Mobile Computing & Disconnected Operation: A Survey of Recent Advances

By Vietanh Nguyen

ABSTRACT

Mobile computing is a new style of computer access emerging at the intersection of the two currently dominant trends: producing portable computers in computer industry and wireless communications in telecommunication industry. This paper discusses some key issues involved in realizing a mobile wireless computing environment by examining the characteristics required of each main component: mobile computer, wireless communications network, and coordination software.

Table of Contents

Introduction

With the recent advances in portable devices and wireless networks, a new paradigm of computing has emerged which is known as mobile computing. What is mobile computing ? There are at least two definitions for this new mode of computer access. First, mobile computing can be seen as an integration of portable computers and wireless networks. Second, it can be seen as a combination of portable computers, modems, and telephone network. In both cases, the connection is temporary with periods of disconnection. There are also other terms that denote this mode of computing such as nomadic computing [12], ubiquitous computing [24], and pervasive computing.

For the purpose of identifying technical issues and research thrusts, this paper will view mobile computing environment as being made up of three main components: a *computer*(that moves), a *network*(that is either wired or wireless), and *coordination software*that ties them together. In the sections that follow, the characteristics required of each component are in turn examined in detail.

Technical Issues and Research Thrusts

This section explores the technical challenges in designing and building mobile computing systems along the line of each component of this new style of computing.

Portable Computer

An integral part of mobile computing environment is computers that are portable or embedded (e.g., navigational computer in a car [20]). Portable computers can get arbitrarily small, down to the size of, say, a walkman, a pocketbook, a watch, or a ring. The implications of portability are small size and weight and dependence on battery.

Small Storage Capacity

Small size and weight of a mobile computer means restricted memory size, small storage capacity, and small user interface. Various methods have been tried to cope with the problems of limited memory and storage capacity including compressing file systems, compressing virtual memory pages, accessing remote storage over the network, using intepreted script languages instead of compiled object codes (since compiled object codes can occupy much space). General Magic's Telescript and Apple's NewtonScript are examples of such script languages.

Small User Interface

Small user interface affects both data entry (keyboard) and data display (screen size). The physical limitations of the mobile terminals necessitate new types of interfaces which do not rely on keyboard and screen size. Among the leading contenders for supplanting the keyboard are pen and speech.The physical limitations of output displays have been tackled by such approaches as approximate answers (subsets of real answers) or intentional answers.

Pen-based data entry is attractive for its ease of use and versatility. In addition, pen positioning can be more accurate with higher resolution than that of a cursor. Handwriting recognition rates can be quite high when trained for a specific user. However, pen usage may lead to such problems as incorrect pointing due parallax between pen tip and the screen image and hand use of pen obscuring more area of the screen.

A new alphabet called Unistrokes was designed to ease the task of handwriting recognition and to make text entry simpler and faster. Each Unistroke character, which can be drawn with a single pen stroke, corresponds to a letter in the English alphabet. Since Unistroke characters are better differentiated than English letters, they require less processing for reliable recognition. [23]

Speech is also considered for data entry and display since its usage requires no surface areas and allows hands-free and eye-free operation. Speaker-independent recognition can be quite accurate. However, speech requires much processing, and may create noise for others as well as compromise privacy. Also, if used for data display, its sequential nature is not suitable for skimming.

Low Power Consumption

In a portable computer, power is consumed mostly by the screen backlighting, the central processing unit (CPU), the memory, the hard disk, the display, and the keyboard. Dependence on battery means low power consumption. Since battery technology is fairly mature and the lifetime of a battery is not expected to increase very much over the next decade, low power consumption makes energy conservation a key issue in both hardware and software.

Power consumption is proportional to the capacitance of the wires, the voltage swing, and the clock frequency. Therefore, in order to save power: [9]

1. Reduce capacitance by increasing the levels of VLSI integration and multichip module technology.

2. Reduce voltage by redesigning chips operating at lower voltage.

3. Reduce clock frequency by trading off computational speed for power savings.

Power management has spawned a new breed of energy efficient CPU's. AT&T's Hobbit chip has two modes of operation: in active mode it consumes about 250 miliwatts, and in inactive mode, it consumes only 50 microwatts. Motorola's Dragon I 68349 CPU draws 300 milliwatts at 16 MHz; when idle it draws 1 milliwatt.

