot be distributed equally among static hosts and mobile hosts.

Mobility has far reaching consequences in system configuration. Mobile clients may find themselves far away from their servers; servers may also move further away from their clients. Thus, system will have to adapt to changing spatial distribution of clients by dynamic replication of data and services (data and services following users) [24]. The basic question to be addressed here is whether the resources should be static or mobile, i.e., follow the moves of the user. Thus, the normally asked question “centralized or distributed”, will now be enhanced to “static or mobile.”

4.3 Scale

The scale of the mobile environment just described will go far beyond any of the existing paradigm. Many predictions call for tens of millions of machines of varying size which can move across a worldwide communication network. Such grandiose scale affects all the issues discussed so far. In location management, the total volume of transactions due to location updates may be by an order of magnitude higher than the capacity of the existing network [20]. In system configuration management, due to frequent changes which may involve wide area moves of large number of machines, scale plays a critical role too. After all, what we have here is a distributed system with not just tens or hundreds of sites. Scale has major consequences on limited bandwidth resources: the increasing number of users requires using smaller and smaller cells. This in turn complicates the location and configuration management due to increasing number of handoffs.

Massive scale of the system also results in its heterogeneity. How do we make sure that a mobile unit sees the same or similar environment regardless of its location? How do we provide a uniform access to information services and databases across the network?

5 Bandwidth and Energy Management

Wireless bandwidth and energy in batteries will be scarce resources which will have to be carefully managed both at the hardware level and at the software level. Wireless links will provide a slow (currently up to 19.2Kb/s for CDPD) “first mile” of information superhighway. Providing a smooth interaction of mobile
4.1 Location Management

In a mobile environment, the location of a user can be regarded as a data item whose value changes with every move. Hence, location becomes a frequently changing piece of data. Given the user’s identification number the network usually establishes his or her permanent (home) address and then finds out his/her current location. The first part of the process is normally performed by name resolution in distributed directory management as exemplified by X.500 system, [5]. Distributed directory management makes it possible to avoid a full directory replication on all nodes of the network but it is designed primarily to handle permanent addresses. The second part in the process of locating the user involves finding his or her current location. How to do this efficiently is still very much an open problem.

The fundamental tradeoff in location management is between searching and informing: If A wants to establish the location of B; should A search the whole network or should A only look at pre-defined locations? Should B inform anybody about his or her moves?

One such method is described below. It corresponds to how location management is performed in the current cellular architectures[19]. It is also closely related to the way mobility is proposed to be handled over the Internet [15].

Assume that each user is attached to a home location server that always “knows” his current address. When a user moves, he informs his home location server about his new address. To send a message (or call) such a user, his home location server has to be contacted first to obtain his current address. In [15] a special form of “address embedding” called IPIP is used to redirect the packets addressed to the mobile user from the home location to his current location.

The major disadvantage of this scheme involves the management of so called “global” moves.

Consider A, whose permanent registration is in California, and B, whose permanent registration is in New York. Assume that both A and B are currently in California. Suppose that A sends a message to B. In order to route the message appropriately, A has to contact the location server in New York which knows that B is in California. B who may be located just “next door” to A, will receive the message only after two messages are exchanged across the whole country.

This scheme works well for users who stay within their respective home areas (i.e., they never move far away from home). As the scenario above indicates, it does not work well for global moves. Another scheme, proposed in [21], which better handles global moves, is based on the assumption that most messages are exchanged between local parties or between a user (in a remote area) and its home location area (in case he or she is outside of his or her registration area).

In this case, before sending a message to the NY home location server, A, who is located in California, will first, broadcast a message to all base stations in the local area (say within the Bay Area) to find out if B is currently located there. Thus, only if B is not found in A’s local area, will B’s home location server in New York be accessed. Assuming that this assumption mostly valid, this locating method should be close to optimal. In general, by taking into account the statistical profiles of user mobility and their calling (messaging) patterns, the overall performance of search and locating methods can be dramatically improved.

