11 MOBILE OPEN SYSTEMS TECHNOLOGY FOR THE UTILITIES INDUSTRIES

Nigel Davies[†], Gordon Blair[†], Adrian Friday[†], Peter Raven[*] and Andrew Cross[*]

[†]Distributed and Multimedia Research Group,
Department of Computing,
Lancaster University,
Lancaster,
U.K.

[*]EA Technology,
Capenhurst,
U.K.

Synopsis

This chapter considers the provision of mobile computing support for field engineers in the electricity industry. Section 11.2 describes field engineers current working practices and from these derives a set of general requirements for a mobile computing environment to support utilities workers. A key requirement which is identified is the need for field engineers to access real-time multimedia information in the field and it is on this requirement that the remainder of the chapter focuses. Sections 11.3 and 11.4 present a survey of enabling technologies to support distributed systems operating in both local and wide area wireless environments. The impact of these technologies on the provision of mobile computing support is assessed in section 11.5. Section 11.6 describes a software architecture which attempts to address the requirements highlighted in section 11.2 and in particular is designed to support real-time access to data in the field. Finally, section 11.7 considers the degree to which utilities workers requirements can be met by the surveyed technologies and considers the likely impact of remote data access on field engineers working practices.

11.1 Introduction

This chapter considers the provision of mobile computing support for field engineers in the electricity distribution industry. The work described has been carried out by the authors as part of the MOST (Mobile Open Systems Technologies for the Utilities Industries) project which is a collaborative venture involving Lancaster University, EA Technology and APM. The MOST project comes after a number of related projects which have attempted to analyse the IT requirements of the electricity distribution industry. The first of these projects was carried out in 1984/5 and examined the broad issues of IT in the engineering function of an electricity board's district office. A further programme of analysis was carried out in 1988/9 as part of RACE 1063 which looked at the requirements for information systems in the field. These projects identified a large number of application areas which required *access to data in the field*. Examples of these application areas include work scheduling to enable craftsman and jobs to be dynamically allocated, remote expert consultation in order to maximise the use of specialist expertise and access to corporate information including network diagrams and map data. At the time, these applications were considered unrealisable due to the limitations of portable computers and data communications technologies.

Since the projects described above significant advances have occurred in these technologies. In particular, the cost of portable computers has fallen dramatically with a significant percentage of all new computers sold in 1992 being portable. Battery life, storage capacity, screen capabilities and overall performance and robustness of these machines have all increased while their weight, size and cost have decreased. In parallel with these advances in computing technology corresponding (if less dramatic) advances in mobile communications technologies have occurred. The result is that it is now practical to provide field engineers with portable workstations which include local mass storage and remote access to corporate information sources.

In this chapter we explore how the state-of-the-art in mobile computing technology can be used to address the requirements of the electricity distribution industry. Particular emphasis is placed on the requirements of *field engineers*. Section 11.2 describes field engineers current working practices and from these derives a set of general requirements for a mobile computing environment to support utilities workers. A key requirement which is identified is the need for field engineers to have *integrated real-time access to multimedia information in the field*. Technologies to address this requirement are reviewed in sections 11.3 and 11.4. Section 11.3 considers the provision of support for the required level of integration and focuses on the emerging ISO standard for Open Distributed Processing (ODP) [ISO,92]. Section 11.4 presents a survey of the state-of-the-art in wireless networks. The notion of *degree of connectivity* is presented and the terms *fully connected, partially/weakly connected* and *disconnected operation* are defined. The section concludes with an examination of a wireless network interface which supports transparent routing of data over the network providing the best *quality-of-service* (QoS) at a given instant in time. The impact of wireless networking technologies on distributed systems platforms is then assessed in section 11.5. The issues are presented in terms of the changes necessary to the computational model and engineering support of the ANSAware distributed systems platform [APM,89] to enable it to operate in a mobile environment.

Section 11.6 presents a software architecture designed by the authors which attempts to address the requirements highlighted in section 11.2 and in particular is designed to support integrated mobile access to real-time data. Finally, section 11.7 considers the degree to which utilities worker's requirements can be met by the surveyed technologies and considers the likely impact of mobile data communications on field engineers working practices.

11.2 Requirements of Mobile Utilities Workers

As a result of the diverse set of activities performed by electricity distribution companies (e.g. power distribution, consumer product marketing and servicing, industrial product marketing and servicing) they have a wide range of mobile IT requirements. Many of these requirements are common to most large organisations, for example, providing access to sales figures and stock levels to travelling salespeople as in Ainger and Maher (this volume). However, the utilities are distinguished from other organisations by the size and complexity of their infrastructures and by the number of engineers required to operate and maintain these infrastructures. The electricity supply industry is an extreme example with an average regional electricity company being responsible for approximately 28,000 substations and around 50,000 km of high and low voltage cable. In addition, the nature of the power network means that maintenance work and repairs must be carried out in a timely fashion without endangering either company workers or consumers.

A detailed study of the requirements of field engineers has been conducted and a number of application areas for mobile computing have been identified. In general, these applications are based on an examination of current working practices. As an example, section 11.2.1 describes the sequence of actions involved when routine maintenance of the power distribution network is required.

11.2.1 Network Maintenance

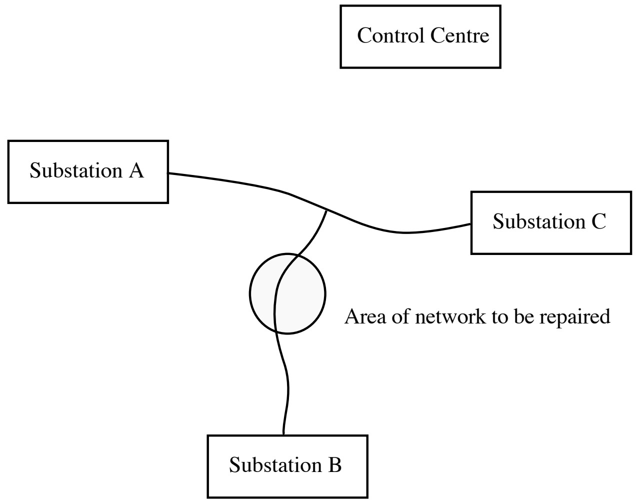

Figure 1 : Maintenance of a Section of a Power Distribution Network

All work within a regional electricity company is traditionally co-ordinated by a single control centre (see figure 1). The engineer supervising a particular repair job files a schedule with the centre some days in advance. This schedule describes in detail the stages involved in carrying out the work and, in particular, the sequence of switching which must be carried out to ensure that the work can be conducted safely (i.e. the section of network being operated on is isolated and earthed) and with the minimum of disruption to users. The control centre checks the switching schedule against its central diagram of the network state (this may be held on its computer system) and approves or rejects the schedule accordingly. Expert systems may be used by both the control centre and the field worker in the development of the switching schedule [Cross,93] although at present these systems are not integrated with the centre's representation of the global network state.