The current memory technology for portable computers is dynamic RAM (DRAM) with power consumption of 0.5 watts for a bank of 4M. Flash EEPROM, a low power, non-volatile, dense storage technology, is considered as a potentially cheaper alternative. Flash memory has a read latency close to that of DRAM and a write latency close to that of disk, and can only withstand a limited number of writes over its lifetime. However, the power required for flash memory access is 20 times more than when it is idle.

In software, energy conservation has lead to new classes of energy efficient systems software, data access protocols and algorithms. Since the CPU power consumption is proportional to clock rate, adjustment of clock rate to avoid CPU idle time has been suggested with a new metric of instruction per joule (instruction/joule) and new CPU scheduling policies. SUN Solaris is an example of an operating system with power saving features built in.

As databases may be stored at stationary hosts, updates and queries may have to be done over wireless links. Since both transmitting and receiving data consume power, energy efficient query optimization techniques also need to be developed. New database query processing cost models have been proposed that are optimized for energy use rather than number of I/Os using a new metric of energy per transaction (energy/transaction). Approximate answers may also be accepted with the quality of the answer depends on the availability of computational and communication resources.

In addition to the physical constraints listed above, mobile computers are more prone to theft or destruction, and thus, pose some risks to data. Solutions for this include data backup, remote data storage, and data encryption.

Wireless Communications Network

The communications infrastructure that ties together mobile computers and software can be either the ubiquitous wired telephone system or wireless networks. The currently available wireless networks come in different types as described below. [10]

Cellular: The existing cellular phone network provides voice and data services to users with hand-held phones. Its coverage only extends for metropolitan areas, and it has problems with scalability (to accomodate a large number of mobile users), data transmission, and low bandwidth for data intensive applications.

Wireless LAN: Wireless LAN can be connected to a mobile computer or a fixed network via a wireless interface card that has an antenna. It only provides coverage for a local area, e.g., inside a building, and does not provide networking support for wide area moves.

Wireless WAN: This special mobile radio network provides wireless email services (e.g., RAM Mobile Data) or wireless access from a mobile host to an application running on a fixed host (e.g., Ardis, a two-way, store-and-forward, packet-based wireless network). Its coverage extends for a wide area, but it has low bandwidth for data services and may have problem with scalability.

Paging: Paging network provides unlimited coverage with low bandwidth, and mostly for receivers only.

In addition, there is on the horizon the arrival of *satellite microwave* networks. Motorola's Iridium proposes a network of 66 satellites for voice and paging services; additional services may include messaging and fax. Qualcomm's Globalstar envisions a network of 48 satellites and TRW's Odyssey with a network of 12 satellites; both aim at providing information services (predominantly voice) around the world.

It is clear that the existing network infrastructure is not capable of providing adequate support for a mobile wireless computing environment.

Limited Bandwidth

In terms of data rate, the data rates for infrared networks range from 19.2 kbps to 1 Mbps, and that for radio networks is 19.2 kbps. Wireless LANs have a data rate of 1 to 2 Mbps and that can be extended to 10 Mbps (in contrast to a data rate of 100 Mbps currently available on the fixed network).

In order to cope with slow data links, compression techniques can be used to conserve bandwidth. File prefetching technique, which can look ahead several references, can be used to smooth out the data flow to the mobile computer for applications having bursty demands. Lazy writeback can also be applied in a similar way. Adaptive communication protocols have been proposed to compensate for the slow speed of some existing mobile communication links and to save the communications cost by reducing link usage. Schemes for aggregating network bandwidth have been suggested to combine several wireless link signals so as to provide higher bandwidth for a period of time to certain receivers.