---

8 In the future each user will have just one PIN (personal identification number) which will point to his permanent, home address in the network

9 Awerbuch and Peleg in [3] provide new type of distributed directory to deal with mobile users
number of new and challenging database problems. These problems involve storage and access (query and update) of location data as well as questions of dynamic replication of resources in response to the changing system configuration.

On the application layer, users will be querying and updating databases over low bandwidth wireless links. *Wide area information services* will provide “general public” information such as traffic, weather, stock quotes etc. The access to such information will typically be *read only* and will involve massive numbers of users\(^5\). Information will be delivered over a macrocell (miles in diameter) and may or may not be location dependent\(^6\). The mobile user upon moving to a new environment will “charge” his palmtop with local information; for example, by filtering the data provided by the local information servers according to his or her predefined profile. The user’s palmtop will in fact act as his personal agent who constantly acquires new data from the surrounding environment, filters and presents it to the user. For example, the user may be informed that she has to leave earlier to the airport due to a combination of deteriorating traffic conditions and long waiting times at the parking lot. While walking in a shopping mall, the user may be attracted by the nearby furniture sale which can be broadcasted by the local furniture store etc. In this vision, the whole environment surrounding the user can be treated a large database which is literally “on ether” and information can be picked “from the air” at any point of time and at any location. This vision provides a true “first mile” of an information superhighway, connecting the user over slow wireless connection (19.2Kb/s) to vast resources on the superfast network.

Mobile users will also produce and consume information stored in databases shared by much smaller groups. Such databases may be stored both at mobile as well as at static hosts and be updated and queried over wireless connections. For example, mobile insurance agents may interact with the databases storing consumer records, while traveling salespeople will access inventory databases in order to check availability of merchandise.

4 Mobility and Scale

Mobility is a behavior with implications for both the fixed as well as the wireless networks. On the fixed network, mobile users can establish a connection from different data ports at different locations. Wireless connection enables virtually unrestricted mobility and connectivity from any location within the radio coverage. In this section, we will argue that mobility is an important new component that will have far reaching consequences for *systems* design. Mobility, in our view affects mostly the network level data management and to much less extent application layer\(^7\). Mobility of clients results in constantly changing topology of the system, calling for mobility of *resources*. Location management deals with mobility of clients, while configuration management refers to mobility of resources.

\(^{5}\) Hence, scale is an important factor here

\(^{6}\) Traffic information, airport data, weather, local advertising will depend on the location of the user, while stock quotations will not

\(^{7}\) Although some applications being inherently location dependent are affected by location
been allocated for PCN, it is expected that sufficient bandwidth will be allocated for the emerging services in PCN. This additional spectrum availability combined with small cell sizes to facilitate spectrum efficiency through frequency reuse should be sufficient to provide support for mobile wireless computing.

2.2 Palmtops

The relentless progress in hardware technology has made possible the availability of hand-held computers or palmtops. When compared to desktops, these palmtops have low speed CPUs with limited memory. However, they are capable of running a wide variety of applications such as calendar, schedule, spread sheets, etc. Examples are HP100LX from Hewlett Packard, Newton from Apple, and EO from AT&T. These palmtops have wireless capability. The HP100LX and Newton use infrared for communication while the communication in EO is via a cellular phone.

Energy conservation is a key issue for small palmtop units. Portability of palmtops requires the manufacturers to use less powerful, lighter batteries such as AA or AAA batteries. Consequently, such batteries have to be either replaced or recharged quite frequently. Recharging, even once a day, is considered to be a nuisance that has to be minimized. This is why there is a growing pressure both on hardware as well as software manufacturers to provide energy efficient CPU’s, memories, as well as systems software. The Hobbit chip from AT&T is an example of such an energy efficient CPU which consumes only 250 mW while in active mode. The power requirement in “doze” mode is only 50 μW. Solaris [4] from SUN is an operating system which has power saving features built in and provides perhaps the first example of the energy efficient systems software.

Perhaps the most surprising limitation, at least to computer scientists, is that the lifetime of a battery is expected to increase only 20% over the next 10 years [18]. This is an extremely slow rate especially comparing to the fast progress in the areas of memory capacity and CPU speed. Hence, there seems to be a long term need for energy efficient solutions in all aspects of the systems software including, as we will show later in the paper, data management. Such solutions will either increase the lifetime of a given battery set or will reduce the size of the battery set without affecting the time between successive recharges. The latter, will increase the portability of the palmtop by reducing the battery weight.