Once the schedule has been approved, the work may be carried out. On the day of the work, field engineers are dispatched to the appropriate switching points (substations A, B and C). The control centre then uses a voice-oriented private mobile radio system to instruct the staff as to which switches to operate. The use of a central co-ordinator helps to ensure the work is carried out in the correct order and allows the centre to maintain an up-to-date picture of the network's state. Once the work has been carried out the engineer must wait until returning to the office before completing the associated paper work.

There are clearly a number of disadvantages with the approach described above, in particular the lack of availability of global network state for the engineers in the field and the reduction in efficiency caused by the bottle-neck of a central point of control. The latter of these points becomes particularly important when faults occur requiring multiple unscheduled work items to be carried out. These disadvantages could be overcome by providing field engineers with a shared up-to-date picture of the network state. Providing such a picture would, of course, require a high degree of real-time synchronisation between field engineers operating on the network to ensure consistency between views of the network state.

11.2.2 Related Applications

Two additional application areas involving field engineers are voice/image store and forward systems and access to digitised maps. These are briefly explained below.

In an emergency situation (e.g. as the result of a local lightning storm) the control centre tends to become inundated by messages from field engineers requesting information about, or permission to start work on, various parts of the electricity distribution network. Since communications between the field engineers and the control centre are synchronous engineers are often forced in such situations to wait a considerable period of time for feedback. A voice/image store and forward (or email) application would enable asynchronous communications allowing field engineers to store messages requesting (for example) the current state of a network subsection on their communications device and have the device forward the message to the control centre at an appropriate time. In a multimedia system the reply might include an image showing the current state of network subsection and an accompanying voice message. This reply could then be processed at a time convenient to the engineer.

An example of an application which requires mobile computing support but only limited mobile communications is the provision of access to power system network diagrams. Currently an engineer carries paper or micro-fiche copies of the extensive networks in his vehicle. These are cumbersome and out of date for most of the time. However, a number of the regional electricity companies are now capturing their network diagrams in digital form, allowing their storage on CD-ROM. Providing support for engineers to be able to carry the network diagram in CD-ROM form is likely to be considerably more convenient and may also be cost effective. When communications is added, then it becomes possible to envisage the engineer being provided with an up to the minute view of the network diagram. Such a view would rely on the bulk of the network information being recalled from local CD-ROM memory with just the information which had changed between CD-ROM issue and the time of use being transmitted over the communications channel. This not only has implications for the ease of use of network data but also has implications for the relative roles of central co-ordinators and field engineers.

11.2.3 The Need for Integration

The regional electricity companies have utilised computer technology to support their businesses since the 1960s. Since then, and as a consequence of the diverse range of activities within the companies, a significant number of computer systems have been installed and are in use in each company. Initially, computing resources were centralised but recent developments have resulted in computing power and systems being applied more widely throughout the organisation. This diversification and dispersion results in a need to be able to treat the multiplicity of applications, data sources and hardware as an integrated computing resource. In addition to the need for integrated access to a heterogeneous computing environment within a company, there will be requirements to co-operate with other utilities in respect of 'Street Works', where initial requirements are for textual message passing related to geography and networks for CSWR purposes. This is anticipated to grow into the need for more complex interchanges involving a 'public view' of each utilities network.

During the establishment of the power industry's private mobile radio system a significant level of co-operation between regional electricity companies was achieved and a common approach adopted. This has resulted in a system in which field engineers can (subject to authorisation) use their radio systems while conducting work 'out-of-area', i.e. within an area whose radio coverage is provided by a different company. A logical extension to this approach is to ensure that access to *data* is also provided for engineers working out-of-area. Thus, the utilities industries have a clear requirement for an integrative or open system standard to address (i) the problems of heterogeneity within each organisation, (ii) interoperability between utilities and (iii) interoperability between regional electricity companies.

11.2.4 Application Characteristics

The scenarios described in sections 11.2.1 and 11.2.2 highlight the key characteristics of mobile computing applications for utilities workers. These are:-

Support for Multimedia Information

> Field engineers manipulate information in a wide variety of media types including bit map (raster) images, vector images and text. In addition, support for voice quality audio is essential to enable communication between engineers and the control centre.

Support for Integrated Access to Multiple Information Sources

Field engineers require integrated access to a heterogeneous computing environment including both mobile and fixed resources within and external to a company.

Real-time Access to Data in the Field

Field engineers require timely access to information in the field. Cooperative applications in particular require low latency communications.

The following sections describe the current technologies which are available to support these application characteristics. Section 11.3 focuses on the use of distributed systems platforms to support integration and section 11.4 presents a review of the state-of-the-art in mobile computing technologies.

11.3 Integration Technologies

11.3.1 Open Systems Standards

A number of standards to support integrated access to resources in a heterogeneous environment are emerging from both commercial and standards organisations.

Commercial Initiatives

The Distributed Computing Environment

The OSF Distributed Computing Environment (DCE) represents a significant effort by commercial and academic institutions through the Open Software Foundation (OSF) to solve the problems of application interoperability in distributed heterogeneous environments [OSF,91]. DCE comprises a number of services for use by distributed application developers. These include a remote procedure call package, a directory service, a time service, a threads service and a security service.

The Common Object Request Broker

The Object Management Group is a non-profit making collection of over 200 corporate, end-user, associate, university and subscribing members. Their aim is to promote the use of object-oriented technology in developing distributed systems. They aim to achieve this goal by publishing guidelines and specifications of common interfaces which can be used by their commercial members when developing software. As part of this process they have recently released a specification for the Common Object Request Broker (CORBA) which provides a standard interface for object interaction [OMG,91].

It should be noted that OSF and the Object Management Group are currently addressing the problem of open systems at different levels. DCE provides low-level tools for building distributed systems and is available as a product from OSF. In contrast, CORBA is a specification of desired functionality and the Object Management Group have no plans to release an implementation themselves. It is anticipated however that the work of these groups will start to converge with the release in 1994 or 1995 of OSF's Distributed Management Environment (DME) which will build on DCE and should address many of the issues currently being tackled by the Object Management Group.