Limited Geographical Coverage

The existing wireless infrastructure provides only a limited geographical coverage. For instance, in the United States, wireless services spring up first in metropolitan areas and then expand outward. There are services available in almost every major city with about 90 percent of major metropolitan locations having coverage. There is also some kind of wireless service available in about 30-40 percent of the rest of the country. That means about 80 percent of the population can get access to wireless networking. However, such a figure can be quite misleading. In a practical application of mobile computing today, a field technician operating in a far-flung rural location can not get access via wireless communications to technical documents stored back at the office. [16]

Heterogeneity

As a mobile unit changes its location, it may encounter different network environments and services, and may have to use different access protocols. For instance, although all the major wireless networks servies operate in adjacent parts of the 800-900 MHz frequency bands, they each use different modulation and transmission methods. Consequently, users must have a different proprietary modem to tap into each service. Such problem of interoperability can affect the scale of mobility.

Another problem with heterogeneous networks concerns access cost. Most wireless networks services charge a flat fee for their service which usually covers a fixed number of messages. Additional charges are levied on a per packet or per message basis. As an example, the current flat fee for data services ranges from 30 to 100 dollars a month. The cost per message goes from 35 cents for a 50 character message to a dollar for a 100 character message. In contrast, the cost for sending data over cellular is based on connection time instead. The typical cuurent rates for cellular range from 15 cents per minute to 60 cents per minute. [10]

Since different services have different access costs, the cost of a query to a centralized database may depend on location of the user. A query asked when a user is connected to a wireless LAN may incur a different cost from that posed when the user is connected to a wireless WAN. Therefore, new methods for dynamic and distributed query optimization will have to be developed to handle varying access costs.

Unreliability

Wireless communication is susceptible to high error rates and transmission interference or interception. Transmission interception may cause security risks. Therefore, there is a need for security measures, which can be achieved via encryption and authentication methods implemented in either software or specialized hardware. An example of such security software is Black Pages. Black Pages is a distributed service for public key data that can bind users and keys to support authentication across administrative domains. MIT's Kerberos is another example. Kerberos is a trusted third-party authentication service. It can authenticate users without revealing their passwords on the network and create encryption keys that can be shared among mutually suspicious parties. It also allows a mobile unit to authenticate itself in a new domain. However, its security is not perfect and is susceptible to off-line password guessing attacks (since Kerberos still relies on well-chosen passwords and their secrecy) and replay attacks (an attacker retransmitting packets intercepted from the network) within a timeout period.

Coordination

Mobile Host Protocols

In order to deal explicitly with the concept of computers that move, new communications protocols are needed. The current assumptions made in protocols for the fixed network may no longer be valid due to the effects of mobility. The developments of protocols for locating a mobile host are currently under way. There have been several proposals for mobile host protocols that are compatible with the TCP/IP protocol suite, of which the best features are incorporated into a proposed standardization document called Internet Draft. [3] These protocols attempt to make the operation and performance of a mobile host indistinguishable from that of a fixed host. This goal translates into two essential requirements: operational transparency and performance transparency above IP. Operational transparency means not having to reinitialize the system or individual applications after a relocation has taken place. One of the factors required to ensure performance transparency is optimal routing of packets to and from mobile hosts. The assumptions made in traffic management for the fixed network need to be rethinked and revised since, for instance, backing off and slowly adjusting transmission rate when congestion detected is not the correct behavior in a mobile network. Characterization of mobile traffic patterns, which may involve temporal, spatial, and statistical rules, needs to be studied. In addition, such issues as compatibility with the existing infrastructure and security risks also need to be addressed.

There is also a need for adapting the simple network management protocol (SNMP), which is the common network management protocol of the TCP/IP suite, to communicate with and manage mobile network devices. An effort in this direction was made in the Walkstation II project at the Swedish Royal Institute of Technology. This project attempts a synthesis of a packet-routing wireless data network and the location update facilities of a cellular radio system. In their work on applying the SNMP to manage a mobile router, in order to accomodate the radio communication channel for the router, an enterprise specific management information base (MIB), which is a group of managed objects that can be accessed via the SNMP, was created to complete the standard MIB. This new MIB consists of a group of objects describing a radio channel and a group of objects describing the state of a radio link. [14]

Frequent or Intermittent Disconnection

In connecting a mobile computer to a network, whether using a telephone network or via radio or infrared links, the connection is temporary with periods of disconnection. In order to ensure operation in spite of intermittent disconnection, two potential approaches have been explored to mask network downtimes. One approach is to make the mobile computer more autonomous - less dependent on the network - by using such methods as file caching or prefetching, and lazy writeback. Another approach is to decouple the mobile computer and the network so that they can operate asynchronously.