Let us now provide some numbers which characterize the sources and consumers of energy. A typical AA cell is rated to give 800 mA-Hr at 1.2 V (.96 W-Hr). The constant power dissipation in a CD-ROM (for disk spinning ) is about 1 W. The power dissipation for display is be around 2.5 W. To increase the longevity of the batteries, the CD-ROM and the display may have to be powered off most of the time. Energy is also consumed by the CPU and the memory of the palmtops. In addition, transmitting and receiving also consume power.

3 Information Services for Mobile Users

Mobile users will be consumers as well as producers of data. The data produced by the mobile users will involve location data which will have to be constantly updated and queried in real time by the network. This is mostly a network level data management problem but as we will see in the next section it involves a

\footnote{For example HP 100LX is powered by AA batteries, Newton, on the other hand, uses AAA batteries}
Other satellite systems that have been proposed are Qualcomm’s Globalstar (48 satellites orbiting the earth), and TRW’s Odyssey, (12 satellites orbiting the earth) which have the same goal of providing information services (mostly voice) around the world.

**Bit Rates**

The possible data rates over wireless in local area networks is much higher than which is available over wide area networks. For local area networks, the available data rate is 1 to 2 Mbps and is expected to increase to 10 Mbps. This is still an order of magnitude less than what is available on the fixed. The data rates for networks based on infrared technology range from 19.2 kbps to 1 Mbps. For wireless widearea data services provided by Specialized Mobile Radio (SMR) networks, the data rate is 19.2 kbps and even for schemes designed for carrying data over cellular (e.g., CDPD, stands for cellular digital packet data), the data rate is 19.2 kbps.

**Tariffs**

For most wireless data services, there is a flat fee for the service which usually covers a fixed number of messages. In addition, charges are levied on a per packet or per message basis. The flat fee for data services varies from 30 to 100 dollars a month. The cost per message ranges from 35 cents for a 50 character message to a dollar for a 100 character message. However, the tariff for sending data over cellular is based on connection time rather than amount of data. Today the typical rates for cellular vary from 15 c/minute to 60 c/minute.

We now describe the general communication architecture for personal communications network (PCN). The assumed architecture in this paper, with the understanding that it is still far from being final, is based on the existing structure of cellular telephone networks [19].

The system configuration of a fixed cellular network consists of fixed information network extended with wireless network elements. These elements include, wireless terminals, base stations, and switches. The whole geographic area is partitioned into *cells*. Each cell is covered by a *base station*, which is attached to the fixed network and provides a wireless communication link between the mobile users and the rest of the network. Currently, the average size of a cell is of the order of 1-2 miles in diameter[22]. It is expected though that the cell sizes will become much smaller (picocells of the size of a building floor). The need for smaller and smaller cells stems from frequency reuse schemes that aim to better utilize the limited radio frequency spectrum available for growing numbers of users.

Each user (sometimes called mobile terminal) will be permanently registered under one of special servers often called Home Location Server. This association of a user with a particular home location server, is fully replicated across the whole network. Therefore the user’s number uniquely identifies the proper home location server. Additionally, a user may also register as a visitor under some other location server. Thus, each location server is responsible for keeping track of the addresses of users who are currently residing in any of the base stations connected to the the location server. Many of the requirements for mobile wireless computing is expected to be handled within the framework of this architecture. The aim of PCN is to provide ubiquitous communication coverage. Network management and control functions will be distributed throughout the network so that the task of providing various services (including data services) can scale to a large number of mobile users and thus enable wide-area network connectivity. Although no spectrum has yet
2 General Architectures: Networks and Machines

The vision of mobile wireless computing requires ubiquitous wireless network connectivity, adequate wireless bandwidth, and small, portable computing platform with sufficient functionality. The necessary networking infrastructure (called PCN - Personal Communication Network) does not exist yet and is still very much under debate, although it will most likely be based on the cellular networking architecture.

