An Architecture for User Privacy in Mobile Networks

Robert John Askwith, B.Sc. (Hons.)

A thesis submitted in partial fulfillment of the requirements of Liverpool John Moores University for the degree Doctor of Philosophy

School of Computing and Mathematical Sciences

October 2000

Acknowledgements

I would first of all like to express greatest thanks to my principal supervisor Madjid Merabti. Throughout the period of this research his advice, encouragement and all round harassment has enabled my progress. I would also like to thank my second supervisor Qi Shi for being patient enough to examine technical details with a fine toothcomb. Also, thanks to my third supervisor Keith Whiteley for support in the early stages of the project.

I would like to thank all my colleagues in the department of Computing and Mathematical Sciences for continued support, especially those fellow researchers who made it more fun than it would otherwise have been, Damien Gregory, David Gresty, Tom Berry, Ewan Smith. The funding for this project was provided by Liverpool John Moores University to whom I am most grateful.

Away from the office I would like to thank all those friends and family who have suffered my inability to avoid technical discussion. It was my father who initially persuaded me to pursue an academic career - like father like son, thanks Dad. The list of friends is too long but special mention is required of Satinder Dharewal, John Cleary and Simon Truss.

Finally I must thank my wife, Sarah, whose seemingly endless sacrifice made the climb considerably easier and the motivation more meaningful. I have a sneaking suspicion Sarah may be the most pleased out of all to see this page in print.

Abstract

During the 1990's the use of electronic communications became a part of everyday life through the explosion of both mobile telephony and Internet use. The next decade promises further rapid development, with multimedia mobile networking and the convergence of existing networks. These developments have not been without impediment, concern over privacy being important amongst these. Privacy in a mobile environment involves the protection of message content, identification and location, as well as service use behaviour. In addition, a user requires this protection against not only parties outside the network but also from the parties operating the network itself.

In this thesis we have addressed the security and privacy requirements within mobile networks. A survey of existing literature shows that these requirements are not currently satisfied. In response to this we present the development of an architecture, the Mobile Network Privacy Architecture (MNPA). The MNPA proposes two novel physical components and two new protocols to perform important network tasks, namely location update and secure billing for services. The two new physical components are the Privacy Routing Capability (PRC), and the Privacy Token Issuing Authority (PTIA). The PRC facilitates network anonymity by enabling untraceable message transport between any two hosts. The PTIA is a third party authority that facilitates privacy-enhanced protocols through the issuing of privacy tokens to the network of each user.

Our analysis and evaluation of the architecture shows that the MNPA protocols and components meet the requirements for achieving privacy. The analysis focuses on potential attacks on the system, including those of collusion between parties. This leads us to examine the most pressing assumption, that of trust between the user and the various components, which must be satisfied in order to complement the technical measures in place. Prototyping experiments are discussed and it is shown how the above system can be implemented. Finally we present a series of proposals for further research work that have been raised by this work, such as user interfaces for privacy, data access for law enforcement and the prevention of denial of service attacks within the MNPA.

Table of Figures

Figure 2.1. The Cellular Concept.....20
Figure 2.2. Wireless Network Architecture.....22
Figure 4.1. Mobile Network Privacy Architecture (MNPA).....70
Figure 4.2. Example Route of a Message Between Two Hosts Using the PRC 77
Figure 4.3. Example Configuration of Six PTIA Elements 84
Figure 6.1. Communications Within the Prototype MNPA 122

Table of Contents

CHAPTER 1: INTRODUCTION 1

1.1 DISTRIBUTED SYSTEMS 2

1.2 MOBILE NETWORKING 4

1.3 COMPUTER SECURITY 6

1.4 PRIVACY 7

1.5 THESIS AIMS 8

1.6 NOVEL ASPECTS OF THIS WORK 10

1.7 SUMMARY 12

1.8 THESIS STRUCTURE 14

CHAPTER 2: BACKGROUND 15

2.1 COMPUTER NETWORKS BASICS 15

2.2 MOBILE COMMUNICATIONS 17

2.2.1 First and Second Generation Systems 18

2.2.2 Third Generation Systems and Beyond 18

2.2.3 Mobile/Wireless Networking Concepts 19

2.3 COMPUTER AND COMMUNICATION SECURITY 23

2.3.1 Security Basics 23

2.3.2 Computer Security 24

2.3.3 Cryptography 27

2.3.4 Network Security 35

2.3.5 Electronic Commerce Security 37

2.4 USER PRIVACY 38

2.4.1 What is Privacy anyway? 39

2.4.2 Privacy Enhancing Technologies 41

2.4.3 The Legal and Political Landscape 43

2.5 SUMMARY 45

CHAPTER 3: SECURITY AND PRIVACY REQUIREMENTS IN MOBILE COMMUNICATIONS 46

3.1 SYSTEM SECURITY: THE SERVICE PROVIDERS PERSPECTIVE 46

3.1.1 Authentication of Users 47

3.1.2 Encryption of Network Traffic 48
3.1.3 Equipment Security 49
3.2 USER PRIVACY: THE USERS PERSPECTIVE 50
3.2.1 Anonymity, Pseudonymity and Different Levels of Privacy Protection 50
3.2.2 Protection from External Attacks 52
3.2.3 Protection from the Network 52
3.2.4 Law Enforcement and Privacy Negotiation 53
3.3 EXISTING SOLUTIONS 55
3.3.1 Privacy in General Networks 55
3.3.2 Privacy in Mobile Communications 56
3.4 SUMMARY 64
CHAPTER 4: THE MOBILE NETWORKS PRIVACY ARCHITECTURE 66
4.1 THE ARCHITECTURE 66
4.1.1 Assumptions 66
4.1.2 Architecture Overview 69
4.1.3 Mobility Management 71
4.1.4 Accounting 73
4.1.5 Private Communications 73
4.2 PRIVACY ROUTING CAPABILITY (PRC) 74
4.2.1 Overview 75
4.2.2 Detailed Operation 76
4.3 PRIVACY TOKEN ISSUING AUTHORITY 79
4.3.1 Overview 79
4.3.2 Subscriptions 80
4.3.3 Tokens 81
4.3.4 Key Distribution 83
4.3.5 Inter-Process Communication 84
4.3.6 Dispute Management 85
4.4 PRIVACY-ENHANCED COMMUNICATIONS 86
4.4.1 Location Management 86
4.4.2 Remote Host Communications 89
4.4.3 Accountability / Billing 90
4.5 SUMMARY 94

CHAPTER 5: SECURITY ANALYSIS OF THE MNPA 96

5.1 THREATS AND VULNERABILITIES 96

5.1.1 Attackers and Attacks 97

5.1.2 Vulnerabilities 99

5.1.3 Trust 99

5.2 ANALYSIS OF THE PRIVACY ROUTING CAPABILITY 100

5.2.1 Protocol Analysis 100

5.2.2 Collusion Attacks 102

5.2.3 Trust 103

5.3 ANALYSIS OF THE PRIVACY TOKEN ISSUING AUTHORITY 104

5.3.1 Subscription 104

5.3.2 Token Issuing 104

5.3.3 Key Distribution 106

5.3.4 Inter-Process Communications 107

5.3.5 Dispute Management 107

5.3.6 Collusion Attacks 108

5.3.7 Trust 108

5.4 ANALYSIS OF THE LOCATION REGISTRATION PROTOCOL 109

5.4.1 Attacks 109

5.4.2 Collusion 110

5.4.3 Trust 111

5.5 ANALYSIS OF THE ACCOUNTABILITY PROTOCOL 111

5.5.1 Attacks 112

5.5.2 Collusion 113

5.5.3 Trust 113

5.6 ANALYSIS OF END-TO-END COMMUNICATIONS 114

5.6.1 Attacks 114

5.6.2 Collusion 114

5.6.3 Trust 115

5.7 SUMMARY 115

CHAPTER 6: IMPLEMENTATION AND EVALUATION 116

6.1 AIMS OF THE PROTOTYPE IMPLEMENTATION 116

6.2 DETAIL OF THE PROTOTYPE IMPLEMENTATION 118

6.2.1 Implementation Environment *118*
6.2.2 Local Network *119*
6.2.3 Home Network *119*
6.2.4 PRC *120*
6.2.5 PTIA *120*
6.2.6 Mobile Subscriber *121*
6.2.7 Service Provider *121*
6.3 EVALUATION OF THE PROTOTYPE IMPLEMENTATION 123
6.3.1 Registration Protocol *123*
6.3.2 Billing Protocol *127*
6.3.3 Prototype Evaluation *129*
6.4 OVERALL PROJECT EVALUATION 131
6.4.1 Comparison with Related Work *132*
6.4.2 Shortcomings of the MNPA *133*
6.5 SUMMARY 136
CHAPTER 7: CONCLUSIONS AND FUTURE WORK 137
7.1 THESIS SUMMARY 137
7.2 CONTRIBUTIONS 140
7.3 FUTURE WORK 141
7.3.1 Integration with Existing Technologies *141*
7.3.2 Application of MNPA to Fixed Networking *142*
7.3.3 Routing decisions within the PRC *143*
7.3.4 Access to Data for Law Enforcement *144*
7.3.5 Usability of Privacy Enhancing Technologies *145*
7.3.6 Reasoning about Trust and Privacy *146*
7.3.7 Configuring User Privacy *146*
7.3.7 Denial of Service *147*
7.4 SUMMARY 148
REFERENCES 150

Chapter 1: Introduction

Modern society recently experienced an enormous and ongoing change in the way in which we communicate, both in our work and in personal lives. This change was the phenomenal growth of electronic communications during the 1990s. First, the Internet grew out of the research community into a peoples network of 370 million [NUA, 2000]. Second, the use of cellular mobile telecommunications grew from just 23 million in 1992 to 600 million [GSMWorld, 2000]. Moreover, it appears that this expansion in utilization will continue for the foreseeable future. Following the deployment of the next generation of mobile communications systems it is predicted that over a billion subscribers will be in place.

However this change has not been totally positive. Increased reliance on computing and communications systems coupled with the inherent rising value of personal information has attracted a similar rise in attention from malicious elements. In addition to the threat of fraud a significant impediment to confidence and uptake of electronic communications is user privacy. The actions we take online inevitably make an electronic footprint that when linked together into an electronic trail threaten to usher in a scenario not dissimilar to the oppression portrayed in Orwell's famously insightful novel '1984' [Orwell, 1949].

The state was the perpetrator in '1984' and to some extent remains one of the biggest threats, despite a global agreement that privacy is a fundamental human right [UN, 1948]. Perhaps a more serious threat to privacy comes from the corporations and their marketing departments who may tap into gigantic databases of user information with scant ethical regard or sufficient restraint. Legal solutions to privacy are very difficult to implement [Agre and Rotenberg, 1998, Froomkin, 2000, Litman, 2000] and it is our view that technical solutions must also exist to support privacy protection. The major aim of a technical solution is to return control of user data back to the user.

Privacy in mobile communications consists not only of the protection of message contents but also the associated data attached to messages. This includes identification, location and behavioural information, such as billing details. In addition to these requirements we add that such privacy must be achieved against not only outsiders but also the network operators. Against this privacy it is a requirement that the operators of the network are protected against fraudulent behaviour using strong authentication and integrity mechanisms. This thesis presents an architecture to satisfy our requirements for user privacy in mobile communications.

The remainder of this chapter is structured as follows. In sections 1.1 and 1.2 we introduce the nature and importance of distributed systems and mobile networking. The next two sections, 1.3 and 1.4, introduce the fundamentals of computer security and privacy from both technical and non-technical views. Section 1.5 sets out the aims of the thesis before we detail the novel results of our work in section 1.6. The chapter closes with a summary in section 1.7 and a description of the thesis structure in section 1.8.

1.1 Distributed Systems

A distributed system is a computing system based upon separate, autonomous but co-operating computers linked by a computer network [Coulouris, et al., 1994]. The aim of a distributed system is to enable the sharing of system-wide resources between individual nodes. Users should perceive a system that looks like a single entity. Various nodes in a distributed system provide services that other nodes may utilize. For example many organizations administer a corporate email server that all other corporate network users have an account with. In turn the email server co-operates with other email servers external to the company in order to deliver email to users whose accounts are administered elsewhere.

The field of distributed systems is a sign of strongly maturing data communications technologies, particularly increasingly available bandwidth and computer processing power in general. The range of applications is constantly increasing, with multimedia services such as video-on-demand [Sumari, 2000]

becoming increasingly possible, as is the move towards electronic commerce. New categories of service frequently add significant complexity to distributed systems due their differing requirements in terms of properties such as robustness, quality of service (QoS), security, and wireless networking [Raychaudhuri, 1999]. Each of these factors of complexity may require new or extended system services and designers are challenged with the task of finding ways to incorporate these into already complex designs.

The key to solving these complexities in distributed systems is the inter-working of the autonomous nodes. The difficulty in inter-working lies in the range of difference found in individual nodes, such as different operating systems, different hardware and different administration policies. These differences can be overcome by standardizing network services. The Transmission Control Protocol & Internet Protocol (TCP/IP) [Socolofsky and Kale, 1991] suite that enables nodes to connect to the worldwide Internet is perhaps the best known of these distributed systems standards.

TCP/IP does not solve all the inter-working problems of the Internet, far from it. In fact, TCP/IP is simply part of a 'stack' of protocols that specify various aspects of interconnection. For a distributed application to operate its communication must be abstracted through various layers of protocols. At the lowest layer standards specify the direct access to the hardware, followed up a layer by control of connection to the neighbouring nodes. At this point co-ordination between nodes is provided by protocols such as TCP/IP. Above this applications may define their own interconnection protocols, such as how to transfer email between nodes.

This layering of protocols is fundamental as it enables flexibility as well as reducing complexity. Flexibility is provided since lower layer protocols are more hardware dependent (enabling different hardware in the system) whilst higher layer protocols are application dependent (enabling different applications). Complexity is reduced since applications only have to assume the existence of other applications operating the same protocols regardless of lower layer

constraints. Each layer in a protocol interacts only with the layers directly below and above. The implication of this is that a user is only aware of the application and the corresponding communications partners, whilst at the lowest level the access protocol is only aware of the underlying hardware and the link layer protocols.

Layering of protocols also allows designers to abstract services to appropriate levels. Other layers can assume the provision of these services or perhaps extend them further depending on the context. Security is one such service that may be provided at various layers and abstractions. The lower layers tend not to provide security services; instead they concern themselves with error correction at the bit level. Network layers may provide authentication and integrity services whilst application layers usually provide end-to-end confidentiality and overall security policies for applications.

1.2 Mobile Networking

An extension of ordinary fixed networking is to allow the movement of nodes within the system, this is known as mobile networking. There are two main developments of importance in mobile networking. Firstly the extension of Internet technology to incorporate mobile nodes and second the expansion of mobile telephony to include data services. Indeed mobile networking is one important aspect of the inevitable convergence between all types of networking so that ultimately it may be that all computers are connected to one logical global distributed system.

As we noted above mobile networks differ from fixed networks in the fact that mobile networks allow nodes to disconnect from one point of attachment and reconnect to another. This convenience for the user means that systems must be able to dynamically reconfigure routing as nodes move within the system, also known as mobility management. In Mobile IP [Perkins, 1997] for example, an association is maintained between a fixed Internet address and a temporary address. Despite the simple description mobility management is one of the most challenging aspects of mobile networking.

Mobile/wireless telephony has been the driving influence behind the development of mobile data networking and will continue to be so for some time. Originally the mobile telephone networks operated as analog voice networks that extended the Public Service Telephone Network (PSTN) by adding a wireless component. In the early nineties second generation mobile telephone networks were developed to pave the way for simple data services through mobile telephone handsets. Since the mid 1990's the vision has been focused towards a third generation of mobile network which can inter-network with data networks as well as telephone systems to provide a single access point for a high-bandwidth multimedia services [O'Mahony, 1998]. The so-called Martini[1] Effect of these systems is to enable any service to be provided at any time and in any location.

To achieve the major goals of the Martini effect means overcoming several major constraints on resources. The most important of these constraints are as follows: first the bandwidth is likely to be scarcer than in wire-line networks, though the gap may close in the future. Second the power availability and computation speed of terminals means that processing must be more efficient. Third, user interfaces are likely to be more compact, which combined with the complexity of mobile access means that well thought out interfaces are needed.

The challenges facing the development of third generation systems are immense and cover not just the technical aspects but also business, legal and political ones. A significant problem being addressed is that of security within these systems [ASPeCT, 1998, Pandya, et al., 1997]. First generation analog wireless networks were notoriously insecure and provided considerable motivation for the design of second-generation systems. However it is recognized that in order to provide adequate security for high value services such as electronic commerce and enough flexibility to distribute these services appropriately then the security of second-generation system is wholly inadequate.

[1] Martini is an alcoholic drink that was once advertised on television using the slogan 'any time, any place anywhere – that's Martini.'

1.3 Computer Security

The fundamental concern of computer security is to provide the trio of services known as Confidentiality, Integrity and Availability [Pfleeger, 1996]. Whilst there are many security services in existing systems, ultimately they all attempt to achieve one or more of these three basic ones. Confidentiality is the protection of the data against unauthorized disclosure, more commonly referred to as privacy. Integrity is the protection of data against modification or deletion, which leads to the enabling of access controls and authentication. Availability is the protection against attacks on the use of data and services, often referred to as denial-of-service or DoS [Needham, 1994a].

Security is a problematic issue for system designers as it is both difficult and costly to design and implement. Costs are incurred in both the traditional sense of having to purchase security components and also due to the constraints it may add to the system in terms of performance and usability. The difficulty of security is well understood by researchers but surprisingly poorly understood by the software community. Indeed, it often seems like every program has at some point been broken at either design level or implementation level. The difficulty in security lies in the seeming impossibility of predicting all possible attacks (and subsequently incorporating defences against these).

There are many mechanisms designers may use to incorporate security into systems, but we are most interested in those for enabling protection in distributed systems. The two most fundamental of these are encryption and authentication. Encryption algorithms produce ciphertext output based upon plaintext input and a (usually secret) parameter known as a key. Symmetric key algorithms use the same key for encryption and decryption of data [AES, 2000, Schneier, 1996]. Two users may share a key in order to establish a secure channel between each other.

Unfortunately this requires a method to distribute keys between parties. This is achieved through the use of asymmetric encryption, or public-key cryptography. Public-key algorithms use two keys; one is kept secret whilst the other can be

made public (e.g. published on a web page). When one party, Alice, needs to establish a secure channel with another, Bob, and therefore share a symmetric key with Bob then Alice can encrypt the symmetric key using Bob's public key. Only Bob can decrypt this symmetric key, therefore the channel is established. The reason that public-key cryptography is not used for all encryption lies in the fact that it is many factors slower than symmetric key encryption [Schneier, 1996].

In addition to encryption, which is concerned mainly with confidentiality, the other major network security mechanism type is authentication. Authentication is concerned with the integrity of data and services. There are many forms of authentication but the overall purpose is to determine the authorization of a particular party by having that party produce some proof of this authorization. A typical authentication technique is to challenge one party to author an encrypted message that only they would be capable of authoring, i.e. using their secret key.

Different applications require different security mechanisms to be used and this has resulted in an enormous variety now embedded into real-world applications. Due to the cost implications of security it is common to see applications, particularly networked applications, that bias their protection towards the system providers perspective. By this we mean that designers often trade off users concerns, most often privacy, against those of the service providers, namely protection against fraudulent and malicious activity. It is important therefore that we understand the notion of privacy, the users perspective in the networked world.

1.4 Privacy

There are numerous activities which people engage in during a typical day, either at work or at home, which they consider private. We might consider these actions private because we feel uncomfortable to be seen doing them, or because knowledge of them could affect other parts of our life. The expectation to be able to perform these activities without the knowledge of others is fundamental to our modern lives. Everyone has different expectations depending on who they are but this does not affect the underlying principle; that there needs to be a private

sphere in which we can live part of our lives [Froomkin, 2000, Warren and Brandeis, 1890].

As we live more and more of our lives electronically then there needs to be a similar ability to protect a space for actions that we wish to keep private. Unfortunately this space seems to be hard to realize for many people. Governments seem unwilling to legislate adequately for data protection and the corporate world values information about potential customers above and beyond the protection of those same customers. Indeed it is arguable that only through intense pressure by civil rights groups that progress is ever made.

Technical measures, known as privacy enhancing technologies (PET's), offer a partial solution to the problem of electronic privacy. Encryption is perhaps the most important tool for privacy protection, allowing data to be either transferred or stored without attackers being able to observe the contents of data. However, encryption is not the end of the story for privacy since associated data may also be sensitive, for example identification data. In email, for example, it might be desirable to send messages anonymously. Other activities we perform online such as accessing services may also carry sensitive details which we would rather were private.

The range of PET's available is also growing as the range of applications in need of protection grows. Perhaps the most important PET's are those that facilitate anonymity. The most common mechanisms for achieving anonymity are based on the idea of routing a message through several intermediate nodes which strip the massage of identifying data, thereby disallowing the recipient, or attacker, from determining the source – i.e. anonymity.

1.5 Thesis Aims

It is the aim of this thesis to describe a framework of privacy-enhancing technologies that combine to provide a mobile network user with capabilities to protect their privacy, should they choose to do so. In order to do this we must first address two issues:

- What are the requirements for privacy within mobile networking? Any requirements should take into account both the user perspective for privacy and the system provider perspective of ensuring fraud prevention and a commercially viable service.

- What existing techniques are appropriate in developing solutions for these requirements? The literature survey shall examine the building blocks of privacy enhancement as well as those efforts that have contributed directly to the knowledge about mobile networking security.

The design of a framework should allow the user to perform all the typical tasks involved in mobile communications, but enable them in a privacy enhanced way. Service providers should be confident that these tasks can be achieved in an authenticated manner. In particular the following are required:

- Mobility shall be allowed to occur without the privacy of the user being compromised. The network will be able to authenticate a user despite their privacy requirements.

- Whilst roaming within the network a user shall be able to communicate with other users without any compromise of privacy, to either the network or the other users, depending on the privacy required.

- Provision of services shall be enabled such that a user may access anonymously. Services that require payment shall not require a user to compromise privacy. Service providers shall be assured that despite anonymity, payment will occur.

Finally the framework should enable us to better understand the complex problem of privacy, both generally, and specifically in terms of mobile networking. The framework should allow us to ask further research questions regarding the field.

1.6 Novel Aspects Of This Work

The problem of user privacy in distributed systems is rarely dealt with in a serious enough manner. During the design of such systems the primary concern when examining security is to prevent attacks from a system perspective. Security is difficult to achieve and privacy may be seen to complicate this. Additionally, privacy appears not to be very well understood. Privacy is often carelessly thought to be simply the protection of message content against external attackers. This sorry state of affairs is unfortunate as it is becoming increasingly clear that privacy is a key factor in preventing many people embracing much of the networked applications that designers and commercial interests see as exciting.

One of these exciting applications is the third-generation mobile network, currently in development. These networks will offer new high-bandwidth multimedia services and allow convergence with other networking technologies such as the Internet. We suggest that privacy will be a major user requirement in third-generation systems, but that privacy must live in harmony with the seemingly opposing requirements for system security. This thesis contributes to our understanding of privacy in mobile environments in the following ways:

- Our first contribution is to provide a set of requirements for privacy in mobile networking environments and examine these against existing research literature [Askwith, et al., 1997]. These requirements enable the network providers to operate a secure system whilst allowing the user of the network to maintain very high levels of privacy against both external and internal attackers. A survey of research literature in the field revealed that no results completely meet these requirements. These techniques focus on either specific application domains or on particular parts of the system security. Others fail because they aim for a lower level of protection against internal attacks. Additionally we also bring together relevant ideas of use in the search for user privacy in mobile environments.

- Using the set of requirements and inspiration from relevant literature this thesis proposes a novel solution to afford strong user privacy to mobile communications users, the Mobile Network Privacy Architecture (MNPA) [Askwith, et al., 1998]. The architecture consists of two physical components that extend the mobile networking paradigm and two new privacy-enhanced communications protocols (many more are possible but these two are most important).

- The first new physical component we have developed is the Privacy Routing Capability (PRC) [Askwith, et al., 2000b]. The PRC allows messages to be sent across the network without being able to be subject to unauthorized tracking. The PRC is an extension of previous work in anonymity, designed to suit the mobile environment and the MNPA.

- The second physical component we have developed is the Privacy Token Issuing Authority (PTIA) [Askwith, et al., 1998] [Askwith, et al., 2000a]. The PTIA is a novel third-party distributed application that manages subscriptions from network providers on behalf of mobile users. The PTIA facilitates behavioural privacy for users by allowing users to access the network anonymously, yet remain accountable. We believe the PTIA to be a particularly novel approach to solving privacy-enhanced network access.

- One of the most important network management tasks in mobile environments is location update management. A new protocol for allowing secure location update management is provided within the MNPA [Askwith, et al., 2000a]. This protocol enables a user to connect to a network without identifying itself to the point of connection, and without revealing location information to the home network.

- Billing is a particularly important aspect of modern distributed systems as the range, complexity and value of services develop (and thus the service-provider costs increase). The MNPA provides a novel protocol post-

payment of services within the mobile network [Askwith, et al., 2000a]. Our billing protocol allows a user to receive service from a service provider without having to identify itself, yet simultaneously enabling post-payment through the home network.

- Analysis of our results examined the possibilities for attack within the MNPA. A range of attack types are categorized and each component is shown to be acceptably resistant to these. We also discuss the additional problems of trust and collusion between entities. Trust and collusion are serious problems since many new attacks potentially become possible if components of the system do not act honestly or competently. Analysis of these issues reveals that with reasonable understanding and knowledge a user may trust network and service providers within the MNPA. The analysis we conducted during development is interesting for the wide view we have taken. Rather than concentrate the analysis on formal examination we have attempted to use a more relaxed but wider approach that tries to determine trustworthiness rather than absolute security. We believe this could lead to a more practical and flexible (in terms of design) solution for the problem.

- Our final contribution is that this research poses some new questions that had not been made explicit before. Among the questions for further work are issues of application to existing mobile environments and potential application to fixed networking environments. Two other important issues raised are those of making complex privacy usable, and related to this is the difficulty of modeling trust within systems. These questions are examined together with an evaluation of the project in terms of the shortcomings of the architecture and a comparison with closely related work.