The ISO Standard for Open Distributed Processing

The emerging ISO standard for Open Distributed Processing (ODP) [ISO,92] addresses the problem of standardisation *within* end systems (c.f. OSI standards which are primarily concerned with communication *between* end-systems). The terms of reference for the ODP standard are as follows:

"This standard will be concerned with, and limited to, the general aspects and common features of distributed systems. It will provide:

a) common definitions of concepts and terms for distributed processing;

b) a generalised model of distributed processing using these concepts and terms;

c) a general framework for identifying and relating together open distributed processing standards."

The standard provides a complete framework for constructing distributed systems and is undoubtedly the most technically advanced architecture available.

11.3.2 The ANSA Architecture

The ANSA architecture is of particular interest since it is has had a significant influence on the development of the ODP standard *and* is supported by an associated implementation called ANSAware. The ANSA/ISA project aims to define a complete framework for the design and construction of distributed systems [APM,89]. The ISA (Integrated Systems Architecture) project is funded within the E.C.'s Esprit program and is derived from the U.K.'s Alvey funded ANSA (Advanced Network Systems Architecture) project. The scope of the architecture being developed is wide and takes on-board the full range of issues from overall business objectives to detailed implementation choices. The complexity inherent in this broad view is managed by partitioning the architecture into five viewpoints: enterprise, information, computational, engineering and technology. The computational and engineering viewpoints are the most mature and these are briefly described below.

ANSA Computational Model

The computational viewpoint provides a programming language model of potentially distributed objects and their modes of interaction. In the ANSA computational model, all interacting entities are treated uniformly as *objects*. Objects are accessed through *interfaces* which define named *operations* together with constraints on their *invocation*. Interfaces are first class entities in their own right, and references to them may be freely passed around the system. Interface references are also the sole entities which may be passed to and from operations as arguments and results. Operations may return different combinations of types of interface reference in different circumstances: these are known as alternative *terminations*.

Services are made available for access by *exporting* an interface to a *trader*. The trader therefore acts as a database of services available in the system. Each entry in this 'database' describes an interface in terms of an abstract data type signature for the object and a set of attributes associated with the object. A client wishing to interact with a service interface must *import* the interface by specifying a set of requirements in terms of operations and attribute values. This will be matched against the available services in the trader and a suitable candidate selected. Note that an exact match is not required: ANSA supports a subtyping policy whereby an interface providing at least the required behaviour can be substituted. Finally, once an interface has been selected, the system can arrange a *binding* to the appropriate implementation of that object and thus allow operations to be invoked.

ANSA Engineering Model

The engineering model sets out specifications, guidelines and concepts by which an abstract computational model may be realised at a systems level [APM,89]. APM Ltd. have released a software system, known as ANSAware, which is a partial implementation of the computational model. In particular it does not enforce the computational model requirement that all operation arguments and results are interface references: most arguments and results are passed by value. In engineering terms, the ANSAware package is a fairly complete implementation of the ANSA engineering model as described in the ANSA Reference Manual [APM,89].

To provide a platform conformant with the computational model, the ANSAware suite augments a general purpose programming language (usually C) with two additional languages. The first of these is IDL (Interface Definition Language), which allows interfaces to be precisely defined in terms of operations as required by the computational model. The second language, prepc, is embedded in a host language, such as C, and allows interactions to be specified between programs which implement the behaviour defined by these interfaces. Specifically, prepc statements allow the programmer to *import* and *export* interfaces, and to *invoke* operations in those interfaces.

In the engineering infrastructure, the binding necessary for invocations is provided by a remote procedure call protocol known as REX (Remote EXecution protocol). This is layered on top of a generic transport layer interface known as a *message passing service* (MPS). A number of additional protocols may be included at both the MPS and the execution protocol levels and these may be combined in a number of different configurations. The infrastructure also supports lightweight threads within objects so that multiple concurrent invocations can be dealt with.

All the above engineering functionality is collected into a single library, and an instance of this library is linked with application code to form a *capsule*. Each capsule may implement one or more computational objects. In the UNIX operating system, a capsule corresponds to a single UNIX process. Computational objects always communicate via invocation at the conceptual level but, as may be expected, invocation between objects in the same capsule is actually implemented by straightforward procedure calls rather than by execution protocols. ANSAware currently runs on a variety of operating systems platforms including UNIX, VMS and MS-DOS).

11.4 Mobile Computing Technologies

The power distribution network encompasses almost all types of terrain and hence field engineers are expected to operate in situations ranging from town centres and office buildings to remote hill sides and river banks. Clearly, providing wire-based communications facilities to engineers in the latter environments is not practical. Equally, it seems inefficient to use poor-quality wireless communications technologies (relative to wire-based communications) to support engineers operating in the former types of environment. It seems likely therefore that engineers will be expected to utilise a range of communications technologies. Figure 2 illustrates the relationship between freedom of movement for engineers and the likely QoS of their communications channel.

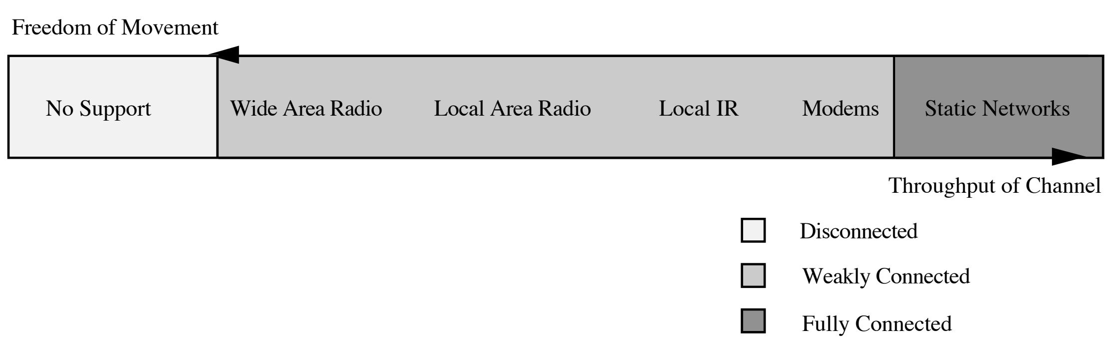

Figure 2 : The Relationship Between Mobility and Communications Bandwidth

Note that the terms *disconnected operation*, *weakly connected (or partially connected) operation* and *fully connected operation* have recently emerged to describe the degree of connectivity (i.e. channel bandwidth and availability) between system components. In general, a mobile is said to be operating in disconnected mode if no communications with its base station are possible. Weakly connected operation implies that limited communication is possible while fully connected operation assumes the existence of a communications link offering comparable performance to a current wire based local-area network.