As an example of the first approach, Coda file system was built to support both voluntary disconnection and involuntary disconnection in the same way using an idea called caching of data, which is widely used for performance. In the Coda file system, a user can cache all the useful files prior to disconnection, work on these files when the network is not available, and later reintegrate. Any conflicts during reintegration due to updates are resolved based on an optimistic concurrency control scheme using logs and manual repair. Caching of data can be either predicted or specified by user; predictive caching involves monitoring and discovering user profile or behavior [11].

There is an implicit assumption in the caching model that a user's working sets are smaller than practical cache sizes. If there is so much data and not all can be cached, then remote access is required, and different software techniques are needed. A design known as reconfigurable services [8] has been proposed, which is based on the idea of maintaining a larger second-level cache that can follow a user around. In this design, services can dynamically reconfigure themselves, taking advantage of the unique aspects of each service, to always be available near a mobile client.

The remote access model also applies to the case that data is being shared and being regularly removed from caches by cache consistency mechanisms. In disconnected operation, the traditional notion of strong consistency of databases using atomic transactions has been modified to become less restrictive. This is motivated by the fact that communications links are slow and expensive and thus make it difficult or uneconomical to maintain closely synchronized copies of data. Consistency may have different levels. For example, a strict read operation will require the most recent version of a file be returned whereas a loose read operation will accept any available copy. The level of consistency depends on the strength of a connection. The weaker the connection is, the weaker the level of consistency is. There is also a trade-off between consistency and availability of resources.

Bayou replicated data manager that offers session guarantees is an example of this new way of implementing consistent data access for mobile users. In Bayou, a session is defined as an abstraction for a sequence of read and write operations performed during the execution of an application. Unlike atomic transactions which ensure atomicity and serializability, sessions with guarantees are intended to enable a client to observe a database that is consistent with its own actions even if it reads and writes from various potentially inconsistent servers, and sessions permit control over the scope and selection of the guarantees. The purpose of guarantees is to make users and applications less confused about the ordering of read and write operations in weakly consistent replicated data systems. [22]

High Bandwidth Variability

In addition to the intermittent disconnection, there is also the problem of wide variations in network bandwidth. This can be caused by going from having wired to wireless access. Due to such variability, adaptation has become an essential factor in communication protocols and systems software for mobile computing. As an example of adaptation in systems software, a generalization of the Coda architecture called Odyssey has been suggested to support the capability of "application-aware adaptation" of mobile clients. The strategy embodied in this approach is that "access to shared data in mobile computing systems is best achieved by a collaborative partnership between the operating system and individual applications." [18] This collaboration means a split of functionality between the operating system and individual applications. The operating system is to sense external events such as connectivity and physical location changes and to monitor resources such as network bandwidth and allocate them accordingly, whereas each individual application is to adapt to changing conditions by using the resources provided by the operating system. Due to the diversity of data format (video images, maps, other spatial images, hypertext-like data in the web etc), the nature of adaptation depends on the type of application and is aimed at providing a client with a variable level of detail or quality.

Location Dependent Information

The mobility of a computer user suggests a new class of applications called context-aware computing. These applications are made aware of the context in which they are run, based on a limited amount of information covering a user's proximate environment, to exploit the constantly changing environment. There are three important aspects of context: where one is, who is around, and what resources are nearby. Being aware of it may help promote and mediate one's interactions with devices, computers, and others, and help navigate unfamiliar places. [19]

Such applications can help answer question like, what's the nearest server? The nearest server may cease to be the nearest due to migration. Since a physical distance may not correspond to a network distance (e.g., when crossing administrative domains), the communication path may grow out of proportion with respect to actual movement. A longer communication path not only consumes more network capacity but also has more intermediaries, and thus, longer latency and greater risk of being disconnected. A context-aware application can avoid such problem by dynamically transferring service connections to closer servers.