1.7 Summary

Media coverage of security incidents is often hysterical, giving rise to concern about ones safety in the online world. Whilst it is true that coverage and

subsequent concern often focuses on the potential financial losses caused by malicious attackers significant weight is also given to privacy. Privacy violations may come from many sources, such as external attackers (often called, somewhat misleadingly, 'hackers'), corporations, and government agencies. Whilst there are numerous commercial interests proposing self-regulation schemes and governments attempt privacy legislation, consumers appear to remain extremely worried about privacy. This worry is likely to translate into reluctance on the part of the ordinary user to immerse themselves into cyberspace as much as they might (whether this is a good thing or not is left to the reader to decide).

With this in mind we set out on this project to examine technical solutions to creating that relate specifically to the privacy of global mobile networking. This chapter has introduced the problem area and presented an outline of our results. We began by introducing distributed systems, before moving on to discuss the special nature of mobile networking. This was followed by an outline of the pertinent aspects of computer security, before we briefly looked at the nature of privacy.

Having addressed this background work the aims of the thesis were presented. These aims range from setting out requirements for user privacy in mobile network, through developing a solution capable of meeting both mobility management and service provision requirements of network operators, towards ultimately contributing a new understanding of the of the initial problem domain.

The remainder of the chapter was devoted towards a summary of the novel aspects of the work. Our contributions to the knowledge were stated as, firstly, an examination of the requirements for security and privacy in mobile networks against a comparison of existing literature. This was followed by the introduction of the Mobile Network Privacy Architecture (MNPA), which encapsulates our solutions to mobile network privacy. The MNPA contains two new physical components, the Privacy Routing Capability (PRC) and the Privacy Token Issuing Authority (PTIA). These components allow the implementation of novel privacy-enhanced protocols for location update and post-payment billing. Analysis of the

MNPA enabled us to make statements about the trust requirements in mobile environments. Finally we propose a series of new research avenues resulting from the development of the MNPA.

1.8 Thesis Structure

The remainder of this thesis is structured as follows:

- Chapter 2 presents the background to the field looking at computer networking, mobile networks, computer security including cryptography, and finally privacy and its associated protecting technologies.

- Chapter 3 examines the requirements for security (and privacy) from both the system and user perspective. Following this we examine the literature to determine the current state of the art in mobile network security and privacy.

- Chapter 4 presents our architecture, the Mobile Network Privacy Architecture (MNPA). First we give an overview before detailing the entities and protocols involved.

- Chapter 5 provides an analysis of the MNPA. The elements of the MNPA are analyzed in terms of attack potential, collusion opportunities and trust requirements.

- Chapter 6 is an evaluation of the MNPA that includes discussion of prototype and overall project evaluation.

- Chapter 7 looks back at the achievements of the thesis and concludes what we have learnt from field of user privacy in mobile networking. The chapter is then able to pose some further research questions.

Chapter 2: Background

This chapter examines the background to our research. First, we shall set the scene for computer networks in general including their uses and why they have become such an important part of everyday life. Following this we move on to look at mobile communications and particularly the evolution from early systems to the intended future generations soon to appear.

Having considered the background to computer networks we then look at computer and communications security. We outline the basic goals of security before discussing computer security, cryptography and communications security. Then we take a detailed look at privacy, the techniques used to enhance privacy and the legal and political outlook for privacy. We close the chapter with a summary of networking and security.

2.1 Computer Networks Basics

Computer networks are ubiquitous in the modern 'information society' although their importance has to some extent appeared very suddenly. The Internet [Leiner, et al., 1997] began in the late 1960's yet only in the last five years, particularly with the advent of the World Wide Web, has it become a part of everyday life. Similarly with telephony, the age of digital mobile communications from the early 1990's has revolutionised the way we think about communicating with others.

The Internet has been shown to grow at an exponential rate (though it is not clear how long this can continue) [MIDS, 1998]. Currently there are estimated to be around 370 million users (September 2000) [NUA, 2000]. The huge growth in cellular mobile communications has been corresponding and there were estimated to be around 650 million users [GSMWorld, 2000] in early 2000. So, having established the popularity of computer networks we need to examine what the basics of networking involve.

Essentially a computer network is any set of autonomous computers that can communicate with each other via some medium. We term a computer that is connected to a network a host and the connections between hosts we term links. The media used to make network links can range from copper cables (typically found within office networks), optical fibre (typically for connecting networks), and microwave links (for wireless/mobile communications).

In order to allow hosts to communicate on the given links the hosts must implement a set of protocols, such as the Open Systems Interconnection (ISO/OSI) [Halsall, 1996, Tannenbaum, 1996] or the Transmission Control Protocol and Internet Protocol (TCP/IP) [Socolofsky and Kale, 1991]. The protocol stack is created which places one abstraction on top of another such that it becomes relatively simple for programmers to implement networking in their applications. An important concept in networking handled by these layers of the protocol stack is addressing. Each host in a network must, in order to communicate usefully, be addressable by other network hosts.

The simplest network is the Local Area Network (LAN) that has few users and covers only a small area administered by a single authority. Interconnection of LAN's is generally termed Wide Area Network (WAN), if control is still under a single authority or more commonly an inter-network where the control is federated. Inter-networks tend to be a heterogeneous mix of hardware and software rather than a series of similar hosts. The most well known inter-network is the Internet.

Today's Internet grew out of DARPA research projects in the late 60's. It is now in excess of 370 million users strong and growing at a considerable (and steady) rate. Initially the main uses of the Internet were email (electronic messaging), Telnet (remote access to computer systems) and FTP (file transfer protocol). Now the major service is the World Wide Web, or WWW, which many would argue has been the catalyst to getting people online. The combination of the WWW, and the Internet in general, with business activities has led to the arrival of Electronic

Commerce. Many believe that the success of the Internet is key to the economic growth of many nations. We do therefore, live in a 'dot com' world.

Networks like the Internet allow the creation of Distributed Systems that allow the processing and storage load of a system to be placed at the convenience of the network rather than the user, whilst keeping the distribution transparent to the user.

Computer networks allow users to share computing resources more freely, distribute workload, and create more reliable and easily expandable systems. Looking more broadly, networks and particularly the Internet, allow users to conduct their business over great distances without the need for face-to-face contact. They can bring together people with similar interests from all parts of the globe and spread information almost instantaneously.

As we move forward into the twenty first century the uses of the Internet will continue to increase as more and more bandwidth becomes available. Perhaps more significantly, communications networks in general will begin to converge so that the Internet becomes the carrier for telephony as well as computing data based services. Part of this convergence will involve increased mobility for users, which we look at next.

2.2 Mobile Communications

Communications that are not bound to a physical location are very attractive to users. Although mobile communications predate computers (radio communications were in fact 100 years old in 1999 [Morton, 1999]) it is only since the late 1970's that people have been able to connect mobile devices to computer networks. Initially this was brought about by the development of the Cellular concept by AT&T but more recently this has moved into digital telephony such as the GSM network and mobile connection to the Internet via the MobileIP protocol. This subsection looks at the evolution of mobile telephony since it is envisioned that future networks will become integrated as general data networks.

2.2.1 First and Second Generation Systems

The Cellular concept developed by AT&T [MacDonald, 1979] solved the problem of sharing limited radio spectrum amongst large numbers of users by splitting the network into geographic 'cells', which then share the spectrum over a smaller number of users. The first generation cellular systems were analog in nature in that they could only transmit analog voice information rather than digital data. The dominant standard was called AMPS.

Towards the late eighties moves were being made to introduce a second generation of mobile communications system that still operated in a cellular manner but transmitted digital data. Digital systems enable more sophisticated telephony services to be provided. A particular advantage to such a system is increased fraud protection through the use of encryption over the air interface, which we shall discuss in the next chapter.

Once again no single standard has been adopted worldwide although the European standard Global System for Mobile communications (GSM) [Rahnema, 1993] is dominant. GSM, initially a pan European standards effort, now has in excess of 360 million users (August 2000) [GSMWorld, 2000] in many countries throughout the world and is considered the most important standard. Particular interest is paid to GSM since it is likely to form one of the most important stepping stones to future third generation systems.

2.2.2 Third Generation Systems and Beyond

Over the last few years considerable effort in both the research and commercial communities has been channelled towards the developments of standards for future mobile communication systems, generally known as third generation systems (or 3G systems) [Groves and Clapton, 1996, Kazovsky, et al., 1998]. The principal aim of 3G is to provide 'any service, any time, anywhere'. In practice this means a permanent global reach for high bandwidth multimedia.

Now, in 1999, these efforts come in two flavours. The international standards effort is named IMT-2000 [Pandya, et al., 1997], or International Mobile Telecommunications 2000, whilst the European effort is termed UMTS [Cullen and Lobley, 1996] [O'Mahony, 1998] or Universal Mobile Telecommunications System.

Unlike GSM and other standards, these standards are more concerned with laying out a framework for implementations of 3G systems to fit into. This involves detailing aspects such as what services are possible and how systems should interconnect. A consequence of this approach is that existing systems will be able to evolve into 3G systems rather than be replaced.

Introduction of UMTS/IMT-2000 is likely to begin very soon, though considerable effort is being targeted for 2002 with major deployment of full systems by 2005. In early 1999 commercial interests such as Ericsson began performing local trials of UMTS services in the UK.

In terms of the future beyond 3G systems it is difficult to predict since it is dependant on the relative successes of 3G systems. Two things seem particularly likely though. Firstly that the penetration rates for data communications of any type will continue to rise and secondly that convergence will occur between computer networks such as the Internet, especially in view of the recent development of Mobile IP [Perkins, 1997], and telecommunications networks such as GSM.

One thing is for certain; any fourth generation network will then have to be 'all things to all people'. The shape and form of these is left truly to the imagination, and beyond.

2.2.3 Mobile/Wireless Networking Concepts

Mobile networks are ones in which it is possible to change the point of attachment of a particular (or all) hosts. Wireless networks are a special case of mobile network in that the point-of-attachment is performed via transmission of

radio waves rather than the more familiar cabling connection. Wireless communications takes place by utilizing select radio frequencies from the entire spectrum. The range (spectrum) of frequencies is finite which has led to regulatory frameworks by government to determine access. Spectrum is therefore a scarce resource. A major breakthrough in wireless communications allowing mass use of the spectrum was the development of the Cellular concept by AT&T. This was the precursor to the first analog mobile phone system (AMPS).

Cellular systems partition the geographical area into hexagonal cells, each of which has a range of the available frequency to work with. By allocating different frequency ranges to a cells' six neighbouring cells then signals from no two cells interfere with each other. Repeating the seven cell clusters across the entire geographical region enables the operators to reuse the available spectrum. See figure 2.1 below.

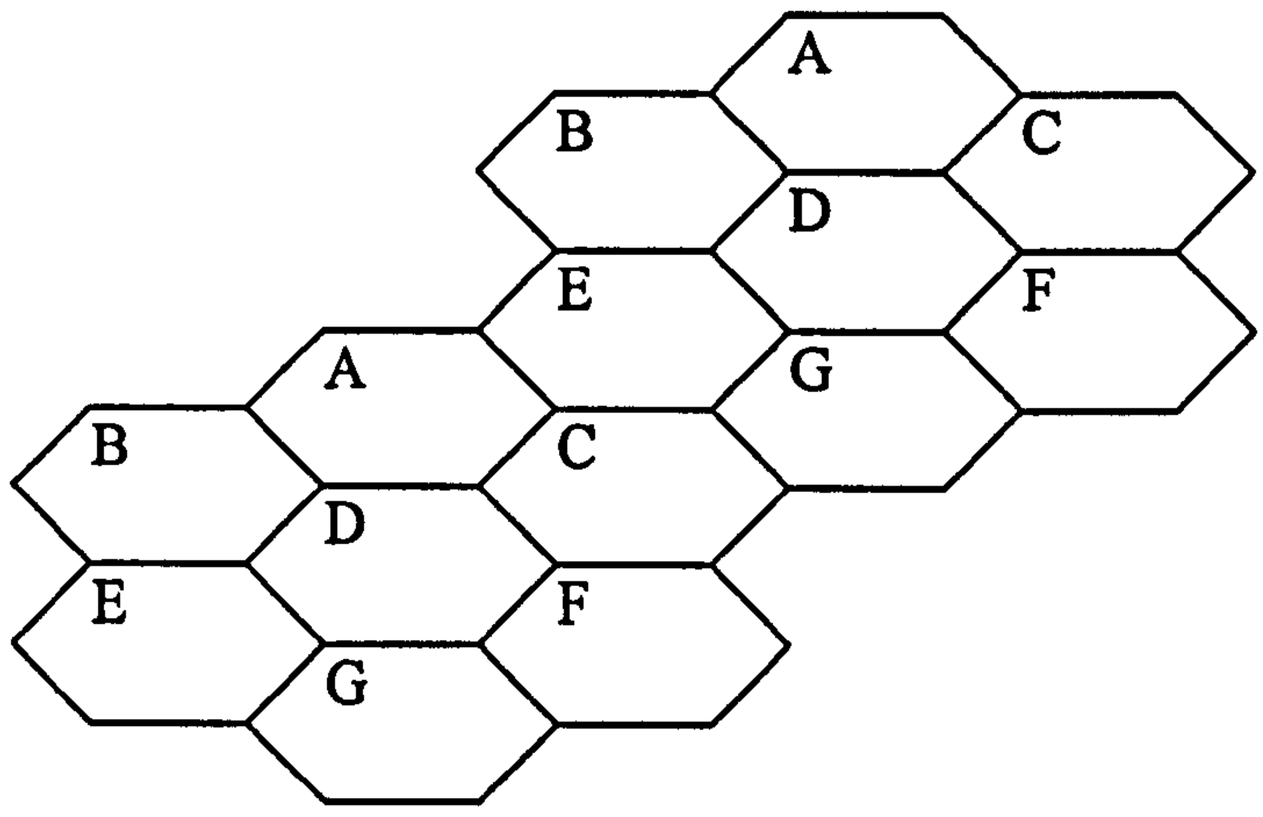


Figure 2.1. The Cellular Concept.

In Figure 2.1 we denote each frequency range with a letter ranging from A to G. Each range is used exactly twice and no letter appears adjacent to a cell containing the same letter. No two cells using the same frequency can thus interfere with each other.

There are several major types of wireless network, the aims and uses of which vary considerably. The most basic is the paging network. This operates only as a broadcast network with paging terminals only able to receive messages. This

makes for a simple and cheap to operate network, but at the general cost of usefulness. Another simple type of wireless technology is the cordless phone. These operate under the principle of using short-range base stations catering for single users. Cordless telephones are cheap and uncomplicated but are inflexible in terms of accommodating users and allowing roaming.

Roaming is better catered for in cellular telephony systems such as GSM. The aim of GSM (and similar systems) is to cater for a mass market. This results in a requirement for maximizing the cell capacity. The cost of this is expensive, low bandwidth, high complexity terminals, with long delay. Wireless data networks, or Wireless LANs are intended to offer a wireless option to the typical fixed desktop computer. This is most applicable where wireless is required but the likely mobility is low (e.g. same building). The final type of wireless technology is Satellite. In these systems there are only a few, very expensive, base stations (i.e. the satellites) but these cover a vast area, possibly global.

Regardless of the type of network involved the architecture is essentially the same. A Mobile Terminal (MT) connects to the local Base Station (BS) over the radio interface. Each BS is connected to a Mobile Switching Center (MSC) that controls the interface to the backbone internetwork and performs routing within the domain. The MSC maintains a register of the locations of all its registered users and their current locations in a database called the Home Location Register (HLR). A second database, entitled the Visiting Location Register (VLR) records the identities of all MTs within the geographical control of the MSC. A series of MSCs may be connected over an internetwork, e.g. PSTN. This is shown in figure 2.2 below.

The management of wireless/mobile networks consists of three major problems;

- **Radio Resource Management.** Since the resource is finite it must be carefully allocated to the users in the system. Radio resource management is mainly controlled at the base station and is the wireless part of a network (rather than the mobile part). There are three basic schemes for

radio resource management; TDMA, FDMA and CDMA. See [Li and Liao, 1997] for further discussion.

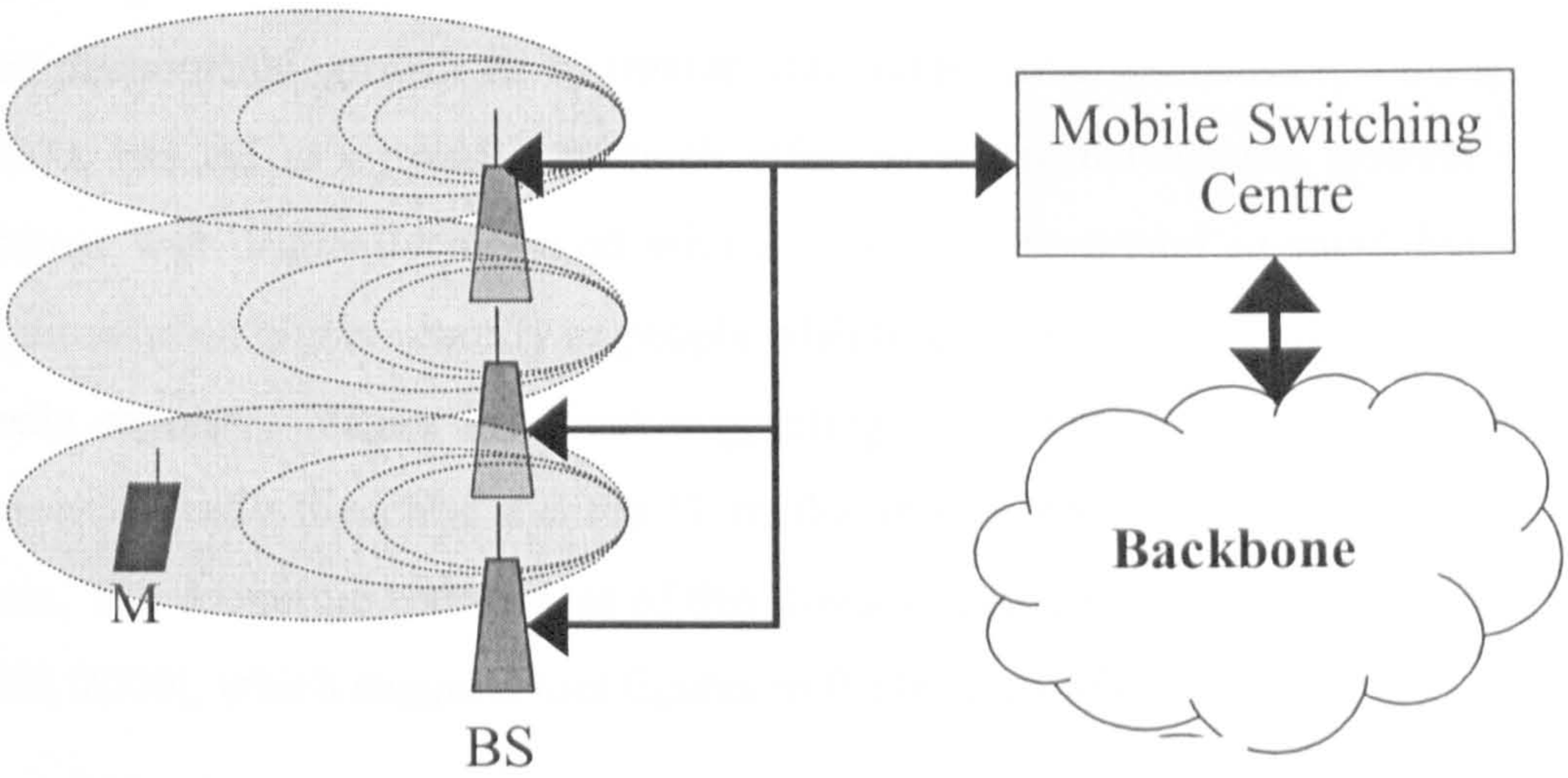


Figure 2.2. Wireless Network Architecture.

- **Mobility Management**. The primary difference in mobile networking over fixed networking is the ability for hosts to easily alter their point of connection. The first problem this poses for the management of the network is actually locating users (Paging). When a user moves between cells two issues need resolving; how to decide when to connect to a new base station (Location Update) and then actually managing the transfer of the connection (Handover). Finally the wider problem of handing over connections between wider areas, i.e. between MSCs or inter-domain is called Roaming. See [Akyildiz, et al., 1999] for further discussion.
- **System Management**. The remaining problems in managing mobile networks lie in the overall system issues. There are three particularly difficult problems for system management. Firstly the naming scheme must allow users to be identified with a meaningful static address yet allow the logical address to remain dynamic. The second problem is implementing efficient signalling protocols to interconnect with the backbone. Finally, and the concern of this research is that of securing the network. Security must provide privacy for the user and fraud protection

for the network. We look at the requirements in greater depth in the next chapter.

2.3 Computer and Communication Security

The phenomenal growth in computer communications, particularly during the 1990's, has led to a massive research effort to secure networking systems. The Internet was originally designed with a low security model in mind but now requires much higher security as people wish to conduct business over it. General media regularly feature headlines suggesting that there is a regular stream of serious security breaches and the IT media reports security news on a weekly basis. The economic seriousness of this threat is borne out by empirical research [CSI, 2000], which suggests loss figures in the billions of dollars per year.

As we progress towards integration of networks and increasing the range of high value services such as electronic commerce there is a strong desire amongst commerce and users to solve the security issues. Whilst the majority of breaches tend not to affect the ordinary user trust is required in order to progress. This section looks at what security is, and how computers and communications can be secured.

2.3.1 Security Basics

Security of computer systems attempts maintain three properties of the given system; **Confidentiality, Integrity** and **Availability** [Pfleeger, 1996, Voydock and Kent, 1983]. These properties may be compromised through **threats** that exploit **vulnerabilities** in a system. Note that threats do not necessarily have to involve intruders but could indeed come from accidental compromise or from natural sources such as power outages. The three security goals can be described as follows:

- **Confidentiality:** The resources of a computer system should remain *readable* only by authorised users. The most common reference to confidentiality is to applications such as email, where only communicating parties may read messages, but it more generally refers to any resource that requires read

access. The threat to confidentiality is interception, and the main vulnerabilities are weak access controls to computers and insecure communications links.

- **Integrity**: The resources of computer systems should remain *writeable* only by authorised parties. In other words unauthorised parties should not have write access to resources. In terms of communications then any received message should be the same as the sent one. The threats to integrity are either based on modification or fabrication of data. The vulnerabilities are similar in nature to those for confidentiality.

- **Availability**: The resources of a computer system should remain *accessible* to authorised users. In addition to basic access, general properties of availability are timeliness, fair allocation, fault tolerance and usability. The threat to availability is interruption (often termed Denial of Service or DoS). The vulnerabilities for availability are often very different to those for confidentiality and integrity since exploiting valid actions of a program enable many such attacks. For example, by bombarding a Web server with requests an attacker can at least slow the server down and possibly cause it to become overloaded and halt totally. An interesting discussion of Denial of Service is made in [Needham, 1994b].

In order to achieve security in a computer system those responsible must first assess the threats and vulnerabilities and identify a set of controls to implement. However, like in any business activity, risk must be assessed in order and balanced against the cost of the measures. There is no economic sense in implementing controls that will cost more than any cost of compromise. The next three subsections deal with the controls for computer and communications security.

2.3.2 Computer Security

Computer security relates to the protection of hardware and software on individual computers. Hardware may be protected by controls that relate more to

other types of physical security. Software is more complex to secure and is best considered as operating system security and application security.

Operating Systems: The most critical software on a computer is the operating system since it has the highest degree of access to the system resources. If an attacker can compromise the operating system of a computer then he may be able to easily perform devastating attacks on the systems confidentiality, integrity and availability. Indeed there is inevitability about security failure if secure operating systems are not in use [Loscocco, et al., 1998].

Three major controls in operating systems are memory protection, access control and user authentication. Memory protection involves maintaining address space for programs whilst access control is the more general case for system object protection. Access control falls into three main types; mandatory, discretionary, and role-based. Discussion of access controls can be found in [Sandhu and Samarati, 1996, Sandhu and Samarati, 1994].

In order to implement access control the operating system must be able to determine the authority of the user, this is the problem of user authentication. Typically users are authenticated via passwords [Abadi, et al., 1997, Morris and Thompson, 1979] though the use of Biometrics is an emerging field with potential to combat many of the problems of password-based authentication. Biometrics are measurements of unique physical attributes of a person such as fingerprints, voice and face recognition.

Since no system can be said to be 100% secure (quite often due to the human element of a system) there has been an effort to try to design Trusted Operating Systems. Such an operating system tries to posses trust rather than achieve a level of security. This trust is based on the presence of features appropriate to the intended use of the system and to any assurance given by external analysis. In both Europe and North America government assurance schemes, ITSEC [ITSEC, 2000] and TCSEC [NIST, 2000] respectively, are in place to grade operating systems according to sets of rigorous criteria. An attempt to harmonize these

schemes called the Common Criteria [CCITSE, 1996] has recently been developed.

Application Security: In order to be useful a computer system should allow users to run application programs. However applications may contain bugs that can be exploited by an attacker or may even be malicious in nature, such as viruses.

Viruses are programs that attach to other programs with the intention of causing some unintended actions on the computer system. These may spread by any method that data can transfer and may perform any action possible for that system. As such they present a very difficult problem for security. Applications called virus-scanners are used to try to detect their presence. A recent virus that received much publicity was the Melissa virus [CERT, 1999].

Other malicious code, or Malware, includes trojan horses (programs that perform undocumented secondary tasks), logic bombs (programs that perform event driven malicious actions), trapdoors (undocumented access point into applications) and worms (self-replicating network aware code). Perhaps the major problem surrounding viruses and malware is the human element. Since the explosion of Internet use people feel free to download software from a variety of sources without any consideration as to the origin and potential effects. This underlines the idea that systems include the people. It is imperative that good security practice, through education, becomes part of the culture of computing and communications technologies.

A particularly interesting class of malicious code threat that has become more important recently is mobile code. Programming languages such as Java allow programmers to create code that can move between hosts in a network. Much work is being done in addressing the security issues of mobile code [McGraw and Felten, 1998, Rubin and Geer, 1998].

Another major area in program security is Database security. Database security must be maintained in a similar manner to other software, however the complex

nature of databases makes them a special case. The logical structure needs to be maintained as well as the integrity of individual elements. Extensive logging and complex access control can control these.

A unique problem in databases is inference that is to say the deduction of sensitive information based on other related (less sensitive) information. Concealing or suppression can control this. Concealing can be achieved by returning query results that approximate or hide the sensitive values whilst suppression involves rejecting attempts to query data that might lead to sensitive values being inferred.