11.4.1 Mobile Computing Technology

Before examining the communications technologies which can be used to support access to data in the field it is useful to consider the capabilities of the end systems. Portable machines have an inherent limitation in that they must always trade-off computing power against size and battery-life. In addition, size restrictions make large displays impractical. However, in recent years technological developments have enabled portable machines to close the gap with their desk based counterparts. These developments include the introduction of thin film transistor (TFT) displays which can compete with desk-top video display units in terms of image quality and low-voltage processors which extend battery-life. Despite these advances, devices which are taken as standard equipment on a desk-top machine, e.g. large hard disks and expansion cards, are difficult to support in a mobile environment. Magnetic disks for instance are a major drain on a portable's batteries with, perhaps surprisingly, the extent of this drain being directly related to their storage capacity [Douglis,93].

For the purposes of this chapter we assume that power consumption is not a critical factor. In our work on MOST we address the issue by assuming that the engineer's mobile computers will be based primarily within their vans and thus have access to an external power source. Disconnection from this power source will be for strictly limited periods of time.

11.4.2 Local Area Communications Technologies

Within a local area there are three main categories of communications technology available to provide connectivity: traditional wire/fiber based networks (e.g. Ethernet, FDDI, DQDB), infra-red wireless technology (e.g. Photonic's Photolink) and radio based wireless technology (e.g. NCR's WaveLAN). There are numerous sources of information on wire (or fiber) based networks available and in this section we concentrate on a brief survey of the main features of wireless technologies.

Infra-Red Networks

Optical networks operate with very high carrier frequencies (typically 300THz) giving a huge theoretical bandwidth. Finite transceiver performance places limits on the bandwidth, but it is still far in advance of that available in the other main local wireless technology: radio based systems. In addition, an optical system has several inherent advantages; it is confined by any opaque barrier offering good security and reducing the load on software security measures, it can be based on reliable cheap hardware such as LEDs, CD lasers and plastic optical components and the power consumption of the hardware is around an order of magnitude less than an equivalent radio system.

In infra-red networks optical propagation is via two main technologies:

Multi-directional (diffuse)

> These systems support complete mobility but the bandwidth is fundamentally limited by multi-path signals to 260Mb/m/s. Such systems have been demonstrated at 10Mbps in a 10m x 10m x 3m room.

Line of sight

> Line of sight systems support bit rates of hundreds of Mbps due to the confinement of energy and absence of multi-path effects. Their susceptibility to path blockage is limited by the beam size of the transmitters and field of view of the receivers. An adaptation of this technique whereby all transmitters are aimed at a common receiver (for example on the ceiling) can be used to help alleviate path blockage effects.

Infra-red technologies are becoming increasingly available in the commercial arena; Apple's latest computer, the Newton (a pen based handheld unit), utilises infra-red technology to connect to other Newtons or an AppleTalk wired backbone. Olivetti's active badge system, which allows the location and paging of badge wearers within buildings, also uses infra-red technology, albeit in a very low bandwidth diffuse form.

Radio Networks

Local area radio networks are designed to connect computers over relatively short distances, typically within an office or plant. There are three main reasons for their use:-

- As a substitute for wired backbones where it is expensive, complicated or undesirable (for example in a listed building) to install cables.
- To facilitate office re-organisation.
- For networking portable computers.

Due to the high amounts of interference experienced in working environments local area radio systems often utilise spread spectrum technology (SST). There are two main SST forms, direct sequence and frequency hopping. Direct sequence systems use a spreading code to give the bit stream a more complex form which is then broadcast over a larger bandwidth. The code incorporates a high degree of redundancy allowing recombination on reception despite the interference. In practical environments this form of spread spectrum technology is found to be more susceptible to interference than frequency hopping and also gives rise to self-jamming when a number of transmitters are operative in a close proximity. In frequency hopping the bit stream is transmitted with error detection data. If the transmission fails then a new channel is picked using a predictable pseudo-random sequence. This form of SST is found to be more robust in practical situations. Multiple transmitters can use different pseudo-random sequences to give higher throughput over the same bandwidth.

There are a number of commercially available wireless LAN systems, including NCR's WaveLAN [Tuch,91] offering 2Mbps and Motorola's Altair [Buchholz,91], [Freeburg,91] offering 15Mbps.

11.4.3 Wide Area Communications Technologies

Once engineers leave their local environment they must rely on wide-area communications technologies to provide them with a communications channel. If they are working near to a population centre then they are likely to be able to utilise an existing wide-area wire-based network such as the PSTN. Such networks typically provide a respectable quality communications channel with a throughput of the order of 19.2 Kbps. However, much of the work that is carried out by field engineers is in areas where it is simply not practical to obtain a wired link. Nor is it practical to use infra-red based technologies in the wide-area (aside from the need for numerous transceivers and an associated backbone network, infra-red signals are adversely affected by sunlight, rain etc.). Hence, if field engineers are to be supported in the wide-area radio-based communications technologies must be utilised.

Currently, almost all wide-area communications are provided by *cellular* radio systems. The concept of cellular radio was developed to allow the wide-spread deployment of mobile telephones. Previously such telephones were almost impossible to acquire since the limited number of channels available for this application would only support a small number of simultaneous conversations and hence few subscribers [Walker,90]. Since trials in the US in 1979 which demonstrated that cellular radio was a viable proposition there has been a massive growth of the cellular radio industry with mobile phones becoming common-place in Europe, Scandinavia and the US.

Cellular radio systems are able to support a large number of subscribers using frequency reuse. The area to be covered is divided into a number of cells with each cell being serviced by its own low-power transmitter. Frequencies can then be re-used in cells which are sufficiently far-apart. Figure 3 shows a cellular system based on a seven cell repeat pattern (other repeat patterns are possible), i.e. the area is divided into seven cell clusters and all of the available radio channels are used by each cluster.