Practical Applications

A natural question that has often come up upon the arrival of a new technology is, what are its practical applications? For mobile computing, novel applications have been and will be proposed.

Projects

A sampling of projects on mobile computing applications are briefly described here.

A rather ambitious and influential project in mobile computing, which is known as Ubicomp or the Ubiquitous Computing Project, has been pursued at the Xerox Palo Alto Research Center (PARC) and, as the result of a reverse technology transfer process, at some other academic sites as well. The underlying idea behind this work is to make computers become invisible or blend into human environment so as to make "using a computer as refreshing as taking a walk in the woods." [6][24]

SciencePad Project - An Intelligent Electronic Notepad for Ubiquitous Scientific Computing - at Purdue: "How will mobile computing affect the way people learn and do science?" is the question this project tries to answer. A classroom of the future has been set up where SciencePads are being used on an experimental basis. SciencePad considers such issues as user interface for walkstations, the ability to find information across heterogeneous geographically distributed information systems, the dynamic reconfigurability of computations between the mobile client and the stationary servers.

Mobile Wireless Access to the World Wide Web (WWW): There have been several approaches to extending and customizing the web browser for mobile computing environment [7, 1994 Workshop on Mobile Computing Systems and Applications]. These projects try to address such issues as load balancing and location-dependent information.

Products

Though mobile computing is still an emerging technology where much research still needs to be done, there have been some commercial products.

A notable one is a wireless mobile unit introduced by Hewlett-Packard that will allow doctors to receive vital data of patients for remote diagnosis in emergencies. This application of mobile computing is considered as the first of its kind in the area of health care. [17]

Another product called Mobilizer for Windows was introduced by Digital Equipment Corporation. The concept used here is similar to that of the Coda file system. [25]

In order to shield mobile users away from the complexity of networking and communications and allows them to focus more on their work, the network can be tranformed from being a mere conduit for applications carried out at its end points to a host that itself can perform distributed applications. Such a network makes use of store-and-forward messaging transport instead of just real-time, connection-oriented sessions and is often based on an application of artificial intelligence known as agent technology. A sampling of intelligent networks include AT&T's PersonalLink and IBM's proposed Intelligent Communication. [15]

Conclusion

The arrival of mobile computing signals a paradigm shift in human-machine interface and in machine-machine interface. As such, it is expected to create new fundamental research problems in intelligent user interface and networking, some of which are discussed in this paper. While presenting many technical challenges, mobile computing also brings with it a promise for a great enhancement of human productivity.

References

1. Mobile Computing Pointers to conferences, institutes, on-line journals, etc. on mobile computing.

2. Mobile IP A description of the current mobile IP draft, which proposes protocol enhancements for transparent routing of datagrams to mobile hosts in the Internet.

3. Ubiquitous Computing The home page of Ubicomp or the Ubiquitous Computing Project at Xerox PARC.

4. The WWW Virtual Library: Mobile and Wireless Computing. A weekly updated index of ongoing projects on mobile and wireless computing.

5. Dan Duchamp, *Issues in Wireless Mobile Computing,* Proceedings of Third Workshop on Workstation Operating Systems, IEEE, Key Biscayne FL, April 1992, pp. 1-7.
   This paper mentions several aspects of systems software design that need to be rethinked to provide proper support for a wireless mobile computing environment.

6. George H. Forman and John Zahorjan, *The Challenges of Mobile Computing,* IEEE Computer, April 1994, pp. 38-47.
   A survey of software design challenges for mobile computing along its three properties: wireless communication, mobility, and portability.

7. Tomasz Imielinski and B. R. Badrinath, *Mobile Wireless Computing: Challenges in Data Management,* (To appear in Communications of the ACM).
   This paper formulates a number of open problems in data management for a mobile wireless computing environment: management of location dependent data, information services to mobile users, frequent disconnections, wireless data broadcasting, and energy efficient data access.