Of course not all security compromises are due to malicious activity. Poor programming and accidental events caused by users are undoubtedly the biggest risks to systems [Risks, 2000]. The recent and ongoing phenomenon that is 'Y2K' (Y2K is the affectionate name for date related computer bugs related the roll over from 1999 into 2000, caused principally by using two-digit codes for storing year data) is testimony to this. Jones provides an interesting survey of date related problems in [Jones, 1998].

We have stated that computer security does not attempt to achieve the goal of 100% security, as this is unrealistic (and indeed not helpful as an engineering aim). In recognition of this much research has been carried out in detecting intrusions, either after the event or more proactively in real-time. Such Intrusion Detection Systems (IDS) [Graham, 1998, Mukherjee, et al., 1994] examine system activity and report on either anomalous or misuse behaviour. The nature of attacks on systems is often complex and detecting them is non-trivial. Recent results have been made using Mobile Agents to distribute the detection process across the network [Gregory, et al., 1998].

2.3.3 Cryptography

Cryptography is the science of secret codes. Although as a science it has been in use since at least Roman times, it is since the birth of computers that it has become of major scientific interest. Indeed, it has been said that the British allies

would likely not have won the Second World War had it not been for the efforts of the cryptographers at Bletchley Park [BP, 2000] who cracked the German Enigma machine using some of the first computers ever built.

It is perhaps no surprise then that governments, or more precisely military and secret service agencies, still research cryptography and indeed it is one of the more political of computing fields with frequent controversy surrounding its use and development.

In this section we shall discuss the basic types of cryptography, keeping in mind our research topic. We shall discuss both symmetric and asymmetric encryption, how encryption can be used for data authentication through Digital Signatures and finally how cryptography is applied in systems and its problems. Readers are referred to [Schneier, 1996] for an in depth treatment of Cryptography. An excellent survey of the mathematical underpinnings is presented by both [Rivest, 1990] and [Goldreich, 1997].

Basic Encryption Terminology. An encryption function outputs ciphertext based upon some plaintext input and a second parameter known as a key. Thus if each user chooses a different key then the same algorithm may be used by all users without the same mapping being generated. Knowledge of the secret key allows the authorised user to access the plaintext given the ciphertext using decryption. Typically decryption is the same algorithm as encryption.

The length of the key determines the attackers search space, so that if a 128 bit key is used the maximum number of possible keys is 2^{128}. Such a number is considered large enough to deter even the thought of a brute force attack on most common secure secret key algorithms. The search space may not be quite of this order depending on whether some numbers are not safe to use as keys. There are two classes of encryption, symmetric (single key) and asymmetric (two keys). It is important to note that key length isn't necessarily a good metric of security since many other factors contribute to overall security.

Symmetric Encryption. This class of cipher requires a single key to operate both encryption and decryption. The implication of this is twofold, the key must be kept secret to prevent unauthorized parties from accessing messages and also that two parties must be able to share a key in order to communicate securely.

Today, the most common implementation of symmetric encryption is the Data Encryption Standard (DES) [NBS, 1977]. This was developed by IBM in the 1970's as the US encryption standard and is still in use throughout the world today. Despite this DES is widely considered to be at the end of its useful lifetime. This is mostly due to the key-lengths used in DES, either with 40 or 56 bits. The former can be cracked in hours by a single machine whilst the latter has recently be attacked within 24 hours by a team of activists who built a specialist machine called DES Cracker [EFF, 1997].

The implications of this are far reaching. Firstly all exported products containing DES were originally required to run at 40 bit key lengths whilst domestic use permitted up to 56 bits. Recent alterations to the US regulations appear to de-regulate the exportation of strong cryptography, with certain conditions. Second, if a team of researchers with a moderate budget (DES Cracker cost $250,000) can crack 56 bit DES then it is easy to suggest that a more determined attacker, for example a large multinational organization, might already have more powerful machines. It has been speculated that the National Security Agency (NSA) in the USA has many machines capable of cracking 56-bit DES in minutes. Whether this is true is hearsay, but the possibility remains high.

In 1996 efforts began to replace DES with another more secure algorithm, to be known as the Advanced Encryption Standard, or AES. This will operate at 128 bits and above and will be more flexible in its design making it suitable for many applications. In early 1999 the selection procedure was well under way which involved lengthy peer-review of a series of proposed algorithms by many of the worlds top cryptographers. Such an open process is encouraging and should result in a standard that suits everyone (except perhaps the security agencies). Within a matter of days prior to the completion of this thesis NIST announced that the

winner of the AES selection process was the Rijndael algorithm, developed by a research team based in Belgium [AES, 2000].

Asymmetric Encryption. Also (more commonly) known as Public Key Encryption, this class is distinguished by requiring two keys, one for encryption, and the other for decryption. Public key encryption works by keeping one key secret (the *private key*) and publishing the other (the *public key*). To send a confidential message you encrypt the message with the recipient's public key.

Public key cryptography was first published in 1976 by Diffie and Hellman [Diffie and Hellman, 1976] with the first practical scheme appearing a year later, known as RSA after inventors Rivest, Shamir and Aldeman [Rivest, et al., 1978]. It has recently come to light that the British government security research department CESG came across asymmetric encryption in the 1960's [Ellis, 1997]. Due to the military nature of the research this was kept secret at the time.

The RSA algorithm is generally accepted as the most secure practical public key scheme and is widely used. Despite 20 years of analysis no serious flaws have been found [Boneh, 1999]. The strength of RSA lies in its simplicity and that the problem is believed to be as hard as factoring prime numbers (NP-complete). However, patent issues surrounding RSA have meant that the Diffie-Hellman scheme is more popular amongst many programmers. Diffie-Hellman is believed to be similarly secure. The patent on RSA expired in September 2000 so this situation may change.

Another technique for public key encryption is that of elliptic curve cryptography [Menezes and Vanstone, 1993] [Araki, et al., 1998]. The importance of elliptic curve cryptography is their potential efficiency due to reduced key lengths, making them particularly suitable for power-limited mobile devices. Currently elliptic curve cryptography is in the research stage and confidence about the security of these schemes is not enough to deploy widely in real systems.

In mathematical terms public key methods are one-way functions that have a private backdoor. For the user this means that if you publish one key (the *public key*) and keep the other secret then other users can encrypt messages using your public key which will enable only you (the holder of the *private key*) to decrypt them.

This has the advantage that users need not share a secret key before communicating. This allows complex distributed protocols to be produced that would possibly not be possible using symmetric encryption alone. However, two disadvantages overshadow the use of public key cryptography; speed and trust.

Public key methods are typically many orders of magnitude slower, e.g. DES is in the region of 1000 times faster than RSA [Schneier, 1996]. This is due to the computational operations involved, typically modular exponentiation to large powers (RSA uses between 512 and 2056 bit exponents). Whilst this means that it cannot practically be used for message encryption it has found great real world use in protocols to authenticate users and for symmetric 'session' key distribution. Symmetric keys, typically used only for one communication 'session' can be distributed by encrypting them with the recipient's public key before sending to the other party. A well-known application of such session key distribution is used in the secure email program Pretty Good Privacy (PGP) [PGP, 1999].

Certification. Trust becomes a problem in public key cryptosystems because although two users can share keys without having met neither party can determine anything about the other party. In other words there is no secure binding between a users claimed identity and the public key. This might not make immediate sense, but consider an attacker who simply replaces a users public key with his key. If such an attack were to succeed then the attacker could receive subsequent session keys.

This problem can be relieved by the implementation of a certification authority (CA) [Chokhani, 1994] [Perlman, 1999]. The function of a CA is to provide the necessary bindings between public keys and identities (and usually several other

credentials such as organization, but we ignore this technicality). Thus a CA 'certifies' a users public key. This certification is achieved when a CA digitally signs public keys (we discuss digital signatures shortly). The public key of the CA is widely known to its users. There are several major problems with CA's, specifically; structure, revocation [Cooper, 1998] and real-world implementation. A particular problem in electronic commerce with certificates is the common practice of linking identities rather than permissions to keys, as is more commonly required, see [Gladman, et al., 1999].

The number of certificates required is potentially large (with the additional influence that a user may possess many public keys used for different tasks). Therefore the task needs to be distributed in some way since one CA could not handle all possible public key certificates. Typically this is addressed via a hierarchical structure so that if a key does not belong to that CA it can be traced through the tree.

It is possible that the trust in a public key becomes reduced beyond acceptability. For example if a user has his computer broken into or a user leaves the organization he works for. CA's provide cover for such eventualities by allowing certificates to be revoked [Naor and Nissim, 2000]. This means that an attempt to look up the certificate will result in a negative reply indicating to the user not to trust the public key. However, it is not necessarily a trivial matter since most damage is likely to occur between the time of compromise and revocation. Mechanisms to better cope with this are still to emerge.

The last significant problem with CA's is actual real-world implementations. There is currently no global infrastructure supporting certification, due partly to political and partly to commercial reasons. It remains to be seen exactly what the business model for certification is and what physical provisions will be made to allow users to adequately trust providers of certification.

That is not to say business isn't attempting to forge ahead. Commercial certification has typically become known as Trusted Third Parties, since they tend

to offer a range of services related to the operation of Public Key Infrastructure (PKI). The UK's Royal Mail has recently begun offering certification on a large scale through a service known as ViaCode [ViaCode, 1999], based on technology from Canadian company Entrust [Entrust, 2000].

Cryptographic Authentication. Perhaps the most important application of public key cryptography is for authentication of either entities or data origin. We shall now look at the basic techniques for authentication, digital signatures, and hash algorithms. Finally we shall take a quick look at how these are typically applied in protocols and how the use of cryptography is more difficult than it looks.

It is intuitive to think of authentication as establishing the provenance of some claim. In cryptography this is indeed the case however the way it is applied varies according to the claim being made. The fundamental concept in cryptographic authentication is to prove the possession of a cryptographic key, typically the secret key of a public key pair. Authentication allows us to provide integrity in computer systems, rather than confidentiality. This distinction is important since we need not be able to recover the original form of a message during authentication, rather we need to be able to prove that a party authored a message.

Authentication can be used in two logically distinct ways, to prove identity and to prove origin of a message. Let us briefly explain these in turn. To prove identity we can request that a user encrypt some message that we choose using their secret key. If we decrypt this with their public key we should see our original message and, if we make the assumption that only this person knows the secret key, be confident of the identity of a party. Similarly to prove message origin (and integrity) we may add a signature to a message. A signature is usually a hash of the message encrypted with a private key, where the hash is the result of some collision resistant one-way algorithm. We can write this more formally as follows:

$$Alice \rightarrow Bob : M, A_{priv}\{h(M)\}$$

Where Alice is sending Bob a message, M, and a signature of this message. A_{priv} is the secret part of her public key pair and h(M) is the hash of M. Note here that Bob would use Alice's public key A_{pub} to verify the signature. For simplicity, and in order to avoid exact implementation details, we may abbreviate the above statement to:

$$Alice \rightarrow Bob : M, A_{sig}\{M\}$$

A special type of signature known as Blind Signature was introduced by Chaum in [Chaum, 1982]. These signatures allow one party to provide another with a message to be signed without letting the signing party see the message. Chaum suggested this would be a useful way to create digital coins by allowing a customer of a bank submit coins for a bank to sign. The bank would not be able to subsequently trace the spender of the coin. A merchant receiving the coin would be able to verify the coin be checking the signature.

The simplest explanation of Blind Signature is that a message contains a blinding component before signing which is removed after signature. Although there are have been many proposals for Blind Signature we shall briefly describe the basic operation of Chaum's method, based on RSA. Assume one party, *A*, has a message, *m*, which he wants another party, *B*, to blind-sign. First *A* sends the message *m'* to *B* where *m'* is equal to $(m * r^b)$ where *r* is the random blinding factor chosen by *A* and *b* is the public key of *B*. The next step is for *B* to sign *m'* and return it to *A*. The random blinding factor can now be divided out using *b* revealing a signature, m^b on the message *m*. Note that *B* did not see *m*.

In a distributed system the authentication of users is required to allow access to the system. Authentication servers are an essential component of a secure system. This type of authentication is often referred to as *authorization*. Two well known authentication services are Kerberos [Bird, 1995] and KryptoKnight [Neuman and Ts'o, 1994]. Distributed services authenticate their clients via the authentication server, i.e. the client authenticates to the authentication server that

provides the client with authorization to use the service (typically by generating a session key).

Finally we should consider the difficulty of cryptography. The research literature is littered with broken cryptographic algorithms and protocols, even ones designed by highly respected researchers. Indeed it may be suggested that algorithms and protocols designed by people without an understanding of this difficulty will inevitably contain flaws. The following quote from Schneier suggests the reason for this, *"Conventional engineering is about making things work. It's the genesis of the term 'hack,' as in 'he worked all night and hacked the code together.' The code works; it doesn't matter what it looks like. Security is different; it's about making sure things don't NOT work"* [Schneier, 2000].

This notion causes problems on many levels. Firstly it is still not yet properly understood how to formally define cryptographic processes. Logics for authentication, such as BAN [Burrows, et al., 1990], exist but have been widely criticized, e.g. [Boyd and Mao, 1993]. At a higher level many real-life systems are broken due to inaccurate implementation (due either to poor specification or poor programming). It is probably true that most security breaches occur by discovery of implementation flaws. Finally, incompetent management and operation of secure systems frequently leads to failures [Anderson, 1994]. By definition then, "no system is 100% secure". With this in mind security engineering is about reducing risks as far as possible.

2.3.4 Network Security

Cryptography alone does not provide network security, although it is probably the most important aspect. In this subsection we attempt to tie in the discussion in the previous subsection and look at the securing of networks as a whole. We do not attempt to tackle mobile networks at this stage, as we will look more closely at these in the next chapter.

We start by looking at the possible threats in open networks and then briefly examine the controls available; access control, authentication, firewalls and

intrusion detection. Finally we examine the subject of Electronic Commerce security that addresses many of the network security challenges in interesting and novels ways.

The threats to networks are many and varied, but ultimately can be classified into those which breach confidentiality, integrity or availability. Threats to confidentiality might include disclosure of email, company documents or database records. Confidentiality is dealt with by a combination of encrypting communications and applying access controls to stored data. Access controls are an operating system mechanism that allow the owners (or administrators) of data to set permissions on the respective data. Research in access controls is currently focused on Role-based Access Controls, e.g. [Tari and Chan, 1998], known as RBAC. These types of controls allow a more dynamic and flexible approach by enabling different roles to be applied to users in order to control their access depending on the current business role they are in.

Integrity in networks is controlled through authentication, such as those methods discussed in the previous subsection. The protection of integrity involves preventing impersonation of users. For example a user who is able to guess another users' password will have legitimate access to their files and network services – in the eyes of any access control.

Control of confidentiality and integrity cannot always be provided by these methods. Suites of security applications are often used in the case of more open networks, such as when e-commerce services are provided (e.g. Web Servers), security is usually monitored by security suites. These suites tend to combine several technologies, namely, Firewalls, Anti-virus protection, vulnerability testing and Intrusion Detection. Firewalls are a way of implementing policies to network traffic by defining what is and isn't allowed through the network border. For example a Firewall might prevent connections to ActiveX controls. By combining Firewalls with encryption it is possible to create Virtual Private Networks (VPN) over open networks. This is particularly useful for enabling businesses to create secure links with their partners. Anti-virus software scans

code entering the network against known virus signatures. Vulnerability testing software examines the configuration of a network for known problems. Intrusion Detection software, as mentioned in 2.3.2, monitor apparently legitimate traffic for patterns of anomaly or misuse, and prevent intrusions taking place.

2.3.5 Electronic Commerce Security

A major implementation issue in networking technology is that of creating an environment for Electronic Commerce (also referred to as E-commerce). Of particular concern is that involving business to customer relationships rather than just the traditional business-to-business relationship. Aside from creating the business models to enable such developments the main problem lies in the information security.

We are interested in the techniques used because they confront several of the same problems we face. Firstly, E-commerce schemes strive to give the user a high level of privacy in order to mimic typical payment structures in the physical world (especially physical cash). Second, there is authentication and trust issues, the use of digital signatures is paramount in E-commerce. Finally, and more generally, E-commerce, by its very nature requires a high level of security. There is little point in having payment systems that are trivial to break.

In order to examine E-commerce schemes we shall divide them into two classes, those that aim to provide e-cash and those that aim to produce payment systems.

E-cash schemes all derive from the seminal early eighties work of David Chaum [Chaum, et al., 1988]. In this scheme allowing a bank to use blind signatures on digital 'coins' produces anonymous cash. These can then be traded with vendors who are able to verify the 'coin' as coming from the bank. Since the banks signature is blind neither the vendor nor the bank are able to trace the coin.

Many authors have extended this work by addressing the shortcomings or imposing new attack model requirements. Brands et al developed the concept of 'wallets with observers' [Brands, 1993] that introduce off-line coins and prevent

double spending by allowing a double-spent coin to become traceable. Other developments include divisible coins [Chan, et al., 1998] [Okamoto and Ohta, 1991], undeniable payments [Chen and Mitchell, 1997], electronic checks [Chaum, 1989], escrow cash [Fujisaki and Okamoto, 1998], fairness [Petersen and Poupard, 1997] and other privacy controls [Davida, et al., 1997] [Radu, et al., 1997]. There has been much research interest in this area and it is not our intention to cover every aspect, this is merely a considered summary.

Simon provided an interesting view on E-cash schemes in [Simon, 1996] by noting that if we make assumptions about the anonymity of the underlying network then anonymous e-cash schemes become a simpler task. Similar work has been carried out by Jakobsson [Jakobsson and Juels, 1998, Jakobsson and M'Raihi, 1998b] who looks at the use of anonymous communications in e-cash schemes. Syverson in [Syverson, et al., 1997] also considers how to achieve un-linkable serial transactions by assuming anonymity in the underlying network.

E-cash schemes do not inherently consider the underlying network but much research has been carried out to look for suitable network-based payment schemes. Again similar properties are required, high security, privacy and trust. Again this is a broad field and we do not wish to survey it completely but some notable works include SNPP [Dukach, 1992], NetCash [Medvinsky and Neuman, 1993], NetBill [Sirbu and Tygar, 1995], iKP [Bellare, et al., 1995], and 'ticks' [Pedersen, 1997].

Such schemes are interesting to us because they provide many novel methods of privacy-enhanced interaction between clients and servers; in particular we have been inspired by the use of Blind Signatures.

2.4 User Privacy

In this subsection we shall take a deeper look at privacy, introducing some notions of what privacy is, and why it is becoming such an important issue in the information society. Following this we examine efforts to achieve privacy in a

variety of settings, typically focusing on the Internet. Finally we examine the legal and political situation regarding privacy.

2.4.1 What is Privacy anyway?

Privacy is generally considered to be the 'right to be left alone' [Warren and Brandeis, 1890]. Provision for privacy has been widely accepted in the civilized world. Indeed, Article 12 of the Universal Declaration of Human Rights states:

> *"No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks"* [UN, 1948].

Despite such commonly accepted rights an individual's privacy is increasingly under threat in the emerging information society. The principal reasons for this are firstly that information is easier than ever to gather [Clarke, 1997] [Froomkin, 2000] [Privacy-International, 1997] and second that regulation is considerably behind these developments [Agre and Rotenberg, 1998, Bainbridge and Pearce, 2000, Litman, 2000].

Modern techniques for data mining and market research mean that vast quantities of data can be brought together to form electronic trails or dossiers. Potentially, every on-line action a user makes can be recorded and related together. Although there is nothing new about having data about us recorded, it is the ease with which this can be achieved that is truly frightening.

The picture is complicated further by many governments' reluctance to allow citizens the sufficient rights to privacy. Whilst it is difficult to assess accurately it would seem this reluctance may stem from government intelligence agencies inability to find a significant role in post-cold war society. These agencies appear to have considerable influence on government's communications policy. Three major recent events support this view: ECHELON [Campbell, 1999b], ENFOPOL [Campbell, 1999a] and the UK DTI E-commerce consultation

document [DTI, 1999] [UK-Crypto, 1999]. The latter has resulted in the widely criticized Regulation of Investigatory Powers Bill [Gladman, 2000].

ECHELON is an alleged secret global surveillance network operated by the US National Security Agency that is capable of spying on much of the world's electronic communications. There would appear to be little justification for such an operation. ENFOPOL is the proposals for security agencies access to communications networks. The proposals could make it mandatory for service providers to build in access points. Again there appears to be no adequate justification and even less consideration for the enormous cost of modification and operation that would be needed.

The controversy surrounding the UK Governments recent consultation document for building a safe environment for e-commerce suggested to many that their hands are tied somewhat by pressure from the intelligence agencies. Whilst the Government seemed relatively willing to back down over proposals for Key Escrow the manner in which it was handled left little room for proper consultation regarding alternatives (the implication of no alternative being found was a potential return to a policy of Key Escrow). This seems particularly surprising given the almost total opposition by industry and researchers to Key Escrow.

Privacy advocates are frequently accused of paranoia in these circumstances but we feel this is unjustified. How are consumers supposed to gain confidence in electronic communications if their elected government is conducting such widespread clandestine activity? More importantly, in the light of the possibilities for privacy abuse, how are consumers supposed to gain confidence that commercial entities are acting in an ethical manner with their data. It is widely known that a major inhibitor to increased use of the Internet is concern for security and privacy [Wang, et al., 1998].

Interestingly it has become apparent that Privacy may be a marketable commodity [Szabo, 1995] [Laudon, 1996]. A scheme known as P3P allows Web surfers to

negotiate privacy levels with a Web server. The implication being that users personal information actually carries some value. For example, it is typical at many news service Web sites to require the user to register their name and e-mail address before accessing content, suggesting that users "don't get something for nothing" - the price being your personal information.

This notion that privacy is a commodity is especially disturbing in view of the already apparent gap between information haves and information have-nots. A recent magazine article [Economist, 1999] suggested that privacy would not become a problem for those with wealth whilst the less financially fortunate will see increasing erosions to theirs. The article closes by noting a theory that suggests we should have no privacy whatsoever as this would make those in power (and thus more likely to be afforded privacy) more accountable. We accept this is an interesting idea but is perhaps a little too radical at the present time.

2.4.2 Privacy Enhancing Technologies

In this subsection we review some of the proposed methods for enabling anonymity and privacy protection within open networks. The goal of technical solutions to anonymity is to hide the identity of a sender from all other parties (though not necessarily including the recipient). Typically this is achieved by sending the message via one or more intermediaries in order to confuse any traffic analysis.

Chaum [Chaum, 1981] formally described this idea, which he called a digital-mix. A digital-mix is a network host with an input message set and an output message set. The transformation performed at the mix is to decrypt the input message set and then batch, pad, and re-order messages to produce the output message set. By considering each message and its associated header information in this process, the relationship between input and output is known only to the mix itself. By chaining a series of mixes together a user can foil traffic analysis within the network so long as at least one digital mix remains uncompromised.

Since the early work of Chaum many authors have tackled improving the digital-mix. Many researchers have focused on the formal aspects and strengthening the security model. Work such as [Franz and Jerichow, 1998, Kesdogan, et al., 1998] examine specific problems and pose interesting solutions to these, whilst others such as [Jakobsson, 1998, Jakobsson and M'Raihi, 1998a] propose more general solutions. Jakobsson's solution utilizes the idea of shared split-key cryptography to reduce opportunities for collaboration. However these fail to address the practical implementation problem.

Whilst digital-mixes have proven to be useful in store and forward applications, e.g. Babel [Gulcu and Tsudik, 1996], they have some major drawbacks for use in real-time communications. The batching and reordering of messages place potentially unacceptable delays on messages. Computation in terms of the message senders being required to perform multiple layers of encryption on the initial messages is a particularly inefficient distribution of load. This point is particularly pertinent in a mobile environment where terminals may have certain constraints on resources such as computation speed and energy consumption.

Some mix-based solutions have attempted to address these concerns. An ISDN based system known as Real-time Mixes [Jerichow, et al., 1998] has recently been developed and been shown to operate favourably. This system uses features of the ISDN system to create mix connections. Unfortunately this system is not suitable for non-ISDN modes of communications systems.

Another recent solution is the Onion Routing project [Goldschlag, et al., 1999, Reed, et al., 1996]. This is an Internet based technique that draws the computational complexity away from the user and into the network by proxying the anonymity process. Each trusted network operates a proxy onto the outside network that performs the layering of encryption into 'onion' data types, which then form an anonymous connection. The proxy is required to be trusted since it has full knowledge of the source and destination addresses. This causes a problem since we assume that users may be in an untrusted network at any time. Nevertheless, onion routing has been demonstrated at near real-time speeds.

In addition to the mix-based solutions to privacy there are some other novel approaches. Chaum proposed an anonymous broadcast technique called the Dining Cryptographers. This method allows a sender to broadcast messages to the other network users without his identity. Although a very elegant solution it suffers from being terribly inefficient and doesn't scale well, thus rendering it impractical for large-scale real-time communications. Another anonymity method is the CROWDS project [Reiter and Rubin, 1998, Reiter and Rubin, 1999] which allows a user to join a crowd of anonymity seekers. A sender forwards messages into their crowd, which in turn eventually forwards to the recipient. Each recipient of a message cannot then be sure who sent the message unless they have access to every crowd member.

Other significant research in anonymity includes anonymous auctions, e.g. [Franklin and Reiter, 1996], and e-cash systems. This latter category tends to focus on achieving anonymity within the protocols and the coinage rather than the network itself. A common theme is the use of blind signatures, which we use in our architecture.

Practical schemes for privacy also take the form of agreements such as TRUSTe [Reagle and Cranor, 1999] and P3P [Benassi, 1999]. The former scheme is a privacy policy monitoring scheme. In return for a privacy 'clean bill of health' companies are audited on their data collection practices, while the latter is a markup language for privacy preferences. A user wishing to access a web page, for example, can only do so if their P3P settings are lower or equal to those of the web server. From a practical viewpoint these schemes are important to increase the profile of privacy but technically offer little protection. As a security engineer one must assume rules and policies are going to be broken, a technical solution is much easier to feel safe with.

2.4.3 The Legal and Political Landscape

Technical solutions must operate within legal and political constraints. The legal landscape for privacy is currently confused with few well defined provisions and

even fewer positive legislation. Political pressures are polarized between civil rights activists e.g. [CRCL, 2000], and law enforcement proponents. This section looks at some of the issues we consider important to our research with respect to the legal and political situation.

Legal. Many countries, particularly those in the European Union (including the UK), are in the process of implementing strong data protection legislation [France, 1999]. Although these Acts do not necessarily prevent the collection of data there are many prohibitions on what can be done with the data. One important aspect of the Data Protection is the ability for the person to gain access to complete information businesses hold about them. Any protection for user privacy is welcome but such legal provisions need to be backed up by strong technical measures in order to ensure a progression towards ubiquitous use of consumer networks.