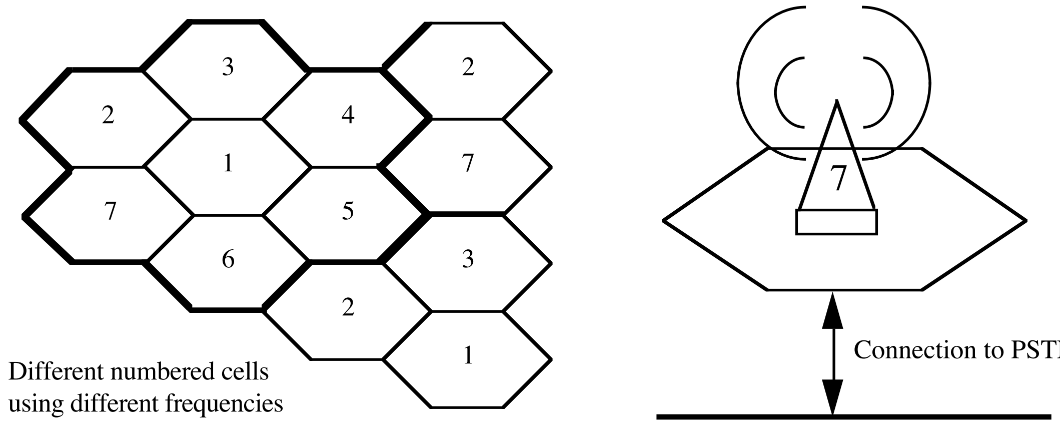

Figure 3 : A Cellular Radio System Based on a Seven Cell Repeat Pattern

Mobiles communicate with the base station of the cell in which they currently reside and the base station relays the message using a wire-based network. As mobiles approach a cell boundary the signal strength fades and the communications link must be switched to the adjacent transmitter (this process is termed handoff or handover).

The regular re-use of frequencies achieved by a cellular plan is used by modern private mobile radio (PMR) systems as well as the public service. PMR systems are used by the utilities and the emergency services because of the need for reliability during crisis conditions and because of the different coverage requirements which a utility has compared to the general public. The rate of development in mobile communications is high and a 2nd generation public cellular service is currently being introduced across Europe. This uses a number of digital techniques to support both higher frequency re-use and an improved quality of service. PMR is also subject to developments using digital techniques. In both areas the channel structure is digital resulting in an improved capability to carry higher speed data. The basic digital circuit carries voice in a digitally encoded form. There are research projects currently underway to produce 3rd generation mobile communications systems and these promise to push data rates even higher, with 64kbits/sec being predicted for the early years of the next decade.

The combination of improvements in both modulation techniques which allow more bits/sec to be transmitted in a given bandwidth and the error performance of these systems when in an interference limited condition (allowing more frequent spatial re-use) results in significantly higher traffic capacities than had earlier been thought possible.

Both public and private cellular systems have a number of common characteristics. In particular, they both use a trunking scheme to allocate radio channels. In such systems one channel carries the control traffic and service traffic is carried by one of a number of additional channels. The allocation of mobiles to traffic channels is carried out using the control channel and usually lasts for the duration of a call. In quiescent conditions all mobiles will listen to the control channel. This channel is typically structured and provides a timing reference to the mobiles so that transmissions can occur during a specified time slot. Contention between mobiles for access to the control channel is dealt with by message collisions being detected and requests for retransmission being issued. Mobiles re-transmit with variable delays to avoid repeated collisions (a slotted Aloha scheme). In many systems multiple mobiles may be assigned to a single traffic channel to allow multi-party communications. Currently, both public and private systems are capable of carrying voice and data traffic and both offer data throughput in the region of 9.6 Kbps.

11.4.4 Network Interfaces

Currently most end-systems support either a single network interface or multiple interfaces with messages having to be explicitly routed via the selected network. The Walkstation II project

[Hager,93] is an attempt by researchers to provide a user transparent network interface unit (MINT) which is capable of dynamically utilising the best available communications network. The prototype unit will have a standard Ethernet connection in addition to directed infra-red and wide-area radio transceivers. Connection between the unit and its host computer will be via an Ethernet interface in order to minimise changes required to the host. All of the protocol processing required to interact with the different networks will be carried out on-board the unit using a 68030 processor. The developers propose that the same unit will be used for both mobiles and base stations with only the controlling software differentiating them.

It is the authors belief that the future of mobile computing will be heavily influenced by the successful development of devices such as MINT which allow users to transparently exploit the best available communications channel. However, in the near to medium term the transparent exploitation of different communications networks using technologies such as MINT is clearly infeasible for field engineers. Despite this, it is clear that applications designed to support field engineers will have a number of different communications channels available to them during a typical working cycle. It *is* realistic to imagine a scenario in which a field engineer uses his mobile computer in the office (connectivity provided by a wire-based LAN), in the field (connectivity provided by a radio-network) and at home (connectivity provided by a modem and the PSTN) with periods of disconnected operation occurring throughout the cycle. Given such a scenario, section 11.5 examines the impact such a diverse range of communications capabilities will have on key aspects of a mobile computing support system, focusing in particular on the impact on distributed systems platforms.

11.5 The Impact of Mobility on Distributed Systems Platforms

The literature on mobile computing contains a number of general overviews of the impact of mobility on computer support (e.g. [Duchamp,91], [Duchamp,92], [Douglis,92], [Imielinski,92]). In contrast, in this section we concentrate on *the impact of heterogeneous networks on distributed systems platforms* and in particular, on the impact of wireless networks and weak connectivity on ODP based architectures such as ANSA (see section 11.3). The authors believe that the issue of weak connectivity lies at the heart of mobile computing: all of the wireless communications technologies discussed in section 11.4 are characterised by having substantially lower bandwidths and higher error rates than their wired counterparts. In addition, while total disconnection is rare when a wired network is used for communication, it may be relatively common if radio or infra-red communications are used to provide connectivity. We focus on the impact of these types of network on distributed systems platforms for two reasons. Firstly, because distributed systems platforms can be used to support operation in a heterogeneous environment (see section 11.3) and secondly, because experience of based wired environments has shown that such platforms are essential to support the development of complex distributed applications (e.g. those identified in section 11.2) [Birman,91].