Below, we describe the current status of wireless networks and their shortcomings. We also provide a short description of the PCN. Finally we discuss palmtops and their physical limitations.

2.1 Networks: Current and Future Status

Below, we list the currently used wireless networks and briefly discuss their shortcomings:

1. Cellular: The cellular (in future, microcellular) architecture (analog and digital cellular phones) is capable of providing voice and data services to users with hand held phones. Continuous coverage of cellular service is restricted to metropolitan regions. Wide area moves require the user to inform the network of the new location. Existing networking support in the cellular network does not scale to massive numbers of mobile users and there is also very little experience with data transmission. Furthermore, the available bandwidth is very low for most data intensive applications.

2. Wireless LAN: This is a traditional LAN (e.g., ethernet) extended with a wireless interface to service small low powered portable terminals capable of wireless access. The wireless LAN is further connected to a more extensive fixed network such as LAN, WAN, Internet, etc. Wireless LANs are already available. Examples include NCR's wavelan, Motorola's ALTAIR, Proxim's Range LAN, and Telesystem's ARLAN. The main component in the wireless LAN is the wireless interface card that has an antenna. This interface card can be connected to the mobile unit as well as to the fixed network. Wireless LANs have limited range and are designed to be used only in local environments (from inside buildings) and does not, as yet, provide networking support for wide area moves.

3. Wide Area Wireless Networks: These are special mobile radio networks provided by private service providers such as RAM and ARDIS. The infrastructure provides nationwide coverage for low bandwidth data services. For example, RAM network offers wireless e-mail services nationwide and ARDIS provides wireless access from a portable unit to an application running on a fixed host. There is no support for ubiquitous network connectivity and it is still unclear how such networks will scale with massive number of mobile users.

4. Paging Networks

Such as SkyTel and Motorola. Here, coverage is not a problem but the service is usually receive only and has very low bandwidth.

In addition, new satellite services are being proposed. Motorola's Iridium, with 66 satellites that orbit the earth at a low-earth-orbit (LEO), is one such proposal. The initial applications for satellite systems are predominantly voice and paging. Additional services planned include messaging and fax.
architecture is intended to support both smaller units which we call *dumb terminals* and larger units which we call *walkstations*. The dumb terminals will rely completely on the MSSs, will have no disk, and usually a very limited amount of RAM. Dumb mobile terminals (sophisticated pagers as well) will participate in any global distributed environment only via “proxies” acting on their behalf and residing on the fixed network. Here the problems are related to *ubiquitous* networking. The walkstations, on the other hand, will do a significant amount of processing locally and only occasionally use the resources of the MSS. Walkstations will have their own disk where they can cache a portion of the data, which can be queried, updated and shared among many users.

Another model where users do not carry any devices but use what is available in the environment is the Xerox’s model of ubiquitous computing [25]. Devices such as different types of *Tabs*\(^3\) will be available in the computing environment. The same user may be using different tabs at different locations.

Mobile computing will bring about a new *style* of computing. Due to battery power restrictions, the mobile clients will be frequently disconnected (powered off). Most likely, short bursts of activity, like reading and sending e-mail, or querying local databases will be separated by substantial periods of disconnection. Also, quite often, the mobile client will “wake up” in a totally new environment in some new location far away from home. Finally, due to mobility, the client may cross the border between two different cells (coverage areas) while being active (the so called *handoff* process). Handoffs are relatively straightforward in cellular voice communication due to a higher loss of information that can be tolerated. In case of data transfers, where the loss rates must be extremely low, the handoffs are more challenging.

Mobile computing poses new challenges to the data management community. How will mobility of users affect data distribution, query processing and transaction processing? What is the role of wireless medium in distribution of information? How can one query data that is broadcast over the wireless? What is the influence of limited battery life on data access from a mobile palmtop terminal? How should prolonged periods of disconnection of the mobile machines be handled?

It is useful to group the major challenges to data management brought by the vision of mobile computing into the following categories:

1. Mobility Management and Scalability

2. Bandwidth Management

3. Energy Management

They will be discussed in more details in the following sections.