The other aspect of law relevant to privacy is the range of controls on cryptography [Koops, 1997b] [EPIC, 1997]. Although the situation is in constant flux the broad trend is towards allowing law enforcement access to communications. In the UK for example, legislation [Gladman, 2000] has recently passed through parliament that would allow law enforcement access to cryptographic keys of criminal suspects. This proposed legislation has been widely criticized by industry and the research community [FIPR, 2000]. A major problem is that it encourages the escrow of key material. Key escrow was initially proposed in the USA [Denning, 1994] but public pressure forced the government to drop the proposals. The cost to networks of key escrow would be prohibitive; indeed it is not clear that Key Escrow is possible to operate securely [Abelson, et al., 1997]. Research into the use of encryption by criminals appears to be inconclusive [Denning and Baugh, 1997], which would appear to add weight to the view that controls should be relaxed.

Engineering computer systems to include access for law enforcement is almost certain to introduce weaknesses, and would certainly add significant complexity to an already complex system. With this in mind we do not attempt to introduce it

into our requirements, though we shall discuss the possibilities for law enforcement where relevant.

2.5 Summary

This chapter presents background analysis for the field of research we are concerned with in this thesis. We began by examining the basics of computer networking, before extending these ideas by looking at mobile networking. The evolution of mobile networking towards a third generation of system is examined as well as the general architectures used and the main issues that are particular to mobility.

The next part of the chapter discussed computer and communications security, covering the basic concepts of confidentiality, integrity and availability, followed by some analysis of the main issues involved in securing systems. We followed this by examining the field of cryptography, concerning ourselves not only with the technical detail but also legal and political influence that are very much part of the field.

Having covered the main areas of basic security and cryptography we presented some information on network security and e-commerce security since these are pertinent to our work. Finally we provided some insight into the notion of privacy, examining both the threats to privacy and the technologies that attempt to redress the balance for the user. Some discussion is given of the political and legal situation regarding privacy.

Chapter 3: Security and Privacy Requirements in Mobile Communications

The previous chapter introduces both the technologies of networking and the security solutions that may be used within them. In this chapter the requirements for security and privacy in mobile communication networks are discussed. These requirements fall in to two broad categories **System Security** and **User Privacy.**

System Security we take as meaning the prevention of fraudulent activity within the network. This typically means such intrusions as false authentication of users in order to receive service without payment. Although this category affects both the users and service providers we chose to consider this a service provider problem since they take the responsibility for maintaining fraud prevention (since it is possible that the user themselves may make such attacks).

The second category, User Privacy, concerns the protection of a users data from unauthorised parties. The obvious case for user privacy is the protection of message contents from external parties. However, there are many other requirements for user privacy that we shall discuss.

In section 3.1 we look at System Security or equivalently the service providers requirements. We discuss the authentication of users, encryption over the air and security requirements for terminal equipment. Section 3.2 looks at the User Privacy requirements including both internal and external parties. We also discuss possible law enforcement requirements in mobile networks. In the final section, 3.3, we look at related work in the area including privacy techniques for general communications networks and those specifically aimed at mobile networks. A summary of the chapter is given in 3.4.

3.1 System Security: The Service Providers Perspective

In order to run a successful business network service providers need to prevent fraudulent activity within their administrative domain. Of primary concern is that

a user account is accessed only by that accounts' registered user and that their activity is correctly accounted for. As with all networked computer systems, the computers involved in running the system may be vulnerable to attack. We shall consider these to be beyond the scope of our work, but would suggest that these will be subject to protection via secure operating system controls.

User accounts may be secured through three main controls. These are: authentication of users actions, encryption of traffic on the network and through securing users' hardware terminals. We shall deal with each in turn.

3.1.1 Authentication of Users

In order to account for users actions (to provide billing to customers) authentication must take place. Such authentication must occur whenever any of the following actions takes place:

- A user becomes visible to the network. This may take place when a user activates their terminal or when the user recovers from some connection failure. The network process resulting from network connection is called **Registration**. In addition to authenticating a users' connection the location of the mobile terminal is recorded in order to route messages to the user. This is a two-part process since both the local connection point (local network) and the registered home domain (home network) must jointly update locations for this user.

- A user moves between separate sections of the network. The network process resulting from managing movement, called **Location Update,** is part of a wider group of techniques known as mobility management. Such a location update can be one of three types:

- **Inter-cell.** This is the smallest logical movement in a network. When a mobile terminal moves around a network the most suitable local cell will serve it. Typically when a user moves to a location that would be better

served by another cell service is handed over between the two. Note that intra-cell (i.e. movement within a cell) is of no relevance since the logical point-of-attachment remains constant.

- **Inter-switch.** A single switch provides the infrastructure to control many cells simultaneously. Each service provider may run many switches under their administrative control. Although moving between two switches still requires a move between two cells it is more significant because inter-cell movement is not detectable outside of the switch concerned. This means that a location update only occurs when an inter-switch (or inter-domain) movement takes place.

- **Inter-domain.** When a mobile terminal makes an inter-switch movement between two switches in different domains then this is called an inter-domain movement. This can either be between the users' home domain and a foreign domain or between two different foreign domains. This is usually more significant than a normal inter-switch movement because a location registration must take place with a new domain and the registration with the previous one must be cancelled. Also billing information may pass between different domains.

- A user activates some service of the network. In telephony terms this is known as call initiation but more generally in packet-switching this would involve sending messages to other end users or connecting to a specific service provided by the network. Typically users would be expected to pay for services and so there is a need to provide secure billing to the user by authenticating the service use. We term this process **Accountability.**

3.1.2 Encryption of Network Traffic

In order to protect user accounts it is necessary to encrypt (at least) some network traffic on the network from evesdroppers. Service providers must be able to conceal any billing and authentication information in order to prevent fraud. In

addition to concealing such data from evesdroppers it may be considered necessary to prevent network elements in other domains from accessing such data.

Encryption requires the management of key data. In order to protect authentication and billing data the service provider must be able to share keys with the mobile users. Strong authentication techniques (i.e. those where authentication is based on the knowledge of an encryption key) typically combine their purpose with that of key distribution. In other words the service provider and user possess keys that allow them to authenticate themselves and to then share session keys.

We note that although it might make commercial sense to offer content encryption to users as a privacy-enhancing feature, it is of no intrinsic benefit to the service provider.

3.1.3 Equipment Security

Strong authentication in distributed systems, and hence mobile networks, can be bypassed if a terminal is stolen or subverted in some way. Thus by taking control of a terminal an attacker may compromise that users' account. Therefore the service provider requires some inherent protection in the hardware terminals used. This can be achieved in one of two ways:

- **Biometrics.** This involves the authentication of the user to the terminal (and henceforth to the network) by some physical attribute such as a fingerprint scan or voice recognition [Jain, et al., 2000]. Biometric devices are becoming increasingly popular for network access although their deployment is far from straightforward. For instance, if biometric data is compromised then the technique is no longer possible for that user. Currently there does not appear to be any trend toward using biometrics in mobile environments.

- **Smart Cards.** These devices are small credit card sized objects that contain a chip with memory and processing capabilities [Guillou and Ugon, 1986]. Sophisticated implementations contain cryptographic functionality. Smart cards can be used to secure a terminal by rendering the terminal useless without a valid smart card. However, in such a simple configuration, the smart card becomes the weak point of attack rather than the terminal [Anderson and Kuhn, 1996]. This can be overcome by allowing the smart card to be accessed via a password, although these too have there own problems [Schneier and Schostack, 1999].

3.2 User Privacy: The Users Perspective

Although the prevention of fraud in mobile networks is the major aspect of security, our main concern is with the privacy of user data. In this section we define the requirements for privacy in mobile networks. First, we discuss the meaning of the relationship of privacy to anonymity and pseudonymity and describe the different types of privacy that user might then desire. Following this we describe what is required to protect user privacy from both external parties and from the network itself. Finally we discuss the possible requirements of law enforcement.

3.2.1 Anonymity, Pseudonymity and Different Levels of Privacy Protection

The terms Anonymity and Pseudonymity tend to get used in an interchangeable fashion in less technical literature. In our work we take anonymity to be untraceable communications whilst pseudonymity to be traceable communications which do not use the participants' proper identities. This leads to the question, are there different levels of privacy? In this subsection we look at two main aspects of privacy, what data is required for privacy protection and what levels of privacy protection are available.

Privacy is the protection of user data. Traditionally this has tended to refer to the contents of users messages, such as in e-mail. However, particularly with the

advent of large scale Internet use, anonymity - the protection of identity, has become a topic of great interest to the user community.

In mobile environments we have an additional problem not faced in traditional fixed networks, that of location. Ideally we would like to hide the relationship between location and identification, as well as keeping our message contents private. The notion of privacy thus becomes one of hiding ones' behaviour as a whole. This is convenient for us since it allows us to consider the aspects of system security as integral to privacy protection (meaning the protection of authentication and accountability data).

Users may view the threats from the systems in different ways, and indeed may vary depending on the situation they find themselves in (e.g. a user may require higher privacy levels when roaming in a foreign network). We shall now discuss the variety of privacy levels available to the user.

In [Samfat and Molva, 1994a] five categories of anonymity for mobile environments are described. These are given as follows; note that each successive class includes the previous:

- *C0* : No privacy.
- *C1* : Hiding user identity from eavesdroppers
- *C2* : Hiding user identity from foreign authorities
- *C3* : Hiding the relationship between the user and the home authority
- *C4* : Hiding the identities of the home authority from foreign authorities
- *C5* : Hiding user behaviour from the home authority

These categories attempt to simplify the complex description of privacy and anonymity in mobile environments. Clearly in *C0* the user is not provided with any privacy at all whilst in *C1* encryption may be used to prevent messages being read by external attackers. Temporary identities are also used in *C1* to protect the user identity. In *C2* these temporary identities are incorporated into authentication procedures to prevent the local network from determining the true identity of a

user. This level of protection is currently provided by GSM. Level *C3* protects the home network association from external attacks whilst internal attacks are protected at *C4*. Finally the highest level, *C5*, protects the privacy of the user from all parties **including** the home network. It is this last level we are attempting to achieve.

3.2.2 Protection from External Attacks

External parties are considered to be those only with access to the network media (either the wireline or over-the-air segment). In order to protect from such attacks it is sufficient to encrypt all traffic between network nodes. Whilst this approach is suitable for content data it may fail in two respects to user information;

- **Initial connection setup.** Is it possible for an eavesdropper to gain information from secure connection establishment, i.e. prior to a session key being shared?

- **Control/Header Information.** If the system continues to use identification information to route messages to the user then an external party may simply read this despite all content being obscured. Additionally it may reveal location information that is deemed sensitive.

3.2.3 Protection from the Network

The above problems remain true for attacks at the network level, however there are additional considerations to bear in mind. In order to achieve higher levels of privacy it is necessary to protect certain information from the network itself. In order to do this we need to consider what information is allowed to be available to which nodes in the network and which information requires protection. The three main tasks of the network (either local or home) are:

- **Authentication.** In order to connect to any part of a network a user must be able to mutually authenticate himself with the network. Where the user is roaming in a foreign network this means that the local network will need

three-way authentication between the user, local network and home network. For user privacy we must ensure that the location of the user is kept private from the home network and the identity is secure from the local network.

- **Service Provision and Secure Billing.** Once a user has established a connection then, typically, billing will occur for any services provided. This applies to services of foreign networks, which may be paid either locally or post-service through the home network. We must keep the service provision and billing aspects separate so that no behavioural links can be made. Note, we must also maintain location and identification privacy as above.

- **Routing.** In order to receive messages a user must allow the network to associate a location with it. This is apparently contrary to our privacy goals. However, if we unlink the location from the identity then we can allow the local network to manage the precise location of a pseudonymous user and the home network can manage the pseudonymous location for the users identity. We call this Privacy Routing, which we describe a solution for in Chapter 4.

Common to all operations provided in the network it must be possible for the user to provide unique messages. This can be achieved though the use of public key cryptography. We note however that the user must be able to keep their secret key safe at all times. An implication of this is that the network provider cannot be involved in key-creation operations. The manufacturer of the equipment usually performs this, however this clearly raises trust problems. We shall discuss this later but consider it to be generally outside the scope of this research.

3.2.4 Law Enforcement and Privacy Negotiation

It is possible that in some jurisdictions constraints on privacy may be applied. This can affect the user and the service provider in several ways. As we discussed in chapter 2, this may be key-length restrictions, key escrow/recovery, or other general limitations on privacy (most likely on anonymity).

With respect to mobile networks law enforcement may require access to data above and beyond message content, as this alone might not be useful. Specifically location and identification information would typically be of use. Of course, as in current telephony systems, details of billing would be subject to request. This poses the potential problem of linking up information that would otherwise be protected.

We note again that many countries' electronic communications laws are in a state of flux and even where clear trends are apparent it would not be wise to make exact statements about what is required. Indeed, it is not even clear if law enforcement is aware of their exact requirements beyond a draconian 'need' to see and hear everything [UK-Crypto, 1999].

This problem may be addressed by grading achievable privacy levels according to the requirements of the likely legal requirement. A user and network can then negotiate a suitable level as necessary. Whilst this is not a trivial matter it is certainly more flexible that a fixed (and thus low) level.

A user then might as part of connection negotiate the encryption algorithm and key-length to be used, the privacy level (*C0...C5*) and the presence of key escrow/recovery. A similar idea of security negotiation is used in many real world security protocols, perhaps the best known is the Secure Socket Layer (SSL) [Netscape, 1996] used widely in Web Browsing software. This protocol is used to secure web connections and requires the negotiation of encryption algorithm used.

Aside from the law enforcement requirements for such a negotiation it may be prudent engineering practice to allow a local service provider and a user to select options based on other policy factors. For example, a user may only possess a particular algorithm or a service provider might not allow a high level of anonymity for internal policy reasons.

For the purpose of this research we shall try to achieve the highest levels of privacy required and then suggest methods for retrograding privacy into the framework, rather than building privacy upwards. It is certainly not the intent of this research to provide solutions to law enforcement but this must remain a consideration.

3.3 Existing Solutions

Now that we have discussed the requirements for security and privacy in mobile networks we can examine what existing efforts have been presented in the literature. We tackle this is two main areas. Firstly we recap from the previous chapter the attempts to achieve privacy in general networks and secondly those specifically aimed at mobile networks.

3.3.1 Privacy in General Networks

In chapter 2 we discussed various privacy-enhancing technologies (PET's). The major techniques of interest are based on the digital-mix invented by Chaum [Chaum, 1981]. A digital-mix is a set of nodes in a network that accept and forward messages from users wishing to attain anonymity. They do this by stripping header information and removing a layer of encryption from incoming messages. Other more complex mechanisms are used to further confuse traffic analysis, such as batching, padding and reordering messages.

Several variants on the mix scheme have appeared, mostly in an attempt to solve the performance problems inherent in Chaum's original idea. Onion Routing, developed by Reed et al. [Reed, et al., 1996] is perhaps the state of the art in anonymity. This scheme works by proxying the some of the encryption overhead to a trusted node. Onion Routing is unsuitable for our work because of this trust requirement. Other schemes were reviewed in chapter 2, but also found to be problematic.

3.3.2 Privacy in Mobile Communications

The protection of privacy in mobile networks has been widely studied. In this subsection we consider firstly the security and privacy of the two most important mobile technologies, GSM (and briefly the successor system UMTS/IMT-2000) and Mobile IP. Following this we examine the major research efforts in the area. We show first that GSM and Mobile IP have low levels of privacy protection and second that although many research efforts surpass these levels they are still left wanting compared to our requirements. Our critique does not intend to show failures in approaches, but rather to highlight the lower requirements set by the various authors.

Global System for Mobile Communications (GSM)

Perhaps the most important mobile communications network is the GSM network, centered on a series of standards developed by ETSI in the early 1990's [Rahnema, 1993]. GSM is currently mainly focused on providing mobile telephony in Europe, although it is becoming increasingly common in North America and Asia. In addition to mobile telephony there are efforts to provide basic data services such as short-message services and low bandwidth Internet access through the Wireless Application Protocol (WAP) [WAPForum, 2000].

A secure design for GSM was a very important driving factor in the standardization process, due partly to the considerable levels of fraud prevalent in 1st generation analog systems. User privacy is also given a greater level of importance in GSM; indeed it offers the highest level of protection amongst the 2nd generation standards. Despite this the privacy levels in GSM are still fairly low, consisting of over-the-air encryption and the basic use of temporary identities whilst roaming. These temporary identities give anonymity against external attackers and the local (foreign) network. Temporary identities in GSM may be attacked easily as the location management protocols allow the network to perform a query upon the mobile terminal asking for the true identity.

More recently GSM has come under attacks from researchers in more fundamental ways. First, that the whole design may be fundamentally flawed

from an engineering point of view as the entire process was conducted largely behind closed doors, aka 'security through obscurity'. Experience has repeatedly shown that a lack of open development leads inevitably to serious flaws being discovered. It is with no great surprise then that the encryption schemes used in GSM have largely been broken [Wagner, et al., 1997] [Young, 1998]. Due to export regulations at the time of standardization implementations of GSM outside of Europe have a crippled level of encryption (40bits) that renders its protection almost useless. Despite this, levels of fraud in GSM are widely believed to be very low (the most significant problem appears to be physical theft of terminals).

3rd Generation Systems

Evolution from 2nd to 3rd generation systems is currently taking place and by 2005 it is expected that much of this road will have been travelled. The standardization process is not yet complete in either Europe (UMTS) or worldwide (IMT-2000). However, it appears that both these efforts (the former is likely to be a subset or a close match to the latter) will contain strong security for both data and voice. In Europe a project sponsored by ACTS entitled ASPeCT examined the requirements and put forward various solutions [ASPeCT, 1998].

The suggestions of ASPeCT include strong authentication and relatively strong privacy requirements. These requirements stop short of providing high levels of protection against internal entities. Billing is also given considerable attention, the suggested scheme unsurprisingly based on binding identities to service contracts [Horn and Preneel, 1998]. ASPeCT is controversial for its suggestion of building in a form of Key Escrow (or Key Recovery as it is often called when not legally required) [Rantos and Mitchell, 1999]. It remains to be seen whether this is carried forward into the standards documents.

Mobile IP

The Internet is currently the largest and most visible network on the globe. The current IP protocol suite does not accommodate mobility. However a

specification for extending it has been proposed, known as Mobile IP [Perkins, 1997]. Its model is similar to that of PCS in that the user has a permanent IP address at its home location and registers for a care-of address when away from home.

When a user roams out of their home domain and into another one they must obtain from the foreign domain a care-of IP address that allows the home network to route packets to the mobile no matter where they are. In order to do this a user has to send to the home domain, as part of a registration packet, a hash of a pre-agreed nonce. The two parties synchronize nonces during registration (they can also re-synchronize at a later stage if necessary). The hash function used is MD5 [Rivest, 1992].

Anonymity is not considered in Mobile IP, however we consider that with some fairly simple conceptual extensions anonymity can be achieved within the model. The use of a MIX type network would be essential to maintain unlinkability between home and foreign domains, in IP this is called tunneling although MIX networks are a specific type of tunneling. Minor extensions to the headers would then be required to allow the necessary registration messages to flow.

Work by Fasbender et al. [Fasbender, et al., 1996a, Fasbender, et al., 1996b] attempts to add anonymity to Mobile IP by using a concept called the Non Disclosure Method, or NDM. It adapts the MIX concept for use in a simplified manner within the Mobile IP structure. A user sends a registration request through several agents before reaching the home agent. The public key of the current agent encrypts each hop. All hops are encrypted before being sent out by the user.

This is essentially a simplified version of a MIX that merely tunnels packets through several agents. They note that it would be quite simple to implement within Mobile IP without requiring any special nodes like in a MIX network. However it would not achieve the same security as a MIX network since traffic analysis measures like batching, reordering and padding are not used. The second of the two papers [Fasbender, et al., 1996a] examines some performance results

using NDM and finds that although it introduces delay into the system it is not as high as might have been expected and would be acceptable in many applications. We discuss this further in chapter 4 when we introduce our general solution to network anonymity in mobile environments.

MIX-based Solutions

Other research has been conducted that has primarily focused on the use of MIX [Chaum, 1981] networks employed in mobile systems. The first results were produced by Cooper in [Cooper and Birman, 1995] which details an anonymous messaging system where a user sends via a MIX but reads via a blinded shared memory concept. Whilst this does demonstrate the idea of using MIX'es to conceal location information outsiders it does not address the issue of insiders or of authenticating mobile users.

The main body of work in this area are three papers from Germany, [Hoff, et al., 1996] and [Federrath, et al., 1995, Federrath, et al., 1996]. These focus on the problem of anonymous location management in mobile networks. [Federrath, et al., 1995] is actually a broader paper on mobile security, detailing requirements and some possible ways forward. Amongst their findings are that MIX networks can be used to unlink a user location with their identity thus protecting sensitive user data.

Federrath in [Federrath, et al., 1995] categorises the security problems in mobile networks as follows; protection of confidentiality, integrity, and availability. Confidentiality is sub categorised as; content, location and address privacy. Integrity is sub categorised as content, addressee, and usage integrity whilst availability is stated as the enabling of communications between parties who wish to and are able to communicate. In order to achieve confidentiality (i.e. privacy) they note the possible use of MIX networks to unlink sender and recipient, their suggestion is based on having a protected home agent that can collaborate with the user whilst they are roaming. We do not consider this proposal to be feasible

in a PCS network since it would be difficult to show trust in such a device to the user.

Hoff et al. [Hoff, et al., 1996] describe an anonymous mobility approach using MIX nodes to register location updates with the local and home networks. Authentication is not considered in this research. By placing the MIX network between the user and the local Mobile Switching Center (MSC) the local network can register the update without having to know the actual location. Whilst this achieves anonymity to a certain degree (*C4*) it does not completely hide information from the home network. It is not clear how the implementation of a MIX network between the user and the local network would operate, we suggest that this is an unnecessary complication. Also it may be possible to use these MIX nodes to fraudulently communicate since at the input to the node the final destination of messages is only assumed to be the local network for location update.

Similar work carried out by Federrath in [Federrath, et al., 1996] places the MIX between the MSC and the home network. It is our consideration that a MIX-type network is always going to be required to achieve full anonymity so this paper offers only little towards the technicalities of implementation, which we consider to be difficult. We also consider authentication to be a greater problem than the actual location update, again this paper offers no solutions to authentication.

Mobile Authentication Schemes

Beller, Chang and Yacobi published the first detailed results in 1993 [Beller, et al., 1993]. Their work proposed an authentication protocol for intra-domain registration using a hybrid public/secret key method. Performance considerations were such that public-key methods were reduced as far as possible. Whilst power continues to be a limitation in mobile devices this is a decreasing limitation, indeed certain classes of device are considerably more powerful than fixed networking devices of that time. Although these results were novel at the time they fall short of our requirements since they offer only intra-domain authentication and propose very limited privacy. In addition no consideration of

location tracking is made, which clearly violates our requirements from section 3.2.

Later work by Carlsen [Carlsen, 1994] noted some flaws in the protocol that allow replay attacks and proposed some new protocols that extend the previous ones with additional ones for end-to-end user security (thus against the network). His work also added in a greater level of anonymity for the user against external parties during registration. Later work on these protocols was performed by Zheng [Zheng, 1996] who noted additional flaws and was able to increase the efficiency of the protocols involved. Mu and Varadharajan in [Mu and Varadharajan, 1996] noted some further flaws and presented a further protocol improvement, which again offers a level of anonymity against system user. An additional paper, [Varadharajan and Mu, 1996], addresses securing end-to-end inter-domain communications. As with their previous work, anonymity is provided against the visited network though this is weak anonymity according to our description in section 3.2.

One of the most influential papers on mobile network security was by Molva, Samfat and Tsudik [Molva, et al., 1994]. Their work was the first to offer an inter-domain capability to mobile authentication. Based on the KryptoKnight [Molva, et al., 1992] family of protocols, it does not however offer higher levels of user privacy. Some suggestions are made at making the user identity confidential from the visited network but the home network is trusted to manage such a process. It is implicit from our requirements that the home network cannot manage such a process.

Further research following on from [Molva, et al., 1994] has been published in [Asokan, 1994, Herzberg, et al., 1994, Samfat and Molva, 1994b, Samfat, et al., 1995]. The work by Samfat et al in [Samfat and Molva, 1994b, Samfat, et al., 1995] defined the five levels of anonymity C1..C5, as discussed here in section 3.2. In their work they define protocols to achieve anonymity to level C2. Of these, the work of [Herzberg, et al., 1994] achieves the strongest privacy, detailing protocols to achieve C3 anonymity. Discussion of possible ways to

achieve higher levels of anonymity is given involving utilization of a 'homeless' mobile user model, where the user is authenticated based on previous location rather than a fixed home network, and also the inclusion of mixes to create anonymity.

The use of the homeless model has the advantage that it is simpler to create anonymity, but is insufficient in that it is only suitable for sending and not receiving messages as there is no method to contact a user for incoming calls.

In parallel to the work just mentioned are a series of papers by Bharghavan [Bharghavan, 1994, Bharghavan, 1995, Bharghavan and Ramamoorthy, 1995]. He approaches the problem in a similar way although attention is only given to achieving anonymity from eavesdroppers. The first of these papers [Bharghavan, 1994] deals only with Wireless LAN's where the anonymity problem is considerably less serious. Later in [Bharghavan, 1995] inter-domain anonymity is considered and the proxy approach to authentication is used. By this we mean that the home passes enough information to the local network to allow it to manage secure roaming whilst in that domain. Finally in [Bharghavan and Ramamoorthy, 1995] accounting is introduced into the solution. By passing accounting information as part of the authentication process networks can be sure of the 'solvency' of a user and begin to charge for services used. Again, the home network is afforded trust by the user to achieve anonymity against other parties. Our requirements clearly do not allow this.

Miscellaneous Related Research

An interesting paper by Hardjono and Seberry [Hardjono and Seberry, 1996] looks at a very wide set of implications for mobile network security. Of particular interest to them is database security and the implications for mobility. Since location management involves large databases being queried at regular intervals regarding sensitive user data this proposes many challenges. For the purpose of our work we are considering only the network aspect and assume some trusted base including a secure set of database capabilities.