11.5.1 Computational Issues

The introduction of mobility affects both engineering and computational aspects of a distributed systems platform. To understand how the computational model can be affected consider the implementation of a simple ANSAware application designed to warn field engineers of approaching bad weather. The application consists of a single, central service which has access to national weather information and a number of client objects (one for each field engineer) which can query the service. Engineers need to be warned when bad weather is recorded in areas adjoining the one in which they are working. Hence, there are two ways of structuring this application: either the client applications register an interest in the weather in particular areas and the server notifies them of any subsequent changes in condition, or, the client applications poll the server at regular intervals for the weather reports of the relevant areas. In a wired environment the former solution would almost certainly be adopted to avoid the communications and processing overheads incurred by polling. However, in a wireless environment there are additional factors to consider. Specifically, a call-back based approach assumes that communications are reliable. If they are not, then an engineer cannot tell whether they have received no notification of bad weather because there is no bad weather or because the server has been unable to contact his mobile.

The example above illustrates how the introduction of mobility can affect an application's structure and functionality even when an abstract computational model is used. The general requirement which emerges from this type of example is that applications and systems software running over mobile networks should be *designed to adapt their mode of operation to variations in the environment in which they are running*. This implies that in order to appropriately react to changes in connectivity, applications must have some control over connectivity and some means of monitoring connectivity so that changes can be immediately acted upon. More specifically, applications require computational level support for monitoring and control of the *quality-of-service* of their underlying communications infrastructure. Details on how this might be achieved are presented in section 11.6.

A further example of the impact of mobility on computational modelling issues is the increase in importance of a service's location. Currently most distributed systems provide a measure of location transparency, i.e. clients need not be aware of the location of services they use. This allows dynamic reconfiguration of applications and the migration of services. However, in a wireless environment service location is clearly a critical consideration of clients when selecting a set of services to use. Recent work in this area [Neuman,93] has begun to explore how location can be used as a parameter in trading to allow clients to use the best located service which meets their requirements.

11.5.2 Engineering Issues

ANSAware is currently implemented using traditional operating systems, file systems and communications protocols. However, the introduction of mobility requires changes to all of these support technologies.

Operating Systems

Two examples of changes to operating systems have been identified, both as a result of the limited resources available to systems running on portable computers.

The first change relates to the requirement to support frequent power-downs to conserve battery life. Operating systems such as UNIX incur substantial overheads on power-down and start-up: typical start-up times for a UNIX system are measured in minutes. In contrast, comparatively simple operating systems such as MS-DOS are substantially faster at powering-down and starting-up. However, operating systems such as MS-DOS do not support the concurrent processing required to support a comprehensive distributed systems platform. Attempts to provide UNIX with checkpointing facilities to allow rapid power-downs and re-starts have had mixed success with average re-start times still being around two minutes [Bender,93].

A second example of the impact of mobility on operating systems is less obvious and relates to traditional operating system's approach to opening object files for execution. Executable files are currently increasing significantly in size year on year. The main technique which has been used to keep down the size of executable files is the use of shared libraries. Using shared libraries commonly used functions such as maths routines, graphics routines and user-interface support can be held in libraries which are linked to the executable at run-time. Libraries may be shared and hence only one copy of the functions is required thus obviating the need for each executable to maintain its own copy (and hence saving substantial amounts of disk space: executables for graphics based applications often reduce to about one tenth of their original size when shared libraries are used [Sabatella,90]). However, current shared libraries have themselves grown in size such that a graphics based application may well include at run-time several megabytes of library code. In a workstation environment this is not a problem with the vast quantities of virtual memory available. However, in a mobile environment such demands are likely to cause significant problems as trade-offs have to be made between virtual memory size and power consumption (disks are a major drain on power with the size of a disk being a significant factor in determining its overall power drain [Douglis,93]).

File Systems

File systems for mobile computing have been the subject of significant research efforts for a number of years. The pioneering work in this area was carried out as part of the Coda file system project at Carnegie Mellon University [Satyanarayanan,90]. Coda was designed as a highly available version of the Andrew File System; servers are replicated and clients cache whole files to allow them to survive temporary network failures which prevent communication with servers. It was soon realised that the same strategy (client caching of whole files) could also be used to deal with voluntary disconnection as experienced when a mobile computer is disconnected from the network. The current implementation of the Coda file system has been shown to work well with moderate sized user caches (60-200 Mbytes) [Satyanarayanan,93]. This performance relies on two assumptions: firstly, that users are able to provide hints to the system informing it which files should be cached prior to disconnection and secondly, that users do not write-share files. Clearly, a file system such as Coda can be used to support a distributed systems platform such as ANSAware. However, a number of research issues remain including:-

- providing support for weakly connected operation where the restricted bandwidth means that caching whole files is not necessarily appropriate,
- supporting files which are heavily write-shared in a weakly connected environment,
- determining whether the user or the system can better determine the appropriate contents of the client cache prior to disconnection, and
- determining the correct relationship between the object-oriented nature of the distributed systems platform and the underlying file system.

Protocols

Mobility also requires new or extended communications protocols. A significant body of work on extensions to the IP addressing scheme now exists (e.g. [Ioannidis,91]) and more recent work in this area has begun to examine how connections for continuous media can be maintained as a mobile moves between cells [Keeton,93]. From the perspective of supporting distributed applications it is safe to assume that most of these issues will be handled transparently by the system. However, as discussed earlier, in many cases it may be useful for applications to be made aware of the QoS that can be provided by the underlying communications medium. In particular, the connection oriented nature of current mobile communications systems is likely to impinge on the implementation of connection-less protocols of the type commonly used in distributed computing. More specifically, the following aspects of mobile communications will need to be addressed:-

(i) *Call connections*

In all wide-area mobile radio systems the time to establish a call is significant (typically several seconds). However, the cost of maintaining an open connection between a mobile and the base station may also be expensive, depending on the type of call charging being used. Applications which require regular short-duration connections are likely to incur significant overheads. An important measure of the system's functionality will be its ability to rationalise communications to keep connection times at a minimum while providing adequate response times.

(ii) *Send/Receive turn-around time*

Many PMR systems are half-duplex with the typical time to switch from send to receive mode being approximately half a second. In such a situation it will be important to keep both protocol and application level acknowledgements to a minimum.

(iii) *Multicasting facilities*

Many mobile applications are co-operative in nature and hence demand group communications. Radio should be an ideal medium for supporting this type of communications since it is inherently a broadcast medium. However, in most practical radio systems multicasting is extremely difficult to implement since it leads to contention for a reply channel between the recipients of a multicast message. The ability to efficiently support multicast will clearly have an impact on potential applications structure.

The use of time division multiple access (TDMA) techniques in 2nd generation mobile communications systems offers the potential for a more complex channel structure to cope with the demands of multi-party communications. While the overall rate of the channel remains limited a TDMA structure would allow a dynamic, demand responsive allocation of channel resources to the mobiles competing to pass traffic back to the base station.