8. J. J. Kistler and M. Satyanarayanan, *Disconnected Operation in the Coda File System,* ACM Transaction on Computer Systems, Vol.10, No.1, February 1992, pp. 3-25.
   A description of a file system that relies on file caching and hoarding to support disconnected operation.

9. Leonard Kleinrock, *Nomadic Computing - An Opportunity,* ACM SIGCOM Computer Communication Review, pp. 36-40.
   This paper describes nomadicity as a new paradigm in the use of computer and communications technology and presents a number of of challenging problems.

10. Henning Koch, Lars Krombholz, and Oliver Theel, *A Brief Introduction into the World of 'Mobile Computing'*,Technical report THD-BS-1993-03, May 21, 1993, Department of Computer Science, University of Darmstadt.
    This brief introduction includes a discussion of the term 'mobile computing' in terms of its definition and technical features, and a mobile computing literature reference graph.

11. Markus Oelhafen, *SNMP Application for the MINT Router (Walkstation II project),* Master's thesis, Department of Teleinformatics, The Royal Institute of Technology, June 1994.
    An attempt to extend the SNMP application to the management of a mobile computer known as the Mobile INTernet router.

12. Andy Reinhardt, *The Network with Smarts: New agent-based WANs presage the future*

of connected computing, Byte, Vol. 19 (10), October 1994, pp. 51-64.
A preview of intelligent networks from AT&T and IBM that may set the model for a future of mobile software agents.

13. Andy Reinhardt, *From Here to Mobility*, Byte, Vol. 20 (6), June 1995, pp. 100-156.
A special report on the current status of mobile computing including wireless links, display technologies, smart batteries, and palmtop computers.

14. Glenn Rifkin, *New Hewlett-Packard Device Sends Vital Signs by Modem*, The New York Times, Wednesday, May 24, 1995.
A wireless mobile unit that allows doctors receive vital data of patients for remote diagnosis in emergencies - this application of mobile computing is considered as the first of its kind in the area of health care.

15. M. Satyanarayanan et al, *Application-Aware Adaptation for Mobile Computing*, Proceedings of the 6th ACM SIGOPS European Workshop, September 1994, Dagstuhl, Germany.
This paper puts forth a strategy called "application-aware adaptation" which advocates a collaborative partnership between the operating system and individual applications to cope with the problems induced by mobility.

16. Bill Schilit, Norman Adams, Roy Want, *Context-Aware Computing Applications*, In Workshop on Mobile Computing Systems and Applications (1994).
This paper contains the definition of context-aware computing, and the implementation of an instance of such applications using ParcTab, a wireless, palm-sized computer.

17. Gene Sloan, *More rental cars will steer you in right direction*, USA Today, Friday, August 11, 1995, p. D1.
The preliminary implementation of global positioning satellite-based navigational computers in rental cars by car rental companies.

18. Douglas B. Terry, *Systems Issues in Mobile Computing*, Slides of a talk at Stanford University, 1994.
A presentation of some current research thrusts or technical challenges in mobile computing.

19. Douglas B. Terry et al, *Session Guarantees for Weakly Consistent Replicated Data*, Proceedings International Conference on Parallel and Distributed Information Systems (PDIS), Austin, Texas, September 1994, pp. 140-149.
This paper proposes some techniques called session guarantees to ensure weak consistency in a replicated storage system.

20. Roy Want et al, *The ParcTab Ubiquitous Computing Experiment.*
Technical report on ParcTab, the smallest device used in the ubiquitous computing project at Xerox PARC.

21. Mark Weiser, *The Computer for the 21st Century*, Scientific American, 265(3):94-104, September 1991.
An exposition on the ubiquitous computing technology and philosophy by its visionary

for the general scientific audience.

22. Joan Wexler, *Digital software lets users mobilize their applications,* Network world, June 26, 1995.
A Window-based software from DEC that supports disconnected operation based on a concept similar to that used in the Coda file system.

Other Reports on Recent Advances in Networking 1995
Back to Raj Jain's Home Page
Last updated: Aug 25, 1995