Aziz and Diffie in [Aziz and Diffie, 1994] examine Wireless LAN security and provide a key exchange scheme for end to end communications that protects against unauthorised access. No privacy issues are considered and this is only dealing with intra-domain security where the mobile user is known. Another scheme by Mohan [Mohan, 1996] looks at the existing US standard IS-41 and proposes some improvements, particularly towards performance. Again privacy is not a considered issue.

Next we note a paper by Fox & Gribble [Fox and Gribble, 1996] that examines the use of inter domain Kerberos in a mobile environment. The main thrust of this paper is to have a local Kerberos server act as a proxy for the user by establishing an association between the local and home servers. Several changes to the typical Kerberos protocols are outlined. However, as with standard Kerberos, the service provides no anonymity. Also the trust model of allowing local proxy servers to have access to information limits the potential for creating anonymity through this method.

An alternative approach to anonymity was provided in [Patel and Crowcroft, 1997] which discusses a method for user privacy that attempts to alter the paradigm for PCS in order to create anonymity with accountability. He considers the notion of a homeless user in that no registration is required with a home network. Services are paid for in advance, in this way a user need not identify them selves. This is analogous to someone using a phonecard at a public phone to communicate. An important note is made that such access even though not proving identity still authenticates since it provides credentials.

Provision of tickets that are unforgeable, resistant to replay, modification, and resale are required for operation. We note that although this is a workable solution in certain situations such as the 'phone box' it suffers from the major shortcoming that a user cannot receive calls except through some form of call back. It is not feasible to make such a radical alteration to the mobile / PCS model. That is not to say that in some situations this could not happen but that really it has to occur within an existing framework.

There have been other research efforts concerning the authentication and privacy issues, we have reviewed the most interesting ones. Two additional areas that are currently emerging within mobile network security are personal reachability management [Reichenbach, et al., 1997] and Intrusion Detection [Buschkes, et al., 1998, Samfat and Molva, 1997]. Reachability management attempts to allow a user to control the flow of messages reaching a user depending on their location, current role and so forth. Mobile intrusion detection is significant from regular intrusion detection since it attempts to profile mobility of users. This has potential consequences for the privacy of users and network operators as likely to favour effective fraud prevention measures over privacy protecting measures.

3.4 Summary

The requirements for security and privacy in mobile networks can be divided into two sections, Systems and User requirements. Service providers need to be able to ensure against fraudulent activity against user accounts whilst the users may desire high levels of privacy.

Security for the service provider consists of being able to register users onto the network (provide authentication at connection time), update locations of users as they move about and manage secure billing procedures for the services the users receive. Location update may occur between cells (inter-cell), between network switches (inter-switch) or even between providers (inter-domain). Each has different requirements.

In addition to authentication of user activity the service provider can operate link encryption to protect external parties accessing information and may also provide some equipment security such as smart-card access and biometrics. Smart-card access looks like becoming the dominant method for securing terminals and is proposed for use in UMTS [O'Mahony, 1998] and IMT-2000 [Pandya, et al., 1997].

Users are concerned with protecting data about themselves. This data includes contents of messages, identification, location and behaviour when receiving service. Protection of these items runs contrary to the operation of the network so we noted that the real issue is unlinking occurrences of the data. For example, a local network is entitled to record a trace of the users movement so long as the identity is not known.

Next we discussed the law enforcement requirements in mobile networks and how this might affect users. Key recovery and restrictions on encryption strength are likely in many jurisdictions. We suggest that a requirement for mobile networks is to provide some negotiation for security parameters so that laws can be abided.

A survey of existing literature concerned with protecting privacy in mobile communications was presented. This literature is wide-ranging, drawing together existing implementations such as Mobile IP as well as theoretical efforts. Existing technologies favour a light approach to privacy whilst many research proposals suggest a need for greater privacy. Despite this recognition it remains that no single proposal attempts to achieve the highest levels of privacy, the levels we have set out in this chapter.

Chapter 4: The Mobile Networks Privacy Architecture

So far we have presented background to our research in the form of both examination of networking and security followed by detailed requirements for security and privacy in mobile environments. In this chapter we introduce our solution to these requirements. This comprises an architecture that allows the secure management of user data to provide privacy. We have called this architecture the Mobile Networks Privacy Architecture (MNPA).

We first outline our architecture explaining how the logical components fit together to provide mobility management and secure communications and system security. We follow this with a discussion of the two main components, the Privacy Routing Capability (PRC) and the Privacy Token Issuing Authority (PTIA). It is then demonstrated how these can be linked together to achieve anonymous communications via new location registration and accountability/billing protocols.

Analysis of the MNPA is not provided in this chapter, which intends only to present the MNPA. Our analysis of the MNPA is presented in chapter 5.

4.1 The Architecture

This section introduces the architecture, MNPA. First we set out the assumptions made before introducing the outline of the architecture. Then we discuss how mobility management, accounting, and communications take place.

4.1.1 Assumptions

There are several principal assumptions we have made in the MNPA in order to develop a complete solution. Some of these assumptions may be considered to introduce weaknesses into the design, which we note, whilst others are simply choices between methods.

- **Encryption**. We assume the existence of secure encryption primitives in order to concentrate on the lager picture. Whilst we make no explicit claim to the existence of completely secure cryptography it would seem that relatively secure primitives would be adequate (e.g. RSA for public-key operations). We should note that a major problem in cryptography is in the correct implementation, as noted by [Anderson and Needham, 1995] [Schneier, 1998]. Anderson also notes that the choice of algorithm can affect the properties of certain protocols. We assume that suitable algorithms are chosen and implemented well. See also chapter 7 for further discussion on encryption.

- **Network Properties.** The networks to which we can apply our architecture are intended to be heterogeneous though we have assumed a minimum set of requirements.

  Firstly we do not assume any particular medium between various network hosts, particularly the air interface between the user and the network attachment point. Whilst the air interface is typical of mobile networks it is not conceptually important to our architecture (i.e. we are considering the more general case of *mobile* rather than *wireless* networking).

  Next we assume that in order to become active in the network the user must register their location with a 'home' network (analogous to the HLR in GSM). We assume that as a user moves around that the location registered will change accordingly. This can be either between cells, switching centres, or networks. The switching centre nearest to a user, which we call the 'local' network, also records the location. These two location stores can then link communications bound for that user. This applies equally to the situation where the user has moved onto another network. This assumption seems sensible in view of the fact that most wide-area mobile networking environments operate under similar conditions. Note therefore that this work is not considering ad hoc, micro-cellular, or ubiquitous computing.

No assumption is made about the implementation of intermediate internetworks travelled across other than these may contain vulnerabilities matching similar to other parts of the network.

- **User Properties.** We assume that the user carries with them a terminal capable of accessing a local network, and ultimately their home network. In addition to this we assume that the terminal is capable of performing various cryptographic primitives without undue performance and power hindrance. This has traditionally been an assumption which researchers has been unable to make and may still not be one suitable for this environment. However, we feel that it is necessary to make it at least to begin with in order to progress the theory. Indeed with the future promise of high-bandwidth mobile multimedia terminals it begins to seem quite a reasonable one.

Another assumption we make is that the user accesses the network (wherever and for whatever reason) via a user account agreed upon with their service provider. This model is how typically mobile phone users access services and how we imagine much internetworking to operate in the future. UMTS provides for a two-tier service provision environment where network connection is provided by one subscription (the network domain) and services are provided by other subscription (service domains).

Another assumption we make is that payment for at least some services is required, and that some of these payments are made post-service. Such an assumption implies the existence of a billing system. For the purpose of this research we have not considered any further how to achieve cash-like payments (commonly known as 'pay-as-you-go'). Many research proposals for electronic cash schemes have already been presented that may be adequate for use in mobile communications. See chapter 2 for some examples.

- **Network Host Security.** Our next assumption is concerned with the security of the hosts in the network. Initially we assume the network hosts to be running secure operating systems. In a real world situation this might be

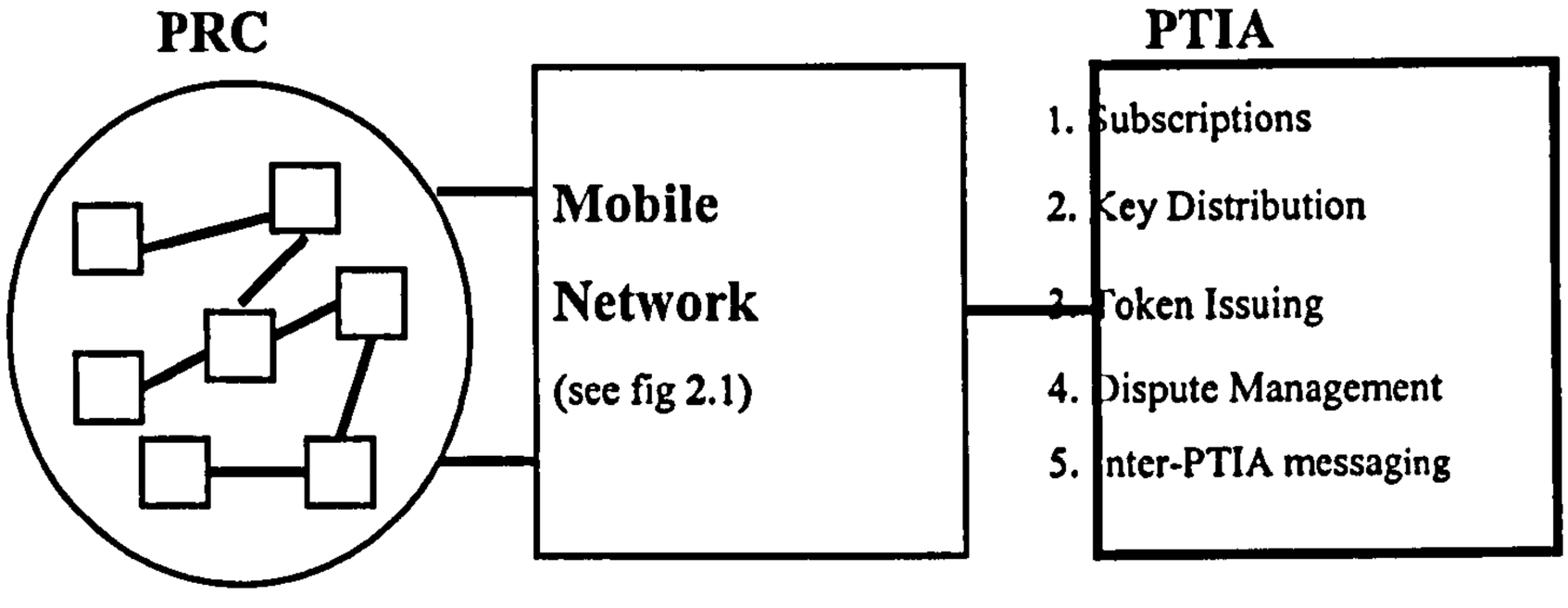
considered a foolish assumption. However this allows us to logically separate external and internal attacks. We can now class attacks on communications as external whilst attacks via the host are internal. An internal attacker might then be someone authorized to access user data but acting wrongly. We shall not consider the wider problem of intrusion detection in our work. Some work on Intrusion Detection in mobile environments already exists though this is primarily aimed at detection fraudulent call connection [Buschkes, et al., 1998, Samfat and Molva, 1997].

Similarly where we describe the operation of security functions such as key management we assume that these can be implemented in a secure manner. Again this is not an entirely practical assumption but it is one we must make in order to be clear about the presented theory. We hope that as technology progresses it will become simpler to construct software in a reliable manner; we most gladly leave this enormous problem to other researchers. In our analysis of the MNPA we shall briefly examine the effects of a variety of possible vulnerabilities.

4.1.2 Architecture Overview

Now we introduce the Mobile Networks Privacy Architecture (MNPA). As we saw from the previous chapter existing techniques are available to gain varying amounts of privacy in mobile networks. However, none of these methods meet the requirements we set out in chapter 3. The MNPA addresses these problems by implementing several new components and extending others in the mobile environment. The logical view of the MNPA is shown in fig 4.1.

The two new components are the Privacy Routing Capability (PRC) and the Privacy Token Issuing Authority (PTIA). The PRC provides a mechanism by which the sender and receiver of communications can be unlinked. The PTIA is a third party that facilitates the privacy process by acting as a broker for access tokens. This builds on the standard logical view of mobile communications as presented in chapter 2.



Key:

PRC	Privacy Routing Capability
PTIA	Privacy Token Issuing Authority

Figure 4.1: Mobile Network Privacy Architecture (MNPA).

The Privacy Routing Capability (PRC) takes the form of an efficient anonymity routing scheme similar in concept to MIX networks [Chaum, 1981]. This allows untraceable communications to occur between two parties and allows our authentication protocols to operate without the identification of parties involved.

The other major addition in the architecture is the Privacy Token Issuing Authority (PTIA). Each domain is registered with one or more of these authorities, however control is left outside of the networks involved (i.e. a different Service Domain). It is our suggestion that these may be partly controlled and/or operated by telecommunications regulators in the country of operation since the PTIA can be considered as a watchdog for accountability in anonymous communications. The PTIA is a totally novel concept that by involvement as a third party acts not only as a facilitator of user privacy but also as a watchdog for network accounting. We discuss the PTIA in 4.3.

The remaining elements of the architecture are logically unchanged from existing networks. Thus the user is connected to the terminal (possibly via a smart card). The terminal connects to the local network Mobile Switching Centre (MSC) via the air interface and base station (BS). Each MSC operates a Visitor Location Register (VLR) recording the logical locations of users currently registered within its domain. The users home domain also maintains a Home Location Register

(HLR), part of which is the current address of its users. In the case of our privacy-enhanced solution this address is an entry point to the Privacy Routing Capability (PRC). A host wishing to communicate with a mobile user can do so transparently using the public international identity of the mobile user.

In order for communications to occur several conditions must be satisfied. First, location registration of the user must occur so that the users' home network can perform routing. This must be performed with privacy for the user up to level C_5. During registration commitment to settle any billing must be achieved. Once registration is complete the user should be able to communicate privately whilst allowing the local network to settle accounts for any service used. Thought must also be given to how to achieve lower level anonymity within the MNPA. How can we, in effect, retrofit weaker privacy for situations that either do not allow or do not require such high levels of privacy? We consider this later for further work, in chapter 7.

4.1.3 Mobility Management

We have seen that in order to communicate in a mobile network environment the roaming terminal (and therefore the mobile user) must have its location registered with the local and home networks in order to create a routing association. This is the main aspect of mobility management. Our concern for mobility management is how to update locations in a secure and privacy enhancing manner.

In order to achieve registration securely the mobile and the local network must be mutually authenticated. A three party protocol is thus required. The local network and mobile user cannot always mutually authenticate themselves together since they may have no previous knowledge of each other (i.e. when roaming) so the home network is required to provide authentication of the local network to the user and of the mobile to the local network. An implication of this is that the home network must provide authentication of itself to the local network to allow the local network to trust the mobiles authentication.

In Chapter 3 we discussed the requirements for user privacy. So-called 'full' anonymity, or C5, requires that no information can be disclosed about the users' activity without the users' agreement. In practical terms there is in fact two pieces of information that can be disclosed during registration, which do not lower the privacy levels but which are required to operate correctly. Firstly the local network knows the physical location of a user, in order to perform local routing to the user. As long as the identification of the user, and its home domain identification, is not disclosed this information is of no use to the local network. Secondly the user must identify itself to the home network in order to update the previous location information. To maintain the highest level of privacy it is important that the location update does not refer to the actual location involved.

The first thing we note from this is that protocol messages between the local and home network should not identify either party. We can realize this using the Privacy Routing Capability (PRC). A message travelling from A to B has source information stripped out so that B can only observe the exit point of the PRC. During registration all messages between the local and home networks must travel through the PRC to obscure the users' home domain relationship. Since the home network must record the location of the user we insist on this location being a PRC address rather than an actual location. Since the local network records the actual location, the PRC forms the untraceable link between the actual location and that recorded by the home network.

At this point we see that although location update can be made private a method for providing authentication is required. We stated earlier that the home network could be allowed access to user identification. During registration if a user identifies himself then the home network can perform simple authentication. The result can be passed back to the local network. However this is not enough since we have already stated that mutual authentication between all three parties is required. This presents a bigger problem since no obvious mechanism is available for mutually authenticating anonymously in this circumstance.

In section 4.4 we outline a new protocols for location registration. This are based on work previously been presented in [Askwith, et al., 1998, Askwith, et al., 2000a, Askwith, et al., 2000b, Askwith, et al., 1997].

4.1.4 Accounting

Service providers in mobile networks will expect users to provide some form of accountability. This would appear to conflict with the desire to achieve privacy whilst obtaining service since to bill someone for services it would seem sensible to log all the service usage for that user for billing at a later stage. Also when a user from a foreign domain is receiving service from a network provider then they must know to whom they should charge.

We demonstrate that this can be achieved within the MNPA. Firstly, before receiving service a user can produce credentials demonstrating they are willing and able to pay (produced by the home network). The network providing service can be confident that service will be paid for since it can verify the credentials but not the identity of the user or who is going to pay on their behalf (i.e. who the users' home network is). Once service has taken place the local network can redeem the credentials in return for an appropriate form of recompense. A mechanism to provide fair play and non-repudiation is needed that should include interaction with the PTIA.

Our mechanisms for providing accountable anonymity are provided in 4.4. This work has previously been presented in [Askwith, et al., 1998], and in more detail in [Askwith, et al., 2000a].

4.1.5 Private Communications

In order to communicate in a private manner the user must not only have their message contents protected but also not have their location and identification data pair disclosed. The local network has access to the users' location but the user remains pseudonymous whilst the home network has access to the identification but has a pseudonymous location for that user. Other parties should have access

to neither the location information nor the identification information of communications that are taking place.

Our work assumes some encryption scheme to protect message content that can be agreed upon by communicating parties either before or during communications. A typical example might be Pretty Good Privacy [PGP, 1999]. Much research has been carried out in this area and it is generally well understood. As such we do not investigate this requirement in order to focus on the more difficult problem of location and identification privacy.

This is achieved by obscuring the routing information via the Privacy Routing Capability (PRC) in our architecture. We recognise that in order to provide true anonymity then messages between two end points must travel through some intermediate points in order to add some obscurity to location and identification.

Also we note that the main flaw with this type of model is that if all the intermediate stages are compromised or cooperate together then it is possible to compromise the users' privacy (of identification and location only). To reduce this possibility we suggest that the intermediate stages be as diverse and independent as possible. This brings about non-technical issues about how to set up a network to provide PRC that a user can trust, to a large extent this is outside the scope of our work but we shall discuss it for completeness.

Now that we have given an overview of the architecture we shall discuss in detail the Privacy Routing Capability (PRC) and the Privacy Token Issuing Authority (PTIA) before presenting how these fit together to allow privacy-enhanced communications.

4.2 Privacy Routing Capability (PRC)

In chapter 3 we discussed various anonymity solutions to privacy and we noted various problems with these solutions. In this section we discuss our proposed solution to this problem, which we call PRC. This section is based upon work presented in [Askwith, et al., 2000b].

4.2.1 Overview

A solution to the PRC problem must solve the following problems with existing 'mix' schemes. Firstly it must remove as much encryption overhead as possible from the user. Secondly it must not rely upon any specific underlying network technology. Finally it must not require excessively trusted components.

Our proposed PRC is, like a MIX system, composed of a series of nodes that are logically linked. Each PRC node is capable of processing messages for PRC users. In order to utilize the PRC a mobile user must have access to a range of PRC nodes. Ideally this would consist of a subset of possible nodes that is greater in number than the required number for any one route. In other words if there are M available nodes in the PRC then the user must have the option of using a subset of N nodes for a route requiring P nodes, where $M \geq N \geq P$.

The solution we have developed mainly utilizes symmetric encryption. Each message is encrypted a series of times with a symmetric key shared between the mobile user and the PRC host. The sharing of keys is achieved through the use of public key encryption. A simplified description of the operation of the PRC is as follows:

Before roaming occurs the user must

- Compute one or more symmetric keys for each accessible PRC node
- Encrypt each symmetric key with the public key of the corresponding PRC node

During roaming, when a message needs sending via the PRC the user must

- Compute a route for the message to take through the PRC
- Encrypt the message with each symmetric key in the reverse order of the route
- Include the encrypted symmetric key of the next PRC node before encrypting
- Send the message to the first PRC node in the route

Upon receipt of a message a PRC node must

- Decrypt the symmetric key at the head of the message
- Decrypt the remainder of the message using the symmetric key
- Send the message to the node identified in the head of the decrypted message

4.2.2 Detailed Operation

Now we shall describe the operation of the PRC in more detail. Pre-computation of keying material enables the user to avoid the significant burden of real-time public-key cryptography. A different symmetric key, K_{UX}, (or optionally, more than one) is computed for each node available to the user within the PRC. The key is then encrypted using the public key, K_X, of node X. This is denoted as $K_X\{K_{UX}\}$, where $K\{m\}$ is the encryption of m using key K.

The pre-computation phase may occur at any time and on any machine under the control of the user. A user may have a computer at home that could handle this task during an idle period and may be set up to allow the user to download the results before roaming. It may be possible to obtain keying material during roaming though obviously some keying material is required prior to roaming.

In order to send a message through the PRC the user must now choose a route and apply the appropriate layers of encryption using the pre-computed keys. The method employed to choose a route is beyond the scope of this research, but a simple method might be to choose at random a route from a set of pre-selected routes. We look at this problem again in chapter 7 along with other further work ideas.

Once a route is selected then the message to be sent is first encrypted with the key for the destination. Then, for each node in the PRC, the resulting message is encrypted with the symmetric key for that node and concatenated with the encrypted symmetric key for that node and the address of the next destination.

The following example shows a message, *m*, being transferred between two hosts *A* and *B* via PRC nodes *X*, *Y*, and *Z*. Figure 4.2 shows a representation of the route taken by the message. In this example it is assumed that A and B share a key, K_{AB}, where this is not the case but A can determine B's public key then the message might contain a session key encrypted with the public key, e.g. $K_B\{K_{AB}, K_{AB}\{m\}\}$.

$A \rightarrow X : K_X\{K_{AX}\}, K_{AX}\{Y, K_Y\{K_{AY}\}, K_{AY}\{Z, K_Z\{K_{AZ}\}, K_{AZ}\{B, K_{AB}\{m\}\}\}\}$

$X \rightarrow Y : K_Y\{K_{AY}\}, K_{AY}\{Z, K_Z\{K_{AZ}\}, K_{AZ}\{B, K_{AB}\{m\}\}\}$

$Y \rightarrow Z : K_Z\{K_{AZ}\}, K_{AZ}\{B, K_{AB}\{m\}\}$

$Z \rightarrow B : K_{AB}\{m\}$

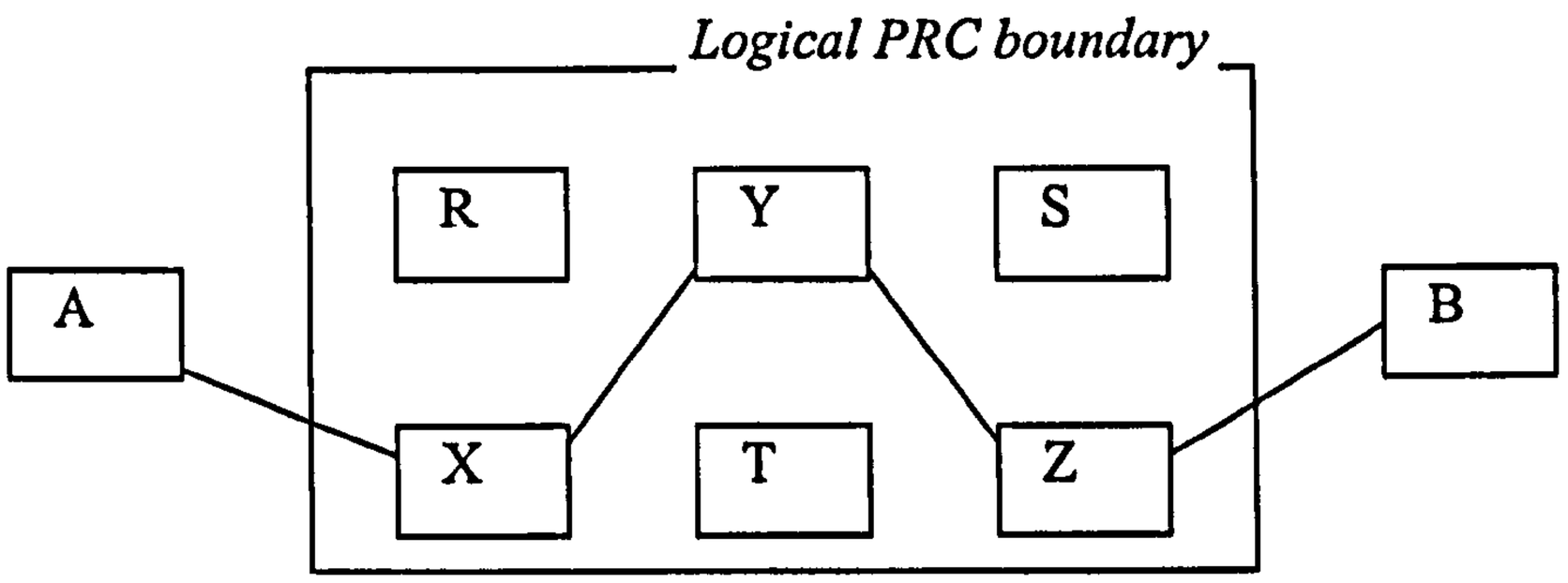


Figure 4.2. Example route of a message between two hosts using the PRC.

Replies to messages sent through the PRC can be achieved in one of two ways. The simplest method is for the mobile user to request the destination host perform a similar PRC computation independently. This assumes that the destination host has access to the location information of the mobile user and can be trusted to utilize the PRC correctly to maintain the mobile users' privacy. Where this is not the case then the mobile user must also provide a return address for the destination.

A method for achieving return addresses is to include the encrypted keys inside the destination message so the recipient can return them through the route. The simplest way to achieve this is to modify the pre-computation stage to include the previous-node identifier in the encrypted symmetric key. This prevents the PRC

having to perform extra public-key computation but does require more careful pre-computation. For each available node in the PRC the user should compute a key for each suitable following node in a route. Note that a different return route from the outgoing route is possible.