This is an area in which the early exploration of the way in which future communications are likely to be exploited may provide important input into the planning and design of the mobile telecommunications infrastructure. Current communications thinking tends to polarise into either speech communications or computer communications. In the field of mobile communications the two types of traffic are likely to converge onto a single network (this may also be true of the fixed networks). Applications such as those being considered in the MOST project are likely to set a different and more complex set of communications requirements than either voice or data traffic on their own and could be used by telecommunications network planners to ensure sufficient flexibility is planned into networks.

As discussed earlier, applications programmers may wish to be informed of the QoS or characteristics of the underlying communications channels. However, it is clearly unreasonable to expect programmers to take account of, for example, the full range of call charging schemes in operation (since these strategies may be subject to change it is probably undesirable to implement support within applications anyway). In addition, many decisions regarding channel optimisation will depend on the mobile's overall requirements rather than on the requirements of any one application. Clearly, therefore, there is a case for systems support for managing these types of issue and presenting the information to application programmers in a form which is useful.

11.6 The MOST Approach

Within the MOST project the authors are attempting to develop techniques to address the issues highlighted in section 11.4. The MOST architecture is shown in figure 4.

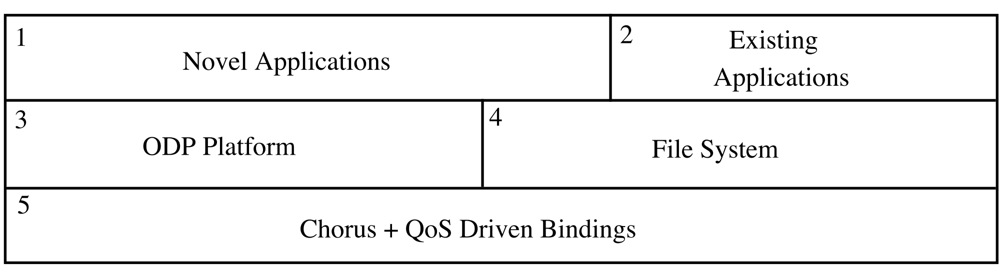

Figure 4 : The MOST Architecture

The architecture is supported [5] by an extended version of the Chorus distributed operating system [Rozier,90]. Chorus has been chosen as the operating system for a number of reasons.

Firstly, the authors have access to the source code which enables us implement any changes required to support mobility. Secondly, the use of modular design and implementation techniques within Chorus allows relatively easy substitution of new components such as real-time scheduling algorithms and memory management functions. Finally, components such as device drivers can be implemented (and debugged) in Chorus as user processes and then simply run as system processes when they are complete. This is in sharp contrast to the effort required to develop new device drivers in, for example, UNIX.

The principal extension we propose to the Chorus system is the addition of QoS driven bindings for inter-process communications. A related project at Lancaster is currently investigating this issue with the aim of supporting real-time continuous media processing [Coulson,93]. The QoS parameters which have thus far been identified within this project are throughput (the average number of bits that may be transferred in a given period), latency (the time the receiver must wait for the first bits of information) and jitter (the variation in delay between packets of information). Our work will build on their results with the addition of new QoS parameters to reflect the impact of mobility. In particular, we intend to support a QoS parameter which can be used to express the notion of degree of connectivity (see section 11.5.2).

The extended version of Chorus will be used to implement both an ODP compatible distributed systems platform [3] and a distributed file system [4]. In the former case the QoS bindings will be made visible to application programmers in the form of object bindings. Application programmers will then be able to specify the desired QoS of a given binding and be able to be notified if this QoS can not be provided. In the latter case (the distributed file system) the QoS driven bindings will be used in the implementation of the file system. In particular, it is anticipated that the file system will make use of changes in throughput, latency and degree of connectivity to determine its caching and cache consistency strategies. The distributed file system will be used to support both mobile aware, ODP based applications [1] and conventional applications [2] to whom mobility must be made as transparent as possible. An item for further study is the relationship between the file system and the distributed systems platform. This is currently being considered in the context of a wired network in projects such as Harness [Balis,93] but is of more significance in a mobile environment where interaction between applications and the file system's management routines offer far more scope for improving performance and availability.

11.7 Concluding Remarks

The widespread implementation of a mobile computing system such as MOST would have a profound influence on the working practices of field engineers within the utilities industries. Current working practices have evolved for work to be carried out with minimal interconnection between staff. The ability for tasks to be undertaken by staff who have access to all the information relevant to that task and to be able to work effectively as a team with other staff who are remote from them, will have a major impact on the way in which work is organised. The balance between central and delegated control is likely to be changed, enabling working practices which are better adapted to the rapid changes in work load which can occur during crisis. However, while as technologists it is possible to see some of the potential for organisational change which future field information and communication systems can enable, it will be through close collaboration between the computing, telecommunication and organisational domains that the greatest benefits will be achieved. Experiences with introducing new technologies into industry have shown that it will be absolutely essential that the social and technical aspects of design are tackled in a balanced way.

From a computing support perspective the utilities present a number of unique requirements. In particular, and in contrast to many mobile computing applications, field engineers require *real-time access to shared data* in the field. Additionally, much of the data manipulated is safety critical in nature so correct responses to failures are critical. Experiences with wire-based distributed systems has shown that to develop co-operative applications of this type programmers must be provided with an adequate tool-kit of simplifying functions.

We have argued that the introduction of mobility requires the design and implementation of new tool-kits (or extensions to existing systems) which allow the development of applications which respond to changes in the QoS offered by underlying communications channels. This requirement is likely to become increasingly important as technologies such as MINT make it possible for applications to experience fluctuations of several orders of magnitude as users cross from one network domain to another. Tool-kits which can accommodate fluctuations of this type can also be used as the basis for re-implementing aspects of conventional computing environments (such as file systems) which require substantial changes to operate in a mobile environment.

As a starting point for the development of a distributed systems tool-kit for mobile computing we have described the MOST project architecture. This uses an extended version of a state-of-the-art micro-kernel to underpin a mobile aware distributed systems platform and associated file system. This architecture is, we believe, capable of supporting both novel, mobile aware applications and conventional file based applications designed to operate in a wired-environment.

Acknowledgements

The work described in this chapter was carried out under the auspices of the MOST project (SERC/DTI Grant No. GR/H87704). The authors would like to acknowledge the assistance of the regional electricity companies in determining the requirements of mobile utilities workers and their co-workers in EA Technology, Capenhurst and the Distributed Multimedia Research group at Lancaster University.