The paper is organized as follows: Section 2 presents an overview of wireless infrastructures both currently available and those that are planned for the future. We also discuss new physical limitations due to the bandwidth and energy restrictions. In section 3, we present the types of information services used by mobile users and in sections 4 and 5, we detail the impact of mobility, bandwidth and energy limitations on data management problems. Finally, section 6 summarizes how different areas of data management will be affected in the mobile and wireless environments.

\(^3\) Tabs are wallet sized devices with a display and can communicate using infrared


**mu** Mobile unit (can be either dumb terminals or walkstations)

**MSS** Mobile Support Station (has a wireless interface)

**Fixed Host** (no wireless interface)

Figure 1: Model of a System to Support Mobility

...ung on different machines. Staying connected, regardless of location will also stimulate more collaborative forms of computing.

Another class of horizontal applications will include information services such as a local yellow pages possibly extended with online information such as movies currently playing at local theaters or merchandise on sale at the local supermarket.

Figure 1 shows our view of the global architecture to support mobile wireless computing. The model consists of two distinct sets of entities: mobile hosts and fixed hosts. We make no assumptions about the types of units that will exist in the system or about the type of data that mobile units will carry. Some of the fixed hosts, called MSS (Mobile Support Stations)[15], are augmented with a wireless interface to communicate with mobile hosts. Additionally, the MSSs will provide commonly used application software, so that a mobile user can download the software from the closest MSS and run it on the palmtop or execute it remotely on the MSS. Thus, the most commonly used applications will be fully replicated. Since not every file will be carried on the mobile platform, each mobile user will also be associated with a Home MSS, which will store information such as user profile, login file, access rights together with user's private files. Our
Mobile Wireless Computing:
Challenges in Data Management\(^1\)

Tomasz Imielski
B. R. Badrinath

Department of Computer Science
Rutgers University
New Brunswick, NJ 08903
e-mail: \{imielski, badri\}@cs.rutgers.edu

Abstract

Mobile computing is a new emerging computing paradigm posing many challenging data management problems. We identify these new challenges and investigate their technical significance. New research problems include management of location dependent data, information services to mobile users, frequent disconnections, wireless data broadcasting, and energy efficient data access.

1 Introduction

The rapidly expanding technology of cellular communications, wireless LAN, and satellite services will make it possible for mobile users to access information anywhere and at anytime. In the near future, tens of millions of users will be carrying a portable computer, often called a personal digital assistant or a personal communicator. Smaller units will run on AA batteries and may be diskless; larger units will run on Ni-Cd packs. These larger units will be powerful laptop computers with large memories and powerful processors. Regardless of size, all mobile computers will be equipped with a wireless connection to information networks. The resulting computing environment, which is often called mobile or nomadic computing, no longer requires a user to maintain a fixed position in the network and enables almost unrestricted user mobility. Mobility and portability will create an entire new class of applications and possibly, new massive markets combining personal computing and, consumer electronics.

There is already a number of vertical, niche applications of mobile wireless computing including: taxi dispatch, mail tracking, point of sale, public safety, trucking etc. Horizontal\(^2\), massive market applications are not here yet but two types of applications which are most frequently mentioned include: mail enabled applications [17] and information services to mobile users.

Mail enabled applications will constitute the core of mobile computing. Users carrying personal communicators will be able to receive and send electronic mail from any location, as well as be informed about certain predefined conditions (such as a plane being late or heavy traffic on the way home) irrespective of time and location. Electronic news services will be delivered and filtered according to individual user profiles. For instance, traffic information or weather reports will be filtered based upon the current position of a user, while stock information will be filtered using the user's portfolio. Electronic mail will link applications run-

\(^1\)The work presented in this paper is part of a larger project called DataMan which addresses all the above issues. The DataMan (a logical successor to WalkMan and WatchMan) project is a joint research effort with Wireless Information Network Laboratory (WINLAB) at Rutgers University. Work supported in part by NSF (SGER) award IRI-9307165.

\(^2\)Horizontal applications are domain independent, as opposed to the vertical applications which are written for a specific application domain.