The following example shows a message, *m*, being transferred between two hosts *A* and *B* via PRC nodes *X*, *Y*, followed by a response message, *n*, from *B* sent via PRC nodes *Y* and *Z*. For simplicity assume *m* contains the return address parameters. The leading parameters (F|R) indicate either forward or return directions.

$$A \rightarrow X : F, K_X\{K_{AX}\}, K_{AX}\{Y, K_Y\{K_{AY}\}, K_{AY}\{B, K_{AB}\{m\}\}\}$$
$$X \rightarrow Y : F, K_Y\{K_{AY}\}, K_{AY}\{B, K_{AB}\{m\}\},$$
$$Y \rightarrow B : F, K_{AB}\{m\}$$

B decrypts *m* to reveal the message plus return keying material for *Y* and *X* respectively; $K_Y\{Z, K_{AY}, K_Z\{A, K_{AZ}\}\}$. The reply now returns to *A* as follows:

$$B \rightarrow Y : R, K_Y\{Z, K_{AY}, K_Z\{A, K_{AZ}\}\}, K_{AB}\{n\}$$
$$Y \rightarrow Z : R, K_Z\{A, K_{AZ}\}, K_{AY}\{B, K_{AB}\{n\}\}$$
$$Z \rightarrow A : R, K_{AZ}\{Y, K_{AY}\{B, K_{AB}\{n\}\}$$

Upon receipt of the latter message *A* is able to decrypt using K_{AZ} to reveal the identity of the next key, K_{AY}, and finally the key of the recipient, K_{AB}. Once again the mobile user does not have to perform any public-key encryption and the PRC nodes only perform one operation each, to recover the symmetric key and forwarding address. The major difference in the return protocol is that the entire route is pre-computed as a single message rather than as a set of parameters.

A concern for implementation of the PRC is the choice of number of hosts to use to balance security against efficiency. Experience with the Onion Routing

network [Reed, et al., 1998] suggests that 5 hops offer sufficient security. We currently see no reason to suggest any other number for use in the PRC. What we would suggest is that each node used be operated by different domains, such as different commercial PRC providers.

We have now described a PRC that enables untraceable routing of messages through the network. This is the first building block to providing privacy services in the MNPA. The other major component of the MNPA is the Privacy Token Issuing Authority, which we describe next.

4.3 Privacy Token Issuing Authority

4.3.1 Overview

The Privacy Token Issuing Authority, PTIA, is a distributed third party application that manages the distribution of user-privacy tokens. The primary objective is to allow a simpler form of authentication by taking the computational burden away from the user. A user-privacy token allows a user to gain access to a network in an un-linkable yet authorized manner, therefore enabling the user to meet their privacy requirements. Assurance is given to the service provider by the PTIA who can trace the user by collaborating with the interested parties if necessary. The service provider – PTIA contractual agreement, legally backs this up. We will demonstrate later in this chapter how user-privacy tokens can be used as the basis for a novel billing technique.

The home network of each user is subscribed to the PTIA and collects tokens on behalf of the user. Any PTIA subscriber can verify tokens. The distributed nature of the PTIA means that a particular token may have been created by one of a number of PTIA processes present across the entire network, each with a different PTIA key. In order to allow a network to verify any token the PTIA must also handle distribution of these keys.

The tokens are collected by service providers who subscribe to the PTIA, on behalf of their users, without a link being made between the user and token (using blind signatures). By pre-computing tokens in batches we aim to relieve the network and the user of this task in real-time. These tokens may be used for various purposes where some form of assurance is required by the network.

Once issued to a user, a token can be submitted for various purposes (such as registration). During submission the token is cancelled from the system, disabling any further use. Attempts to reuse tokens are discovered by the home network and reported to the PTIA. This gives the PTIA a 'watchdog' like position in the architecture. This turns out to be of benefit since anonymous accounts can be arbitrated by the PTIA, a natural role for this third party.

To summarize, the PTIA is responsible for the following functions: managing PTIA subscriptions, issuing and cancelling tokens, key distribution, inter-process communications, and dispute management. We shall now discuss the operation of each of these tasks within the PTIA.

4.3.2 Subscriptions

Here we examine what a subscription is, how it is created and how a network is able to connect with the PTIA to perform token management tasks. A subscription to the PTIA is much like a service account with any other distributed service, i.e. the network is provided with a service address (i.e. username) and an access key.

Account creation begins with the physical process of signing an agreement between itself and the PTIA. This agreement enables the necessary assurance for future tokens. Once this is done the PTIA can send the network or service provider an account name and shared key, both encrypted under the nearest PTIA elements public key. A public-key infrastructure should be in place covering all the constituent PTIA elements. Since the number of PTIA elements is relatively small and under moderately homogenous administrative control the normal problems of Public Key Infrastructure (PKI) are reduced considerably.

Now that an account is created it is a simple task for the network to log in to obtain tokens for itself and for its users. This is achieved by making requests using the shared key provided at setup. This might be achieved as follows, where K_{NP}, is the shared subscription key:

$$Network \rightarrow PTIA : K_{NP}\{Request\}$$
$$PTIA \rightarrow Network : K_{NP}\{Response\}$$

Next we must examine the tokens themselves.

4.3.3 Tokens

In this section we examine what tokens consist of, how they are created, issued and deleted, and how service providers can validate them. A token is essentially a special type of public key certificate. Normal certificates create trust in an identified public key. With PTIA user-privacy tokens we create an anonymous public key certificate for the user that is also bound to the home network public key associated with that user. Only the home network is aware of the binding between the two keys, though the PTIA sees the public key of the home network. This allows tracing at a later stage, should a dispute occur.

The process of token issuing operates as follows:

$$Home \rightarrow PTIA : K_{HP}\{K_H, \{K_H, K_M\}*R^{K_P}\}$$
$$PTIA \rightarrow Home : K_{HP}\{K_P^{-1}\{Id, Date, \{K_H, K_M\}* R^{K_P}\}\}$$

To request a token the home network, *Home*, constructs a message containing its own public key, K_H, and the section to be signed by the PTIA. This latter section contains both K_H and the public key of the user, K_M. These are multiplied by the blinding factor that is computed by raising a random number, R, to the power K_P, where K_P is the public key of the PTIA. The result is encrypted with the shared key between the home network and the PTIA, K_{HP}, to prevent external attacks,

before finally being sent to the PTIA. The home network retains the random number whilst the token is still in use as it may be required during a dispute.

Upon receipt of the token request the PTIA decrypts the message using the shared key, K_{HP}. The remainder of the message is concatenated with a unique token identifier, *Id*, and a date-stamp, then signed using K_P^{-1}. Finally the message is encrypted with K_{HP} before being sent to the home network. The home network produces the blind signature by dividing out $R^K{}_P$, leaving $K_P^{-1}\{(Id, Date)/R^K{}_P, K_H, K_M\}$, which becomes the token. The home network records the token along with the random blinding factor, *R*.

Tokens provided to networks for network use are similar but only provide a single key (for the network) rather than two keys. It is preferable that this key is different to that used for user tokens to lessen the risk of disclosure. When a token is submitted to a network it must contain an identifier for the PTIA element that created the blind signature. This enables the network element to decide which PTIA key to use to verify the token.

Once a network has received a token on behalf of a user then it may issue user-privacy tokens to the user in question. The token acts as a certificate for the network so that a user receives as many tokens as required from the network based on one PTIA token. A token issued by a home network to a user is as follows:

$$\{K_H, K_H^{-1}\{rand, Id, Date, K_P^{-1}\{\{Id, Date\}/R^K{}_P, K_H, K_M\}\}\}$$

Here the network includes its token signing key, K_H, and uses this to sign the reminder of the token. This remainder contains a random number for the token identity, *rand*, the PTIA token identifier, *Id*, the date-stamp, *Date*, and the PTIA blind signature token.

A token presented to a service provider can be validated by evaluating the signatures and checking that the message is formatted correctly in providing the

blinded segment and associated information. When the date-stamp in a token expires the PTIA removes the token number from its records and informs the network of the expiry. The network must then request a new token. The other scenario that causes a token to be cancelled is when a dispute occurs. A network should reject a token with an expired time-stamp.

4.3.4 Key Distribution

A problem in having the PTIA as a distributed application is that networks will come into contact with tokens issued by many different processes within the PTIA. Each PTIA element must be able to operate independently and to do so requires separate public-key pairs. This now brings about the problem of efficiency if networks have to look up public keys each time they want to evaluate a token.

Our solution to this problem is to distribute token validation keys to each subscriber, using the subscription keys detailed above. Integrity of this distribution is achieved by attaching a hash of the new key table to the distribution. We justify this method of distribution by claiming that the amount of keying material is relatively small for the task being undertaken. The number of PTIA elements should be considerably smaller than the number of network and service providers. For example, if each PTIA element provided subscriptions to 10^3 service providers and there were 10^6 service providers in total then there would need to be between 10^3 PTIA elements, and therefore, keys.

New keys will need to be distributed either when existing keys expire or when a key is revoked. It is not envisaged that revocation will occur frequently. A more sensible solution might be to distribute keys once a week or once a month depending on the administrative operating conditions of the respective PTIA. By organizing the PTIA into a hierarchy it is possible to avoid updating entire key tables all in one go, they can simply be done in sections at the respective expiry times. We discuss this hierarchy in the next section.

4.3.5 Inter-Process Communication

By organizing the PTIA into separate sub-domains it is possible to make the inter-process communication tasks simpler to achieve. These tasks are key distribution (described in the previous section) and dispute management (described in the next section). Indeed if sub-domains were created it would be possible for the PTIA to be operated be many independent authorities.

The following, and depicted in figure 4.3, is an example of how a hierarchical PTIA might be achieved. Suppose we have six PTIA elements numbered 1-6. If we split these into two groups of three, e.g. 1-3 and 4-6, we can organize each group as follows:

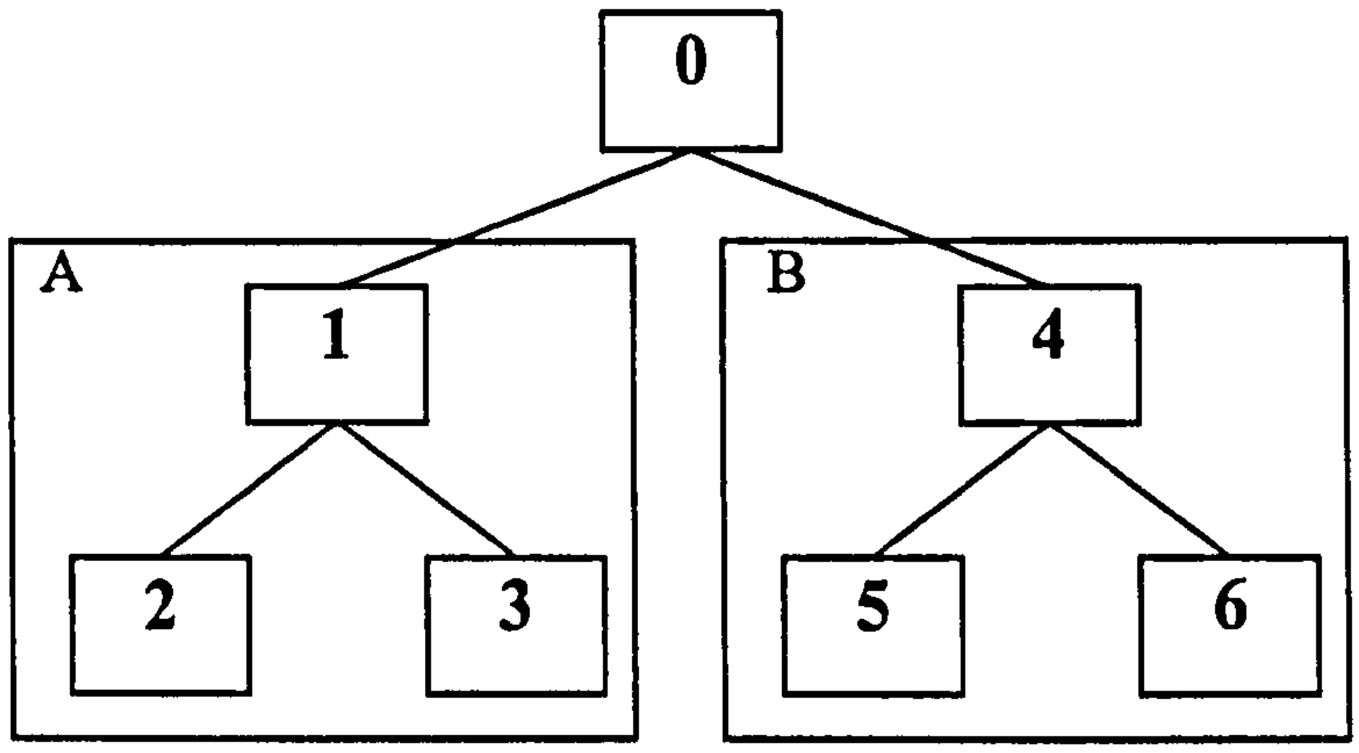


Fig 4.3: example configuration of six PTIA elements

First we place elements 1-3 in the hierarchy labeled 'A' and the element 4-6 in the hierarchy labeled 'B'. These may represent separate domains. Next the two hierarchies are joined by a root element that is responsible only for moving data between hierarchies. Now, if PTIA element 3 is required to deal with a token issued by PTIA element 2 then it knows to pass the token to element 1. Element 1 can identify the token as coming from a child node so it is passed downwards, in this case directly to element 2. Normal tree traversal techniques can be used to locate the correct node. To ease the pressure on the root node it will probably be better to arrange a series of root nodes into a backbone web-like structure.

Under the hierarchy discussed above each PTIA element needs to communicate with only a small number of neighbouring elements. To prevent attacks on communications between these elements each must protect messages using keys shared between their hierarchy partners.

4.3.6 Dispute Management

Disputes requiring the intervention of the PTIA occur when either a service provider or a user does not correctly observe a protocol involving a PTIA token. Examples of disputes might be a user submitting a date-expired token, a service provider over-charging or a user attempting to falsify details. When a dispute occurs the token must be submitted to the PTIA by both the service provider and the home network of the user so the details can be examined allowing the PTIA to determine which party is right.

If a PTIA element is called in to settle a dispute the first action to take is to make sure that the correct element is dealing with the dispute. This is achieved by the receiving PTIA element examining the token identifier and either keeping it (if the token was issued by that element) or passing it across the hierarchy as described in the previous section. Once the correct PTIA element has obtained the dispute information then it must evaluate the dispute. How this is achieved depends upon the nature of the dispute but it will typically involve working through a protocol with the given information. Details of how billing disputes are resolved are given later in this chapter when we discuss our billing protocol.

Recall from the beginning of the PTIA discussion that the user is not identified in the token. In order for some disputes to be resolved identities may be required. The PTIA is able to trace a user identity by co-operating with the home network of the user. Each token identifies the subscribing network and the network can determine the associated user from the token. This raises an interesting point in that where an identity is not required it won't be automatically revealed. The decision to reveal a user identity may be written into service contracts rather than inherent in the technology.

Where entities (providers or users) are identified in disputes it may be appropriate for the PTIA to offer a form of 'credit-rating' for other entities. This has the benefit that the enquiring entities only find out overall ratings rather than full details. We have not provided this facility in the MNPA.

The combination of the PTIA and PRC does not in itself achieve privacy for users within the MNPA. Instead they provide a framework for privacy enhancing protocols. We have developed protocols to achieve location registration, user communications and billing. These are described in the next section.

4.4 Privacy-Enhanced Communications

In this section we detail how the MNPA can be used to implement privacy-enhanced services for the mobile user. First we show how a user can anonymously update his location in both the local and home network databases - the VLR and HLR respectively. Then we show how a user can communicate with another user in an anonymous fashion. Following this, we look at how accountability can be introduced into the architecture by presenting a method of creating a post-payment billing account between a user and a service provider.

4.4.1 Location Management

In order for a mobile user to maintain a presence on the network he is required to register his location at appropriate intervals (e.g. when a terminal is activated). This procedure is termed *location registration*. This is achieved by the mobile terminal making a request for registration, which causes the local network to mutually authenticate the user before recording a location address in the local database. The home network needs to be informed of the registration so it can also record the update.

To maintain the privacy of the user within the MNPA during registration the user submits a privacy token. This token gives assurance to the local network, as shown in the 4.3. The network must now inform the user's home network of the user registration request. In addition a network token is used to provide authentication to the home network, which is then used to form a reply for the

user. Communication between the networks is conducted via a PRC connection to hide the location of each network, and thus information about the user. The aim is therefore to enable identification-location pairs to be recorded as (*Real Identification, PRC Address*) at the home address register, and (*Pseudonym, Current Address*) at the local address register.

We now present the registration protocol followed by an explanation of the contents of each message. Note that other implementation specific message components might be present but we are only concerned with those that are required for the security of the MNPA.

$$
\begin{aligned}
&\textit{Mobile} \rightarrow \textit{Local} \quad : \textit{reg, mt, } K_M^{-1}\{\textit{mt, fma, } K_{MH}\{\textit{rma}\}\} \\
&\textit{Local} \rightarrow \textit{Home} \quad : \textit{reg, lt, } K_L^{-1}\{\textit{lt, mt, rand, } K_{MH}\{\textit{rma}\}\} \\
&\textit{Home} \rightarrow \textit{Local} \quad : \textit{reg, } K_H^{-1}\{\textit{rand, } K_{MH}\{K_L\}\} \\
&\textit{Local} \rightarrow \textit{Mobile} \quad : \textit{reg, } K_H^{-1}\{\textit{rand, } K_{MH}\{K_L\}\}, K_L^{-1}\{K_M\{K_{ML}\}\}
\end{aligned}
$$

To begin with the user, *Mobile*, sends a registration request message, *reg*, to the local network, *Local*. This message contains a token, *mt*, followed by two encrypted sections. The local network first attempts to validate the token. If the token is valid then the public key, K_M, in the token is used to decrypt the first section.

The first encrypted section contains a copy of the token and the PRC forwarding address, *fma*. The local network cannot read the second encrypted message that is encrypted under the key shared between the user and the home network. It contains a PRC return address. This section is forwarded to the home network as part of the second message.

Following the first message the local network sends a message to the address given in *fma*, which is actually bound for the home network, *Home*, via the PRC. This message includes a network token for the local network, *lt*, followed by an encrypted section. The encrypted section contains *lt* and the second encrypted section from the previous message. Also included is a random number used as a

challenge for the home network. These are encrypted under the secret key of the local network, M_L^{-1}.

The token *lt* contains the public key (without identity) of the local network. This enables the home network to decrypt the remainder of the message to reveal the user information. The home network must validate *mt* and *lt* before continuing. If they are valid then the return address *rma* is entered into the user's record in the location database.

Now the home network constructs a message containing two encrypted sections. The random challenge, *rand*, and an encrypted copy of the local network public key (encrypted under the shared key between the user and home network) are encrypted under the secret key of the home network.

Upon receipt of this message the local network first checks the challenge response evaluates to *rand*, and if it does, it then sends the encrypted public key and the session key to the user. The session key is encrypted under the public key of the user and the secret key of the local network. The user can check the validity of this key by using the key encrypted by the home network, $K_{MH}\{K_L\}$.

Discussion of Protocol

The objectives of this protocol are to allow privacy enhanced location registration, mutual authentication between the three parties, and session key distribution between the user and local network. This subsection discusses how this is achieved by the protocol.

In the first message the presence of *mt* enables the local network to recover a certified public key and provides authentication of the user. The signature provides the message with integrity and authenticity. Use of the shared key (between user and home network) provides origin authentication to the receiving party.

The second message provides authentication of the local network to the home network via *lt*. The public key extracted from *lt* allows the home network to witness the integrity of the message. The home network can now provide authentication of the local network to the user by sending the retrieved public key in the third message. This message also provides mutual authentication to the local network via the encrypted challenge, *rand*.

The final message securely distributes the session key to the user. This is achieved by encrypting it with the public key of the user obtained in message one. In order to check the origin authenticity of the key it is signed by the local network. This signature is checked by the user using the key extracted by the home network from *lt* in message two.

So far this analysis demonstrates that session key distribution and mutual authentication occur. We now need to show that location registration occurs in a privacy-enhanced manner. The use of tokens combined with the PRC provides the support for this property. In the first message the local network has access to a certified public key in the user token. This provides no identifying information about the user. Similarly the PRC address, *fma*, provides no information.

Messages two and three supply the respective networks with network tokens for authentication. Contained in these are the public keys for verifying tokens. In view of the relatively small number of service providers it might be arguable that the network could identify the key. However, these keys are only used for tokens since they are blinded by the PTIA no identifying information is revealed. Moderately frequent replacement of these keys would reduce such a risk even further.

4.4.2 Remote Host Communications

We have demonstrated how a user can anonymously register his location. We now demonstrate how a user can maintain user privacy during communication with other users or service providers. Both these operations turn out to be fairly

trivial and do not require any special protocols. Much depends on the exact requirements for privacy of each party.

Several levels of privacy are possible, as discussed in section 2.2. However these levels of privacy do not address privacy between two remote hosts. There are four possible levels of privacy between two remote hosts. From highest level to lowest these are; reveal no information (total privacy), reveal location (identification privacy), reveal identification (location privacy) and reveal both location and identification (no privacy). Note that whilst a user may reveal information to a remote host we assume that they can use protocols that do not leak identification information to the network.

Following registration a mobile user obtains a pseudonym. Thus a user can achieve identification privacy using a direct connection to the remote host (thus avoiding the possible performance cost of the PRC). We envisage this to be the common case where a user is communicating with a service provider rather than another user. To most users this still represents a very high level of privacy since the location is only revealed to the granularity of the location area provided by the local network. Note that if the remote host also requires total or identification privacy, and is not currently pseudonymous, then a PRC connection must be used. Also, if the remote host is aware of the user identification and the user requires location privacy from the network then a PRC connection must be used.

4.4.3 Accountability / Billing

In certain instances a mobile user may wish to acquire services locally from the network and pay for these at a later date through their standard home network service provision account. This poses a problem for both the local network and the user. The user must be able to acquire such service without revealing information about himself yet must also be able to provide assurance of post-payment to the local network.

We achieve this by developing an account of the service provision that we cryptographically tie to a user token. The local network then submits this

resulting account to the home network in order to gain payment. Our protocol to achieve this comes in three phases. Firstly, account setup where the user presents a token that is then cryptographically bound to a service agreement to create the account. This is followed by the service provision phase where the account is passed repeatedly between the user and the network for each unit of service. Each round provides a further binding for that unit of provision. The final stage of the protocol is the settling phase where the account is submitted via the PTIA for payment. The home network can evaluate the account by reversing the cryptographic process to reveal the user token and total billing details. Payment is made if the check succeeds, otherwise a dispute occurs and the PTIA is called to arbitrate.

We now present the three phases of the accounting protocol followed by an explanation of the contents of each message after each phase is presented.

Account Setup

$$\textit{Mobile} \rightarrow \textit{Local} : M, L, K_{ML}\{M, sr, sp, K_M^{-1}\{mt\}\}$$
$$\textit{Local} \rightarrow \textit{Mobile} : M, L, K_{ML}\{L, sc, sp, K_L^{-1}\{K_M^{-1}\{mt\}\}\}$$
$$\textit{Mobile} \rightarrow \textit{Local} : M, L, K_{ML}\{M, sa, sp, K_M^{-1}\{K_L^{-1}\{K_M^{-1}\{mt\}\}\}\}$$

In the first message the mobile sends an encrypted service request to the local network. This request consists of the request, *sr*, any service parameters required, *sp*, followed by a user token, *mt*, encrypted by the private key of the user, K_M^{-1}. This message is encrypted with the key shared between the user and the local network, distributed during location registration.

The second message allows the local network to confirm the service provision, *sc*, and any parameters, *sp*, with a similar encrypted message. The private key of the local network, K_L^{-1}, which forms the basis of the account, encrypts the user token again. Note that the account is bound to both parties and neither can now tamper with it.

The final message in the setup is a confirmation by the user of the terms (a negative confirmation would simply have the *sa* flag set accordingly, e.g. 0). Once again the user applies a layer of encryption to the token. To simplify the expression of this process we shall use the term, *account*, to refer to the layers of encryption performed on the token in previous messages.

The network stores the original token in order to submit it to the home network along with the final account. An account can only be honoured if the token matches the processed account. An account is processed by recursively removing the signatures, leaving the initial token. The network cannot cheat since the signature of the user is required after each round.

Service Provision

$$Local \rightarrow Mobile : M, L, K_{ML}\{M, ss, sb, K_L^{-1}\{sb, account\}\}$$
$$Mobile \rightarrow Local : M, L, K_{ML}\{L, sd, K_M^{-1}\{sb, account\}\}$$

One round of this section of the protocol occurs after each round of service. Firstly the network provides the service, *ss*, and any billing data, *sb*. The network encrypts the account and the billing details together. The user encrypts again and includes a data acknowledgment, *sd*, which may be more complex (e.g. inclusion of service performance indicators). Failure to encrypt the account will result in service termination.

Note that the network can only receive payment for units that have an associated account. A weakness in this method is that by neglecting to sign the last service unit the user can gain this unit for free. We suggest that to lessen this risk service providers should think carefully about the size of service units and particularly those at towards the end of a session.

The protocol is **compact** (only one public key encryption required per round), **tamperproof** (neither party can add, modify or remove a round and defraud the other), **self-contained** (an account is complete after each round) and of **constant**

size (the account data passed remains the same size regardless of the number of units).

Account Settling

$$Local \rightarrow Home : L, H, K_H\{K_{LH}\}, K_{LH}\{mt, K_L, K_M, account, ad\}$$
$$Home \rightarrow Local : L, H, K_{LH}\{mt, nPayment\}$$

Once the service provision has terminated (either naturally or otherwise) the local network submits the *account* with the service details, *ad*, the initial user token, *mt*, and the verification keys, K_M and K_L. These are encrypted with the respective secret key, K_{LH}. The service details, *ad*, are the full collection service parameters, *sp*, from each round of service provision. This message is now sent to the home network, via the PRC.

The home network can now verify the submission by repeatedly applying the verification keys to *account* until *mt* is recovered (or not in the case of misuse). This verification process is compared with the claimed details, *ad*. If successful the home network will forward payment to the local network via the PTIA. We have simplified this for the purpose of this paper to *nPayment*. In reality this will be a more complex expression ensuring security between the local and home networks, of which a variety of methods are possible.

Dispute

$$Home \rightarrow PTIA : K_{HP}\{mt, K_L, K_M, account, ad\}$$
$$PTIA \rightarrow Local : K_{PL}\{mt, Dispute\}$$
or
$$PTIA \rightarrow Home : K_{HP}\{mt, Disagree\}$$

Failure to compute the original token results in the PTIA being asked to arbitrate the account. The PTIA must process the submission, identified by *mt*, in the same manner as described above and either produce a *Dispute* or a *Disagree* message.