References

[APM,89] A.P.M. Ltd. "The ANSA Reference Manual Release 01.00", APM Cambridge Limited, UK., March 1989.

[Balis,93] Balis, A., A. Dede, and K. Folokos. "HARNESS- An Evolving Standard for Distributed Processing." *Proc. Int. Conf. on Open Distributed Processing,* Berlin, Germany, 13-16 September, 1993. Pages 167-175.

[Bender,93] Bender, M., A. Davidson, C. Dong, S. Drach, A. Glenning, K. Jacob, J. Jia, J. Kempf, N. Periakaruppan, G. Snow, and B. Wong. "Unix for Nomads: Making Unix Support Mobile Computing." *Proc. USENIX Symposium on Mobile and Location Independent Computing,* Cambridge, Massachusetts, 2-3 August, 1993. Pages 53-67.

[Birman,91] Birman, K.P. "The Process Group Approach to Reliable Distributed Computing", *Technical Report,* Dept. of Computer Science, Cornell University, U.S.A., July 1991.

[Buchholz,91] Buchholz, D., P. Odlzko, M. Taylor, and R. White. "Wireless In-Building Network Architecture and Protocols", IEE Network Magazine, November 1991. Pages 31-38.

[Coulson,93] Coulson, G., and G.S. Blair. "Micro-kernel Support for Continuous Media in Distributed Systems", *Internal Report MPG-93-04*, Computing Department, Lancaster University, Bailrigg, Lancaster, U.K., February 1993.

[Cross,93] Cross, A.D., J.R. Brailsford, and A.T. Brint. "Expert Systems to Support Network Switching." *Proc. 12th International Conference on Electricity Distribution, CIRED 1993,* Birmingham, U.K., 17-21 May, 1993.

[Douglis,92] Douglis, F., and B. Marsh. "The Workstation as a Waystation: Integrating Mobility into Computing Environments." *Proc. Third Workshop on Workstation Operating Systems,* Key Biscayne, Florida, U.S.A., IEEE Computer Society Press. Pages 17-21.

[Douglis,93] Douglis, F., and B. Marsh. "Low-power Disk Management for Mobile Computers", *Technical Report MITL-TR-53-93*, Matsushita Information Technology Laboratory, April 1993.

[Duchamp,91] Duchamp, D., S.K. Feiner, and J. G. Q. Maguire. "Software Technology for Wireless Mobile Computing", IEE Network Magazine, November 1991. Pages 12-18.

[Duchamp,92] Duchamp, D. "Issues in Wireless Mobile Computing." *Proc. Third Workshop on Workstation Operating Systems,* Key Biscayne, Florida, U.S.A., IEEE Computer Society Press. Pages 2-10.

[Freeburg,91] Freeburg, T.A. "Enabling Technologies for Wireless In-Building Network Communications - Four Technical Challenges, Four Solutions", IEEE Communications Magazine, April 1991. Pages 58-64.

[Hager,93] Hager, R., A. Klemets, G.Q. Maguire, M.T. Smith, and F. Reichert. "MINT - A Mobile Internet Router." *Proc. IEEE VTC'93,* Secaucus, NJ, U.S.A., 1993.

[Imielinski,92] Imielinski, T., and B.R. Badrinath. "Mobile Wirless Computing: Solutions and Challenges in Data Management", *Technical Report* Department of Computer Science, Rutgers University, New Brunswick, U.S.A., 1992.

[Ioannidis,91] Ioannidis, J., D. Duchamp, and G.Q.M. Jr. "IP-based Protocols for Mobile Internetworking." *Proc. SIGCOMM'91,* Zürich, Switzerland, ACM Press, Vol. 21. Pages 235-245.

[ISO,92] ISO. "Draft Recommendation X.901: Basic Reference Model of Open Distributed Processing - Part1: Overview and Guide to Use". 1992.

[Keeton,93] Keeton, K., B.A. Mah, S. Seshan, R.H. Katz, and D. Ferrari. "Providing Connection-Oriented Network Services to Mobile Hosts." *Proc. USENIX Symposium on Mobile and Location Independent Computing,* Cambridge, Massachusetts, 2-3 August, 1993. Pages 83-102.

[Neuman,93] Neuman, B.C., S.S. Augart, and S. Upasani. "Using Prospero to Support Integrated Location-Independent Computing." *Proc. USENIX Symposium on Mobile and Location Independent Computing,* Cambridge, Massachusetts, 2-3 August, 1993. Pages 29-34.

[OMG,91] The Object Management Group. "The Common Object Request Broker: Architecture and Specification (CORBA)", *Report 91.12.1*, The Object Management Group, Framingham, MA, U.S.A., 1991.

[OSF,91] OSF. "Distributed Computing Environment: An Overview", Open Software Foundation, April 1991.

[Rozier,90] Rozier, M., V. Abrossimov, F. Armand, I. Boule, M. Gien, M. Guillemont, F. Herrmann, C. Kaiser, S. Langlois, P. Léonard, and W. Neuhauser. "Overview of the CHORUS Distributed Operating System", *Technical Report CS/TR-90-25*, Chorus systèmes, 1990.

[Sabatella,90] Sabatella, M. "Issues in Shared Library Design." *Proc. USENIX Summer Conference,* Anaheim, California, 1990. Pages 11-23.

[Satyanarayanan,90] Satyanarayanan, M., J.J. Kistler, P. Kumar, M.E. Okasaki, E.H. Siegel, and D.C. Steere. "Coda: A Highly Available File System for a Distributed Workstation Environment." *IEEE Transactions on Computers* Vol. 39 No. 4. Pages 447-459.

[Satyanarayanan,93] Satyanarayanan, M., J.J. Kistler, L.B. Mummert, M.R. Ebling, P. Kumar, and Q. Lu. "Experiences with Disconnected Operation in a Mobile Environment." *Proc. USENIX Symposium on Mobile and Location Independent Computing,* Cambridge, Massachusetts. Pages 11-28.

[Tuch,91] Tuch, B. "An ISM Band Spread Spectrum Local Area Network: WaveLAN." *Proc. IEEE Workshop on Wireless Local Area Networks,* Worcester Polytechnic Institute. Pages 140-145.

[Walker,90] Walker, J., and B.R. Gardner. "Cellular Radio." *Mobile Information Systems Series,* Editor: J. Walker. Artech House. Pages: 59-104.