A *Dispute* means that the local network is at fault. This can either require a reassessment by the local network or no further action. A *Disagree* means the home network is falsely claiming a dispute, which would probably result in payment being made as requested by the local network. The actual mechanisms to solve disputes are dependent on the contractual agreements between the parties and are outside the scope of this paper.

This protocol allows post-payment of services between a user and a local network. It is possible to modify it to provide post-payment between a user and any service provider by modifying the account set up phase. We assume that the new provider is subscribed to the PTIA. The changes to the protocol must allow the user and provider to share a token each and one session key. We can achieve this by modifying the first two messages of the protocol as follows (we have included the last message of the account setup phase for completeness but note that no changes have been made):

$$\textit{Mobile} \rightarrow \textit{Provider} : M, P, K_P\{M, mt, K_{MP}, sr, sp, K_M^{-1}\{mt, K_{MP}\}\}$$

$$\textit{Provider} \rightarrow \textit{Mobile} : M, P, K_{MP}\{P, pt, sc, sp, K_P^{-1}\{K_M^{-1}\{mt\}\}\}$$

$$\textit{Mobile} \rightarrow \textit{Provider} : M, P, K_{MP}\{M, sa, sp, K_M^{-1}\{K_P^{-1}\{K_M^{-1}\{mt\}\}\}\}$$

Note that the initial encryption is done using the public key of the new service provider, K_P. Inside this encryption we have now included a mobile token, *mt*, and a shared session key, K_{MP}. The second message now contains a token from the provider, *pt*. This will contain a different public key from K_P in order to allow the users' service provision to remain anonymous.

4.5 Summary

This chapter has presented our solution to the problem of achieving user privacy in mobile networking environments. This solution is called the Mobile Network Privacy Architecture (MNPA). There are 4 main sections to the chapter as follows: first we present the overall architecture. Sections two and three describe the Privacy Routing Capability (PRC) and the Privacy Token Issuing Authority (PTIA) respectively. The fourth section explains how privacy-enhanced

communications are achieved using the MNPA. We will briefly discuss the outcomes of these sections.

Section one introduces the MNPA. We started by giving a series of assumptions that have been made during the design phase. This leads us into a detailed view of the overall architecture, which contains two new logical entities, the PRC and the PTIA. The operation of both entities is presented in detail. One of the most interesting aspects of the PTIA is the capability to manage disputes. Next we discuss the new protocols that are needed to interact with these entities in order to allow privacy for the user. The protocols presented allow secure location update and anonymous post-payment service provision.

Chapter 5: Security Analysis of the MNPA

In chapter 4 we presented the Mobile Network Privacy Architecture (MNPA). The MNPA consists of two new components that extend the typical mobile computing paradigm. The two components are the Privacy Routing Capability (PRC) and the Privacy Token Issuing Authority (PTIA). The PRC allows communications between two hosts to remain untraceable and uses new protocols based on Mix-networking [Chaum, 1981]. The PTIA is a distributed certification authority that provides networks with certificates that allow them to provide anonymous authorization tokens to their subscribers.

Having provided descriptions of these two components we discussed the implementation of privacy-enhanced communications using a series of new protocols for location registration, end-to-end communications, and anonymous billing. These protocols utilize the PRC and the PTIA within the MNPA.

Here, in this chapter, we provide some analysis of the components and protocols in the MNPA. First we discuss general threats and vulnerabilities in the context of user privacy in mobile communications. This is followed by analysis of the PRC, PTIA, registration protocol and accountability protocol, each in terms of attackers, collusion, and trust required.

5.1 Threats and Vulnerabilities

Analysis of security in computer systems is a very difficult task [Anderson and Needham, 1995, Anderson, 1994, Loscocco, et al., 1998, Schneier, 1998]. It is considered impossible to achieve total security and security engineering is really the task of risk reduction. The difficulty in security engineering lies in assessing the possible **threats** and observing **vulnerabilities**. Threats are events that might occur to compromise a systems confidentiality, integrity or availability. An example of a threat is the compromise of a user password. Vulnerabilities are weaknesses in a system that might lead to a threat being realized. A vulnerability that might lead to the threat of user passwords being compromised is that

password being transmitted over a network connection without being encrypted. An **attack** is the set of actions taken by a person in order to exploit a vulnerability. An attack against user passwords might involve a network packet 'sniffer' able to detect and capture traffic used in protocols carrying passwords.

These examples sound simple and indeed password security is fairly well understood [Abadi, et al., 1997, Morris and Thompson, 1979]. However, in order to achieve total security every threat and vulnerability must be found and prevented. Many researchers have likened security to a chain, which is only as strong as its weakest link.

Experience with system security demonstrates that potential attackers are considerably better at finding these weak links than researchers. Confidence in security takes time and can only really be achieved through openness. Hiding the implementation of a technique, also known as 'security through obscurity', in the hope that attackers won't discover how something works inevitably fails. There is a long list of systems security failures attributed to 'security through obscurity', a pertinent case being the encryption used in GSM [Wagner, et al., 1997, Young, 1998].

In the context of the MNPA we are interested in specific threats to the system, others we are either not interested in or assume to be taken care of elsewhere. In the remainder of this section we try to classify the types of attacker and the attacks that may occur. Finally we examine the notion of trust and how this fits into the MNPA.

5.1.1 Attackers and Attacks

Before being able to assess the security of the MNPA we need to develop an understanding of the threats involved in the system. First we need to classify the types of attackers and the types of attacks. Attackers are the parties involved in mounting an attack, less formally known as 'crackers'. An attack is the series of operations performed by an attacker to compromise some subset of the system.

Attackers come in two basic forms; those with access to hosts on the network and those who can only access one or more links on a network. The first of these we call **internal attackers** (since they may have the same powers as authorized users, indeed they may be authorized users) and the latter we call **external attackers** (since they are not a legitimate part of the system, i.e. unauthorized users). Note that internal attackers are essentially a superset of external attackers since they can also read data on links connected to the host they control. An additional third type of attacker is one, either internal or external, who colludes with other attackers to perform an attack. This type of attacker we call a **colluding attacker.**

The types of attack can also be classified into two basic forms; those against the privacy of a user and those against the system security. The first we call **privacy attacks** and the latter we call **system attacks.** We draw this distinction for two related reasons, to demarcate attacks that affect the privacy of the mobile user (privacy is after all the primary aim of the MNPA) and secondly to distinguish between attacks on a person and attacks on the network itself.

In broad terms privacy attacks are aimed at compromising user data confidentiality whilst system attacks are aimed at compromising integrity. Privacy attacks must involve at least the revelation of a user identity, but are usually more involved. Data associated with a privacy attack within the MNPA can be classified into **identity privacy attacks, location privacy attacks, behaviour privacy attacks,** and **content privacy attacks.**

The first two of these simply refer respectively to compromise of user identity and location. A behaviour privacy attack is one that involves tracing specific activities a user performs. For example recording the services that a user connects to or the billing details associated with these connections. By creating this class of attack we wish to capture attacks on privacy that are not restricted to the other three types. Content attacks concern compromise of the payload of user messages, for example the voice part of a telephone call. Note that attacks imply an identity attack as the starting point.

System attacks are generally thought of as those that could lead to fraud. Note that a legitimate mobile user could therefore have an interest in performing system attacks. As with our discussion of attackers we define a third type of attack called **collusion attacks,** those performed by colluding attackers by sharing information with each other. Collusion attacks could lead to either privacy or system attacks.

5.1.2 Vulnerabilities

At the beginning of this chapter vulnerabilities were defined as weaknesses in a system that may lead to threats being realized. Distributed communications networks are very complex systems and assessing vulnerabilities is a difficult task. Despite this it is possible to pinpoint general areas where vulnerabilities may occur by focusing on the levels of protection required for different types of data.

Vulnerabilities occur for two distinct reasons; poor implementation and misuse, though it is possible to argue that the latter is an implementation issue since the ability to misuse should have been eliminated at requirements capture. However, the nature of open communications systems means that this is probably an impossible task. With this in mind our analysis attempts to assess the impact of vulnerabilities in implementation and operation of the various components of the MNPA.

5.1.3 Trust

A concept unique to distributed information systems is that of trust. Two different meanings can be applied to trust, both of which are important within the MNPA. Firstly there is **cryptographic trust,** which refers to the ability to trust messages according to the cryptography applied to them. For example, if Alice applies a digital signature to a message and Bob is able to verify this message (i.e. Bob is able to gain a certificate for Alice's public-key) then he can formally trust Alice authored the associated message.

A more difficult definition of trust is the broader notion of **system trust.** When a system shares information in order to perform a task it is necessary for one process to trust the other to act in a way that is both correct and honest. This distinction is important since a user may trust a component in terms of honesty but have doubts about it being able to act correctly. Unfortunately system trust is a largely intangible quality. However, cryptographic trust is a component of system trust.

In our analysis we refer mostly to system trust, rather than taking a formal approach. In terms of cryptographic trust we attempt to informally examine the protocols to determine their trustworthiness, in system terms. We also reason about the levels of system trust required in various components and the levels of trust a user might expect of these components, in terms of both honesty and competence. The issue is complicated somewhat by factoring in the possibility with which an element may be compromised (caused by low competence) leading to it acting dishonestly. The analysis first assesses the PRC and PTIA components before examining the registration and accountability protocols.

5.2 Analysis of the Privacy Routing Capability

The Privacy Routing Capability (PRC), as presented in chapter 4, acts as a logical separation between the local network and the communications partners of a mobile user. For example, during location registration a message is passed to the home network to allow mutual authentication to take place. This message travels through the PRC, thus preventing the home and local network learning the identity of each other, thereby enhancing the privacy of the user. The remainder of this section examines the potential attacks, vulnerabilities, collusion and trust in the PRC.

5.2.1 Protocol Analysis

The PRC protocol given in chapter 4 is repeated here. The leading parameters (F|R) indicate either forward or return directions.

(1) $A \rightarrow X : F, K_X\{K_{AX}\}, K_{AX}\{Y, K_Y\{K_{AY}\}, K_{AY}\{B, K_{AB}\{m\}\}\}$

(2) $X \rightarrow Y : F, K_Y\{K_{AY}\}, K_{AY}\{B, K_{AB}\{m\}\},$

(3) $Y \rightarrow B : F, K_{AB}\{m\}$

B decrypts m to reveal the message plus return keying material for Y and X respectively; $K_Y\{Z, K_{AY}, K_Z\{A, K_{AZ}\}\}$. The reply now returns to A as follows:

(4) $B \rightarrow Y : R, K_Y\{Z, K_{AY}, K_Z\{A, K_{AZ}\}\}, K_{AB}\{n\}$

(5) $Y \rightarrow Z : R, K_Z\{A, K_{AZ}\}, K_{AY}\{B, K_{AB}\{n\}\}$

(6) $Z \rightarrow A : R, K_{AZ}\{Y, K_{AY}\{B, K_{AB}\{n\}\}$

In this example a message, m, travels between two parties, A and B, via two PRC nodes, X and Y, before a return message, n, is sent between B and A, via two PRC nodes Y and Z. Without collaboration each PRC node can discover only the immediate source and destination of a message, plus the direction of a message – either forwards (F) or return (R). This can be shown to be true by examining message 2 from the example.

In this message the final destination, B, is protected from X by the symmetric key K_{AY}, shared between A and Y, that is in turn protected by K_Y, the public key for Y. A message travelling between two nodes is different at each stage due to the removal of a layer of encryption; therefore no attacker can correlate a trace based only upon the message content.

In the return direction, for example message 5, both the source and destination are included, yet only the same information is available to each node as in the forward direction. This is because the source, B, is protected from Z by K_{AY} and the destination, A, is protected from Y by K_Z. Traceability is the same as for the forward direction, though note that encryption is added rather than removed. Layering the encryption this way prevents the source from easily tracing the destination location where this is pseudonymous (though end-to-end pseudonymity is optional, and application dependent).

We have shown that the protocol is resistant to traceability in both directions against simple observation of messages by either internal or external attackers. More complex privacy attacks may be possible by sophisticated observations of the PRC network. An attacker capable of observing all links in the PRC can correlate messages based on timing and size in order to follow a message between source and destination. In the original proposal for Mix networks messages are resized into a standard length and batched to foil such attacks.

Our scheme trades this level of protection in order to achieve efficiency. The justification for this is that first without this trade off the PRC would be intolerably inefficient and second the sophistication of such an attack is considered beyond all but the most extreme attacker. It may be argued that this type of attacker could more fruitfully perform these attacks by coercing PRC elements into collusion.

Another type of attack on the PRC is denial of service, a system attack. These attacks are simple to make by any attacker capable of altering messages. For example, by replacing the keying information a PRC node is tricked into applying the wrong encryption on a message. The simplest way to defeat this is for the source node to include a hash value of the message being sent. More simple denial of service that simply attempts to block a system is rapidly becoming a critical operation in network security [Schwartau, 1999]. This is outside the scope of our research though obviously needs to be considered in future work.

5.2.2 Collusion Attacks

If all nodes in a PRC route collude then together they may mount a location privacy attack on the mobile user. Since the PRC only facilitates anonymity it is prevented from mounting an identification attack, as the mobile user is already pseudonymous when using the PRC. Where one node remains uncompromised then the scheme remains secure assuming the remaining nodes cannot correlate input and output messages (see above). This model is the same as most other mix-type schemes, and is the strongest approach we are aware of. Onion Routing

[Reed, et al., 1998] requires the use of a trusted proxy that performs routing and encryption on behalf its clients, though of course the user may control this server in some circumstances.

If the PRC is compromised entirely then it may then collude with other elements of the MNPA in order to reveal the link between the two networks (local and home) to the colluding party. In turn the PRC may learn detailed location and identification information from the local and home networks respectively. So, given a reasonable level of trust in the PRC such attacks seem highly improbable. We examine the trust requirements for the PRC next.

5.2.3 Trust

The user must trust at least one PRC element in order to trust the operation of the overall PRC. There are several recommendations to achieve this. First, the user should decide on a suitable number of elements to use. Using too many will add performance overhead but increase the likelihood of secure anonymity whilst using too few makes the system less trustworthy. It has been suggested by other researchers involved in Onion Routing that five would be adequate. Another measure that can increase trust is to use as diverse a set of elements as possible.

This diversity should be both administrative and geographical. Administrative diversity ensures that collusion becomes increasingly less likely whilst geographical diversity lowers the possibility of local effects such as legal and/or political pressure. Note that diversity in general reduces the potential of successful external attacks.

Finally, an important operational aspect of using the PRC is to vary the elements used on a regular basis. If a compromise of the PRC occurs then all subsequent messages travelling on that route can be monitored. A changing route would mean that an increased number of elements would need to be compromised, therefore reducing the possibility of further attack.

5.3 Analysis of the Privacy Token Issuing Authority

The Privacy Token Issuing Authority (PTIA), discussed in chapter 4, is a distributed authority that allows networks to obtain authorization certificates. These authorization certificates are used to issue tokens to mobile subscribers who use them to access remote networks and service providers in an anonymous yet accountable fashion. In the event of a service dispute the PTIA may be required to process token information to resolve the dispute. Except in the instance of a dispute the PTIA should not learn any information about the user involved due to the application of blind signatures on the token.

The operation of the PTIA can be split into five major tasks, subscription, token issuing, key distribution, inter-process communication and dispute management. The only parties to communicate with the PTIA are subscribers; each subscriber shares a key with the PTIA. This prevents external attackers making an attack other than against keys. Therefore we concentrate on internal attacks taking each task in turn. Following this we examine collusion and trust requirements for the PTIA.

5.3.1 Subscription

The process of subscription to the PTIA involves a network or service provider being supplied with an account name and a key by some out-of-band method (e.g. physical transfer). Only messages between two parties sharing this key are readable, and there is no significance to an external attacker of such communication. A network or service provider cannot impersonate another without this key. As the subscription key is shared the PTIA could impersonate a network or service provider. We show this to be fruitless in the remainder of the PTIA analysis. Therefore we can say that the subscription process is sufficiently secure.

5.3.2 Token Issuing

The most important task the PTIA is involved in is issuing tokens. It has been shown that communications between the PTIA and its subscribers are secure so

external attacks are not considered. The creation of a token is based on a blind signature applied to a message supplied by a PTIA subscriber. Security of this process against misuse by either party needs to be demonstrated.

Token issuing requires three basic operations: request by the subscriber, issue by the PTIA and issue by the subscriber. The first two of these are as follows:

$$Home \rightarrow PTIA \ : K_{HP}\{K_H, \{K_H, K_M\}*R^{K_P}\}$$

$$PTIA \rightarrow Home \ : K_{HP}\{K_P^{-1}\{Id, Date, \{K_H, K_M\}* R^{K_P}\}\}$$

The subscriber, *Home*, provides a message encrypted with the shared subscriber key. This message contains a public key, K_H, belonging to *Home* and a composite message, $\{K_H, K_M\}*R^{K_P}$, made up from K_H, the key of the intended mobile user, K_M, both multiplied by the blinding factor R^{K_P}, where R is a random number chosen by *Home*, raised to the power of the PTIA public key. The second phase, token issuing, involves the PTIA taking the first message and adding an identifier, *Id*, and a date-stamp, *Date*, and applying a signature to the result. This is stored by the PTIA in addition to being sent to *Home*.

The most obvious attack is for *Home* to provide a false message to be blind-signed, since the PTIA cannot read the message it is signing. Fortunately this attack is pointless as the only time it is checked is during a dispute when the PTIA would discover the fact. A second attack involves the PTIA falsely processing the message. This attack is not possible as *Home* can validate the signature has suitable parameters.

The third phase, issuing to the subscriber, involves the network modifying the result of phase two as follows; $\{K_H, K_H^{-1}\{rand, Id, Date, K_P^{-1}\{\{Id, Date\}/R^K{}_P, K_H, K_M\}\}\}$. Here the network captures the identifier, *Id*, and date-stamp, *Date*, removes the blinding factor, generates a random number, *rand*, and signs the result. The token issued to the mobile user consists of this signature and the

public key used to create it. When a mobile user requires service of some kind it submits one of these tokens to the provider.

The first attack to consider is an identification privacy attack on the token. The inclusion of the network public key suggests that the network (not the user) is identifiable. However, the network identity is not included with this key and the key is not available in any Public Key Infrastructure (PKI) system. The identity associated with this key is only available to the PTIA and the mobile user, neither of whom need to make this attack. A second attack involves the network falsely processing the token following phase two. Again this attack fails to be useful as it only becomes relevant in a dispute, when the PTIA would discover the attack.

A system attack is possible where a party replaces the public key with some other data therefore performing a denial of service. This is difficult to avoid but simple to detect as the key is used to validate the signature, something that would fail if the key were not the right one. We can therefore state that token issuing within the PTIA is suitably secure.

5.3.3 Key Distribution

The PTIA is composed of a number of clients each performing the various tasks. Each client possesses a public key that allows it to engage in issuing tokens and dispute management. These keys must be available to PTIA subscribers in order for them to validate tokens submitted to them. This is achieved by distributing a key set to each subscriber either when keys expire or are revoked by a PTIA client.

The most pressing risk in key distribution is failure to update key revocation information in a timely manner [Naor and Nissim, 2000]. The benefit of the 'push' form of key distribution we use, as opposed to the 'pull' of the certification approach, is that notification of any key revocation is more immediate. To combat the problem in the certification approach it is necessary to check a key with the authority every time it is used. Obviously there is likely to

remain some lag between compromise and revocation but this seems unavoidable, and to some extent outside the scope of the research presented here.

The distribution of keys is protected by the subscription keys shared between the PTIA and its subscribers. Assuming these keys to be secure there appears to be no feasible attacks possible from external attackers. Internal attacks may come from either the PTIA or its subscribers. However, integrity is provided by the PTIA to the distribution so no falsification of the tables may take place (which seem like a pointless attack anyway). If a PTIA client is compromised then it might distribute false information, we consider this further when discussion trust later in this section. Aside from these issues we consider the key distribution process to be suitably secure.

5.3.4 Inter-Process Communications

If a PTIA client receives data protected by another clients' public key then it must forward it to the relevant client through the hierarchy as described in chapter 4. Keys shared between the clients adjacent in the hierarchy protect these communications. The only attacks we envisage are denial of service attacks. An attacker might be able to block messages travelling through parts of the hierarchy. A solution to this might be to make the hierarchy more robust by allowing more than one route through it. As mentioned in the previous paragraph a compromised node could create a problem, again we discuss this later. Aside from this issue inter-process communication appear suitably secure.

5.3.5 Dispute Management

When either a mobile user or a network/service provider has a dispute involving a PTIA token the PTIA is required to settle the dispute. This process involves examining evidence collected by the disputing parties to determine a fair outcome. The nature of the dispute will determine the required processing performed by the PTIA. The two instances where we explicitly state disputes may occur are during location registration and during the accounting protocol. We

shall see during analysis of these protocols that dispute management is suitably secure.

5.3.6 Collusion Attacks

Tokens issued by the PTIA are blindly signed, meaning that they cannot be linked to the recipient. The implication of this is that the PTIA cannot collaborate with other parties to reveal information about specific users. During the issuing phase the PTIA records the serial number of each token, which links the token to the user. An attacker who wishes to learn the association between two networks may use this information. The main candidate for such collaboration would seem to be the local network.

This situation is somewhat complicated when the PTIA is involved in dispute management as considerable user data may be revealed. However the only collaboration threat at this stage would be one between the PTIA and a service providing network. If these two parties collude then complete service provision data can be combined with full identification data. We note later that the same risk applies in a collusion between the home network and the service providing network, a collusion that might be argued to be more likely. Alternatively given the relative authority of the PTIA this collusion might be more profitable for law enforcement access. Unfortunately if this were to happen, trust in the PTIA, and therefore the whole system would decrease. We discuss the trust of the PTIA next.

5.3.7 Trust

The PTIA requires a fairly low level of trust on the part of the user. Apart from the instance of dispute resolution the PTIA does not have access to sensitive user data. In assessing trust in the PTIA the user faces the problem of not being able to choose or control the choice of PTIA element to interact with since all interaction takes place on behalf of the home network of the user. Therefore the selection of home network must consider the associated PTIA subscription if possible.

5.4 Analysis of the Location Registration Protocol

The new location registration protocol allows a mobile user to update their location within the network. This is a two-stage process, registering a location with a new local network and updating the current location with the home network. In order to perform this task mutual authentication is required between the three parties.

5.4.1 Attacks

Discussion of the protocol was provided in Chapter 4.4.1 where justification for the security of the protocol was laid out. The protocol is repeated here for convenience.

$$Mobile \rightarrow Local : reg, mt, K_M^{-1}\{mt, fma, K_{MH}\{rma\}\}$$

$$Local \rightarrow Home : reg, lt, K_L^{-1}\{lt, mt, rand, K_{MH}\{rma\}\}$$

$$Home \rightarrow Local : reg, K_H^{-1}\{rand, K_{MH}\{K_L\}\}$$

$$Local \rightarrow Mobile : reg, K_H^{-1}\{rand, K_{MH}\{K_L\}\}, K_L^{-1}\{K_M\{K_{ML}\}\}$$

The mobile is authenticated to the local network using a mobile token, *mt*, which is forwarded to the home network to allow it to update the location register. The origin of the token is proved via the encrypted random number, $K_{MH}\{rma\}$. The local network is authenticated to the home network via the application of the public key to the second message. To do this the public key is first extracted from the privacy token, *lt*, supplied by the local network. Next, the home network is authenticated to the local network by verification of the encryption applied to the third message, the key public key for which was supplied to the local network in *mt*.

The home network supplies the mobile user with the public key of the local network, extracted from *lt*, which is encrypted with the shared key between the mobile user and the home network to ensure only the mobile user can interpret it. The final message also contains a session key, $K_L^{-1}\{K_M\{K_{ML}\}$, which can only be retrieved by an the fully authenticated user.

All public keys used in the protocol are ones used within privacy tokens, the implication of which is that they remain unidentified. Therefore an identification

privacy attack is not feasible by examining the keys alone. The local and home networks are not identified during the protocol, and messages travel through the PRC, so a location privacy attack is not feasible without collaboration. Incorrect use of a token will not result in full authentication so a system attack is not feasible using tokens alone.

External attack might include replay of tokens. However, successful completion of the protocol requires submission of a token containing a public key to which the author has the corresponding private key. An external attacker may also try to subvert the passing of a session key in the final message by inserting a different session key in its place. However, this fails due to the encryption of the key using the public key of the mobile user.

An attack that may succeed is where a local network replayed tokens, possibly in order to gain payment from the home network. However the home network is able to detect replays of tokens by keeping track of available token serial numbers for its users.

5.4.2 Collusion

The location registration protocol requires interaction between the mobile user, the local network, the home network, and the PRC. We consider effects of possible collusion between these parties, apart from the user. The home network and the PRC may collude to link the local network identity with the user identity. Similarly the PRC and local network may collude to link the current location with the identity of the home network. Neither of these cases constitutes a serious breach of privacy and so we consider the risk to be fairly low. The most serious collusion attack would be between the local and home network. These two parties hold all the sensitive information about the user and both are given an anonymous channel (during the protocol, via the PRC) with which to communicate. Therefore the user must have a high level of trust in the home network. We examine this trust requirement next.

5.4.3 Trust

Trust in the local and home networks can be reduced here to trust in the home network, as the two must collaborate to link identity and location. Fortunately the user has considerable choice in the home network he chooses to accept service from. A variety of mechanisms to satisfy an assessment of trust could be employed from media reputation, personal recommendation and formal/governmental certification. As we noted in section 5.3.7 the choice of home network should also take into consideration the choice of PTIA subscription by each network.

A strong case needs to be made to encourage (or mandate) publication of privacy protection records of service providers since many users are inevitably drawn to make choices based on more pressing requirements such as cost and service specifications rather than security issues. This priority problem cuts across many aspects of security; the human element is often the weakest point in a system.

5.5 Analysis of the Accountability Protocol

The accountability protocol presented in Chapter 4.4.3 allows a mobile user to obtain service from a network or service provider and remain anonymous, yet accountable for post-payment. There are three phases to the protocol, setup, provision and settling. The first phase involves the mobile user passing the provider a signed privacy token and any required service parameters. The token is signed by the provider and returned to the mobile user with further service parameters. The final step in the first phase is for the mobile to commit to the service providers parameters and again sign the (signed) token. The result of this signing we call an account.

This repeated signing to create an account forms the basis of the second phase, provision. For each unit of service supplied by the provider a two-message protocol is performed. In the first message the provider adds the billing details to the account and signs it. The new account is sent to the mobile user with the unit of service. The mobile user acknowledges receipt by signing the account and returning it to the provider.

Once service provision is complete the third phase of the protocol takes place. If no dispute occurs then only two messages are exchanged. First, the details of the account are passed to the home network encrypted using a shared key, which is embedded into the message using the public key of the home network (extracted previously from the token). The home network recursively applies the public keys to the account until the token is revealed. If the resulting details tally then the account may be paid. The second message contains the payment for the local network, encrypted with the shared key provided in the previous message. Note that messages between the networks pass through the PRC, preventing linking of these two.

5.5.1 Attacks

The security of the whole protocol lies in the ability of the two parties to create non-repudiation of the service provision. The provision phase requires each party to sign the details of the provision before continuing. This means that no party can add, delete or modify units of service without the other party noticing. Even if one party did not notice then the home network would spot the inconsistency during settling.

The main problem with this protocol is that the final unit of service can go unaccounted for if the mobile user fails to reply to it (deliberately or otherwise). A solution to this is to make final units of service carry no value and therefore unimportant if unpaid for. Privacy attacks are not possible unless the tokens can be attacked, which we have already determined to be suitably secure.

The dispute process involves the submission of the account details by both parties to the PTIA for examination. These submissions are made using the secure subscriber association between provider and PTIA. This leads us to assume that the PTIA will receive the same accounts that were submitted by each party. If we further assume that the PTIA is capable of processing accounts correctly then we can state that the PTIA will arrive at the correct result. Although we have chosen to ignore the details of the concluding phase of dispute resolution it is important

to now note that this phase must include signed evidence on behalf of the PTIA demonstrating the findings of the account examination. This will prevent any further tampering by either party.

An attack on the PTIA by a network attempting to compromise user privacy is possible. A network might request a dispute based on a token in order to determine from the PTIA evidence certain details about that user. We suggest that in order to reduce the likelihood of this attack the following two steps are taken;

- Maintain the anonymity of parties involved when presenting evidence
- Retain identifying details of disputes for examination by cited user

Retention of dispute lists will enable concerned users their alleged behaviour record for anomalies. For example if all user tokens are disputed then the user can safely assume they are experiencing some kind of privacy attack. Publication of the records, with no user identification, may also aid in the trust process outlined in the previous section.

5.5.2 Collusion

The discussion of collusion for accountability follows closely that presented in the previous section. During dispute however the PTIA is also present in the protocol. The PTIA may collude with the local network to link location information and identity (if provided by dispute). The home network can collude with the PTIA to discover the identity of the local network. These risks are similar to those of the PRC collaboration attacks in section 5.2.2.

5.5.3 Trust

The trust required in the various parties involved in accounting procedures is similar to those presented previously. Perhaps the major difference is that the user must trust the parties to perform dispute resolution honestly. This can be simply overcome by ensuring, possibly via auditing, that the user to view evidence in cases when disputes are presented by the PTIA.

5.6 Analysis of End-to-End Communications

Chapter 4 does not discuss at length end-to-end communications as these largely use the PRC. Therefore the security of the PRC is equivalent to the security of end-to-end communications, at least at the network layer. Where information is revealed at the application layer it is important that each party protects this using suitable application layer protocols. Examining all such protocols is obviously outside of the scope of this research.

5.6.1 Attacks

This leads us to document perhaps the only serious attack on end-to-end communications, which is a privacy attack by an end host. In this non-technical attack the end host persuades the mobile user to divulge information. We see no way around this attack other than alerting the user every time sensitive information is given out, and encouragement for education in computer security. A more serious version of this attack is where higher layer protocols (e.g. WAP [WAPForum, 2000]) automatically divulge this information without the knowledge of the user. It is important therefore to ensure that all applications contain configuration options that respect lower-layer setting (those within the MNPA). This leads to a situation whereby default options are determined by the lower layer protocols but can be over-ridden by user intervention at the application layer.

5.6.2 Collusion

The only collusion in this context is that between the PRC and the end-host. This problem is almost identical to the general trust and security of the PRC. The major difference is that whilst the user appears pseudonymous to the PRC under normal operation, the end-host may in this case have access to identification information with which to link network identification information with the PRC. This in turn could lead to collusion with that network resulting in a location privacy compromise. Depending on the power of the end-host this attack seems less likely than one directed by a network element.

5.6.3 Trust

Mobile users need only a low trust in end-hosts. This is justifiable, as in most situations the PRC will take the security burden. When information is revealed to the end-host this is presumably because the mobile user trusts the end-host enough to be supplied with it. The main dimension of end-host trust is more that of judging competence. A user should only release information to an end-host if they consider them capable of handling that information sensibly.

5.7 Summary

Chapter four presented the Mobile Network Privacy Architecture (MNPA), whilst in this chapter we were concerned with analysis of the architecture. We began by extensively defining terminology to be used in the analysis. This terminology involves classifying threats, vulnerabilities and attacks. Trust is an important metric in the MNPA, although we treat it qualitatively rather than quantitively.

Following this introductory analysis each significant section of the MNPA was analyzed according to the potential for attacks, collusion possibilities, and trust requirements. Overall the analysis was positive, showing that given reasonable trust levels in the components a user can achieve high levels of privacy. Our protocols appear to stand up well to scrutiny and the functionality of the elements, given the required assumptions (see 4.4.1), has been demonstrated to not be unreasonably vulnerable.

The major area of concern is that of collusion between elements. On one hand total privacy compromise can result from certain combinations. Unfortunately there does not appear to be any way to ensure that collusion does not occur other than through negative reputations building up from those who are found to have compromised privacy. The bright side is that increased distributed processing of data leading to a larger number of administrative domains means that collusion opportunities may be reduced.

Chapter 6: Implementation and Evaluation

The previous two chapters presented the Mobile Network Privacy Architecture (MNPA). First, chapter 4 laid out the operation of the architecture components, the Privacy Routing Capability (PRC), the Privacy Token Issuing Authority (PTIA) and the protocols that allow the implementation of privacy-enhanced communications for mobile users. Then in chapter 5 an analysis of the MNPA was conducted, examining the security of the system in terms of attacks, threats and vulnerabilities, trust and collusion.

In this chapter we present our evaluation work related to the Mobile Network Privacy Architecture (MNPA). First we discuss the aims of the prototype before detailing relevant parts of the prototype. Following this we provide a more general evaluation of the research.

6.1 Aims of the Prototype Implementation

The theoretical presentation of the MNPA answers many questions about the possibilities for privacy in mobile networks. However, it does not, nor did it attempt to, provide answers relating to how a practical implementation might fare. To examine some of the issues that might face such a practical implementation we performed some prototyping work in an experimental environment. We have limited this work to examining the effects of the protocols presented in chapter 4. The reason for this is first that these protocols require significant parts of each of the physical components to be prototyped and second that they allow us to isolate the main communications effects of the MNPA, in terms of data storage and communication, upon each element. Questions we wish to answer by developing a prototype are:

- *What can be said about transferring the theory of the MNPA into practice?*

Whilst our theory was developed with practicality in mind it is interesting to see how the MNPA will work in practice. We are particularly interested in the communications and storage needs of the architecture.

- *What can be said about the complexity of the system?*

  A central guiding principle for Computer Systems design is the KISS principle (standing for Keep It Simple Stupid) [Lampson, 1983], this principle is particularly important for security. If a system cannot be easily understood and examined it will almost certainly contain weaknesses (e.g. the Risks Archive [Risks, 2000] contains details of many software system failures). There are two main reasons for this; increased difficulty in implementing the theory correctly and increased difficulty in producing sound theory.

- *What can be said about the scaling of the architecture?*

  When developing theory it is often very difficult to foresee practical problems. A common problem is scaling, in which the size of the system contributes its own serious problems. In order to scale well a system should not degrade its performance when users or services are added.

- *What can be said about the effect of security mechanisms on performance?*

  Unfortunately security mechanisms cause a certain amount of overhead to many system operations. Where such overhead is high there it is likely that any implementation will lower the security levels to compensate.

- *What can be said about the overall security of the system?*

  Although there are no meaningful formal measures of security that can be applied to systems we hope to be able to make statements about the possible insecurities of the system based on studying the communications and data storage within the prototype.

Having posed these questions we shall now discuss the design of our implementation.

6.2 Detail of the Prototype Implementation

This section first discusses the design of the individual components of the MNPA. We start by outlining the environment under which we carried out the implementation.

6.2.1 Implementation Environment

The hardware and software used for the implementation consisted of a small number of networked PCs running the Linux operating system. These machines are isolated from the main network in the building and typically have a low usage, principally by other research students. All code was written using the C programming language. The major benefit of this setup was the simplicity offered, whilst also allowing us to easily create networked applications therefore simulating the distributed nature of the architecture.

To simulate a mobile communications network we chose to treat each machine as a separate domain. The implication of this is that each domain must be able to operate each of the components of the MNPA. Therefore each machine becomes a local network (and potentially a home network) into which a subscriber can roam.

So, each machine must contain a PTIA element, access to the PRC, location databases (VLR/HLR), as well as the ability to host processes representing mobile hosts (MH), end hosts (EH) and service providers (SP). Migration of a MH is represented either by changing the port number of the connection (intra-domain) or changing machine (inter-domain). Migration between cells in a mobile network is not represented since this has no consequence to the MNPA. We will now describe the operation of each of the MNPA components within the prototype.

6.2.2 Local Network

A subscriber always connects to the MNPA via the Local Network (LN). The LN is responsible for managing any data connections that are made whilst the subscriber is within the LN area. In terms of the MNPA the LN manages the VLR database for the given local area. The VLR handles address space for a domain. This management involves the ability to update, add and delete subscribers via location registration.

As well as managing the VLR each local network is capable of communicating with the PRC component and directly with other end hosts where necessary. In particular the local network is able to interact with the PTIA using a shared key between the two parties.

When a new subscriber enters a domain and requires location update we model this by allowing the subscriber MH to send location update requests to a pre-defined port at the local network. Completion of location update results in the exchange of a new port number representing the MH location and address.

6.2.3 Home Network

Every subscriber is registered with a home network (HN). The purpose of the HN is to act as the network service provider for each subscriber. Whilst a subscriber physically receives service from a LN the HN manages the long-term service. Two important aspects of this are maintaining a constant network identity and service billing.

The HN implements the Home Location Register (HLR). The HLR records associations between real world identities and pseudonymous locations in order to allow messages to be routed towards a subscriber. Initiation of the registration protocol by a subscriber requires the interaction of the HN of that subscriber. Therefore our HN prototype is capable of authenticating registered subscribers. As with the LN the HN is also capable of communicating with the PRC, PTIA and other end hosts.

6.2.4 PRC

The Privacy Routing Capability (PRC) is a distributed application that allows messages to travel between hosts without being traced. The implementation is made up of a series of PRC processes located on the various machines. Each PRC element processes and forwards messages according to information contained in the input stream (see chapter 4).

6.2.5 PTIA

This is the most complex component of the MNPA. The PTIA is a distributed application, elements of which, although outside the domain of a LN, may logically reside in any location. The main functions of the PTIA are subscriptions, key distribution, token issuing, dispute management, and inter-PTIA messaging.

Each element of the PTIA keeps a database of current subscriber networks (LN/HN). This database stores details of the PTIA accounts held by that element. The most important data in the database is a list of every token currently in circulation for each subscriber to the PTIA.

When networks (HN or LN) require new tokens these are created, logged in the database, and distributed to the requesting network. When a mobile subscriber uses the tokens they are submitted back to the originating for cancelling. This cancellation prevents the tokens from being reused.

If any misuse of tokens is detected (by any party) then a dispute procedure is invoked (by that party). This involves examining the tokens to discover the correct action to take. Notification is then provided to the parties involved. A complete solution to dispute management will actually involve a series of steps outside the scope of the MNPA, which we therefore leave out. An example of such a step might be the provision of written contracts stating the implications of network misuse.

A token that is submitted to a PTIA element may have originated in another PTIA element. When this situation is evident the PTIA must forward the tokens to the

relevant element. This process is transparent to the networks. We achieve this by organizing the PTIA elements into a hierarchy. Our prototype takes a simplistic approach to this naming problem by assigning one member as root elements and all others subordinate by one level. The root element performs only a routing capability, and its address is embedded into each remaining element.

The final operation the PTIA must perform is public-key distribution. Each network subscribed to the PTIA is issued with a set of keys to enable the evaluation of tokens. When new keys are generated or old ones revoked then a new list of keys must be sent out to the networks via the PTIA elements. The distributed nature of the PTIA makes this task fairly simple but synchronization is required so that the operations are performed at the correct time. For simplicity we have chosen to make this every time the PTIA is run.

6.2.6 Mobile Subscriber

We represent mobile subscribers using a single process running on any one of the machines available. As noted in section 6.2.1, migration of a subscriber is performed by either altering the connection port of the process or by invoking a process on another machine (and subsequently killing the current one.)

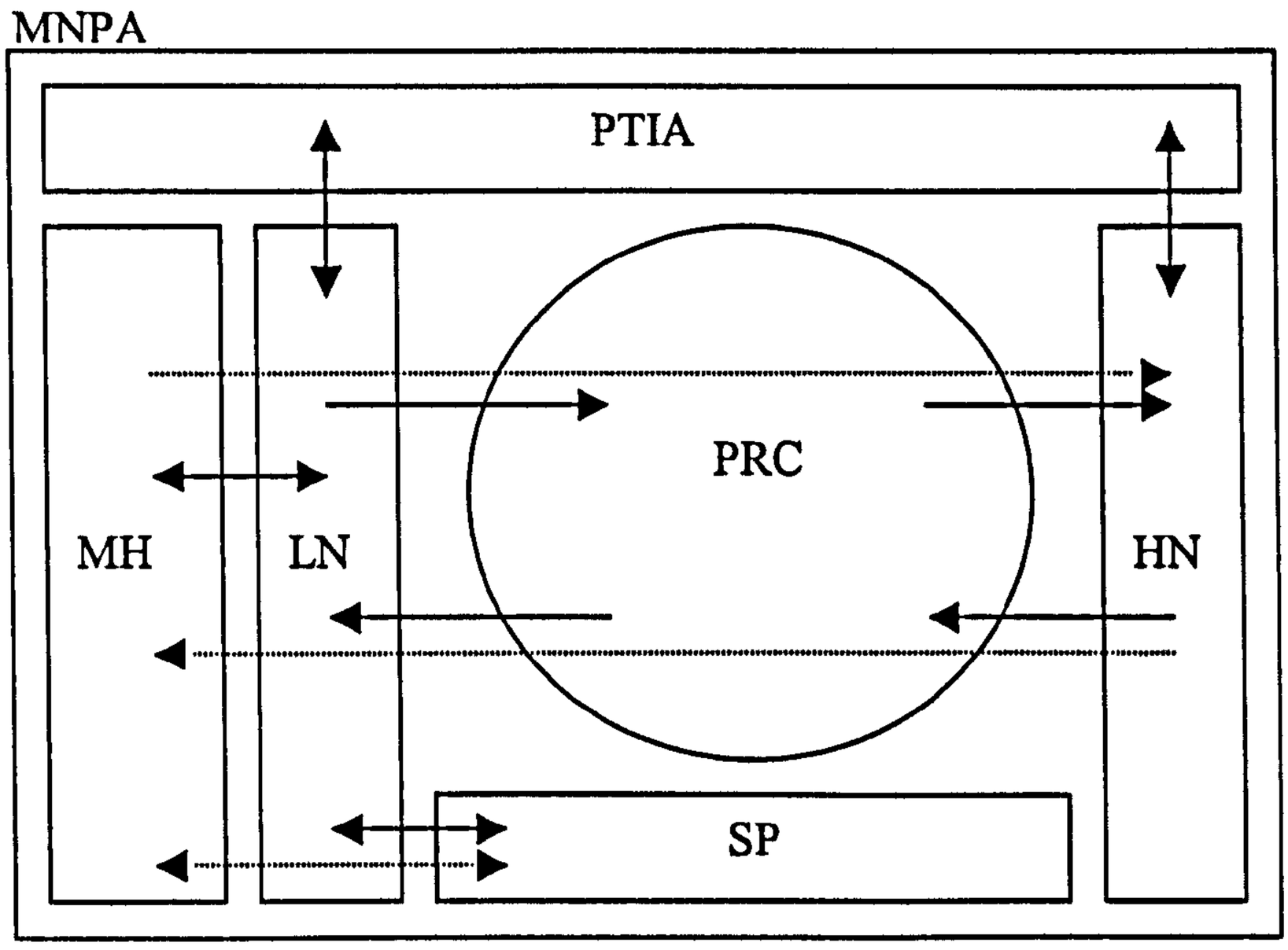
Apart from migration we require the MH to be able to connect to a LN using the MNPA registration protocol. This involves both engaging in the protocol and interaction with the PRC. Once connected then MH communications (both inbound and outbound) to end hosts and connection to services are enabled, though we have not implemented any specific code for this.

6.2.7 Service Provider

Service providers (SPs) in the MNPA prototype simulate a variety of network services by altering various parameters. The important functionality of a service provider is that it is able to interact with the service provision protocol and the PTIA. We need to be able to enable different types of service to subscribers in order to evaluate the capability of the MNPA under different conditions. We have identified three different service parameters:

- **Size** of service data units (SDUs)
- **Number** of SDUs per session
- **Frequency** of SDUs

The MNPA will impose the largest overhead on services characterized by a high number of small SDUs that arrive in a short space of time. Setting these parameters is performed by manual interaction when the SP begins execution.



PRC	Privacy Routing Capability	HN	Home Network
PTIA	Privacy Token Issuing Authority	LN	Local Network
MH	Mobile Host	SP	Service Provider

Figure 6.1 Communications within the prototype MNPA

Figure 6.1, above, shows the components of the MNPA prototype and the associated communication flows. The solid lines show actual flows of data whilst the dashed lines show logical flows of data. The dashed line between the Mobile Host (MH) and the Local and Home Networks (LN/HN) represents the location update protocol whilst the dashed line between a MH and a Service Provider (SP) via the LN represents the billing protocol.

6.3 Evaluation of the Prototype Implementation

Having outlined the operation of the prototype we now provide some evaluation of the work. In this evaluation we are concerned with two major aspects of the operation of the system, data communicated over links and data stored by various elements. To do this we examine the implementation of the protocols for registration and accounting. We note that by examining the operation of these protocols we implicitly catch the operation of all the elements within the MNPA. Finally we attempt to answer the questions posed in 6.1.

6.3.1 Registration Protocol

The registration protocol requires the interaction of several parties. In logical terms the protocol operates between the MH, LN and HN. In addition to these parties the protocol makes use of both the PRC, for messages between the LN and HN, and the PTIA, for messages associated with privacy tokens. In this subsection we examine the data structures used for storage and communication and the processing requirements of each party at each stage of the registration protocol.

Data Structures. In the registration protocol there are four main data structures, one for each message in the protocol. Each message, and therefore data structure only needs to be understood by two parties respectively. Additionally the tokens embedded in messages have their own data structure and there are several data structures representing various key data used in the protocol. Keys are either 256 bytes in length for asymmetric (public) keys or 16 bytes for symmetric (secret) keys. The protocol message data structures are as follows, note that fields are in bytes and some may be variable (the minimum required length is given).

Message 1 between Mobile Host and Local Network

```
Message1 {
        Reg             [16]
        Token           [1064]
        Encrypted       [1192] } // Message 1 length = 2272 bytes
```

Message 2 between local network and home network

```
Message2 {
    Reg          [16]
  Token          [1064]
  Encrypted      [2208] } // Message 2 length = 3288 bytes
```

Message 3 between home network and local network

```
Message3 {
    Reg          [16]
    Encrypted    [272] } // Message 3 length = 288 bytes
```

Message 4 between local network and mobile host

```
Message4 {
    Reg          [16]
    EncryptedPK  [288]
    EncryptedSK  [256] } // Message 4 length = 560 bytes
```

A token has the following structure

```
Token {
    Hkey                 [256]
    EncryptedToken {
        Rand             [16]
    Identifier   [16]
    Date         [8]
    Bsig         [768] }}
                       // Token length = 1064 bytes
```

The reason for the larger message sizes is that public keys require 256 bytes and a token requires 1064 bytes. Note that neither message 3 or 4 carry public keys or tokens so their lengths are considerably reduced. Messages 2 and 3 will in reality be slightly larger since they travel via the PRC. Transport through the PRC would add at least 64 bytes to a message, depending on the number of hops used.

Processing and Storage. Given the protocol communication requirements we now examine the storage and processing requirements. During the protocol the parties involved are required to store various data structures. Construction of messages requires cryptographic operations. Both these aspects will cause overhead to the system. We examine each of the three parties in turn.

Mobile Host. During the protocol the mobile hosts processes messages 1 and 4, a total communication overhead of 2832 bytes. In addition to these messages the mobile host need to store a token (1064 bytes), its public key pair (512 bytes), two shared secret keys (32 bytes), and PRC keying data (64+ bytes). Therefore the storage overhead of the protocol for the mobile host is at least 6008 bytes. The processing cost of encryption to the mobile host is four public key operations and two secret key operations. Preparing the message for transport via the PRC requires one extra secret key operation per hop.

Local Network. The local network is required to process all four messages during the registration protocol, a communication overhead of 6408 bytes. The storage requirement for the protocol involves a public key pair (512 bytes), a shared secret session key (16 bytes), and a token (1064 bytes). Addressing data also requires storage, in this case PRC keying data (64+ bytes) and a pseudonym address for the VLR (16 bytes). Therefore the total storage overhead for the local network in the registration protocol is at least 8080 bytes. The encryption processing overhead to the local network is seven public key operations (including three to evaluate the token). Additional secret key operations are required for preparation of messages for transport via the PRC, as mentioned above.

Home Network. Only messages 2 and 3 are processed by the home network, a total of 3576 bytes. The storage requirements for the home network involve a shared secret key (16 bytes), a public key pair (512 bytes), and the PRC keying data to be stored in the HLR database (80+ bytes). The total storage overhead is therefore 4184 bytes. The encryption overhead involves five public key operations (including three to validate a token) and two secret key operations.

Discussion. The storage requirements for the registration protocol range from 4148 to 8080 bytes. The major concern here is the scaling of the system. In order to manage large numbers of subscribers potentially in a simultaneously active

state the overhead would appear to be very large. We make the assumption that only a proportionately small number of location updates will occur simultaneously, for example hundreds, which would impose short-term storage in the region of a megabyte. The long-term storage for each party is considerably less, 1672 bytes for the local network and 608 bytes for the home network. This would appear to be a reasonably small figure for scaling. Even allowing for tens of thousands of subscribers per mobile switch the total storage would only be in the region of tens of megabytes.

Encryption processing overhead can also limit the potential of a communications network. Secret key operations are of less concern to us as designers than public key operations are. Indeed, Schneier gives an estimate of a difference factor of 1000 [Schneier, 1996]. The mobile host is required to perform four operations during the protocol, three of which are in the last phase (recovering the session key).

Given that location update is a relatively infrequent process it may be argued that the performance need not be as high as in some other functions. The other side of this argument is that location update is most important to maintain continuation of a channel on the move, and that mobile terminals are often less powerful than their fixed relatives. We feel that the number of public key operations is still small and unlikely to affect performance except in the most constrained of environments. Of course the development of more efficient and improved performance hardware will hopefully make this less important.

Public key encryption overhead for the network elements was stated as five and seven operations for the home and local networks respectively. The power constraints on the network are likely to be less important but the problem of scaling comes into play heavily. If we assume that possible hundreds of simultaneous requests are present then the number of public key operations will be 5-7 times this. Clearly this would require dedicated hardware to perform adequately. We note again at this point that advancements in hardware make this, if not immediately possible, then hopefully possible soon.

6.3.2 Billing Protocol

The billing protocol also requires the interaction of several parties. In logical terms the protocol operates between the MH, LN and HN. In addition to these parties the PTIA is required for messages associated with dispute resolution. In this subsection we examine the data structures used for storage and communication and the processing requirements of each party at each stage of the billing protocol.

Data Structures. In the billing protocol there are seven main data structures, one for each message in the protocol, with a further two for the optional dispute phase. Additionally the tokens embedded in messages have their own data structure and there are several data structures representing various key data used in the protocol. We covered this information in the previous section. The protocol message data structures are as follows, note that fields are in bytes and some may be variable (the minimum required length is given).

Messages 1-3 (account setup) between mobile host and local network

```
MessageAS {
            Src          [16]
            Dest         [16]
            Encrypted    [1128] } // Message length = 1160 bytes
```

Message 4-5 (service provision) between local network and mobile host

```
MessageSP1 {
            Src          [16]
            Dest         [16]
            Encrypted    [1146+N] }   // Message length = 1178+N
bytes

MessageSP2 {
            Src          [16]
            Dest         [16]
            Encrypted    [1130] } // Message length = 1130 bytes
```

Message 6-7 (account settling) between local network and mobile host

```
MessageST1 {
            Src          [16]
          Dest           [16]
          SessionKey     [256]
          Encrypted      [2706+(32*NU)]
          }              // Message length = 2994+(32*NU) bytes

MessageST2 {
            Src          [16]
            Dest         [16]
            Encrypted    [1320] } // Message length = 1352 bytes
```

Message 8-9 (dispute management) *between local and home network, and PTIA*

```
MessageDM1 {
            Src          [16]
            Dest         [16]
            Encrypted    [2738+(32*NU)]
            }            // Message length = 2770 + (32*NU) bytes

MessageDM2 {
            Src          [16]
            Dest         [16]
            Encrypted    [1080] } // Message length = 1112 bytes
```

Processing and Storage. In the billing protocol we can observe that there is approximately 3k bytes of data for account set up and approximately 2k bytes per unit of service. Account settling overhead is dependent on the number of service units provided but will be upwards of 4k bytes. Dispute management costs a similar amount of communications overhead as settling. The additional storage overhead for the protocol is as follows: for the mobile host the protocol imposes requirements for a token (1064 bytes), a shared session key (16 bytes), service parameters (64+ bytes), three public keys (756 bytes) and the account details (2378 bytes + 32 bytes per service unit). The local network requires the same amount of storage per client. The home network is required to allow for a session key (16 bytes), a public key (256 bytes) and the payment details (256+ bytes). In addition the dispute management phase requires the local and home networks and the PTIA to store a secret key each (16 bytes). Note also that long-term storage of billing and dispute data is likely to occur for legal reasons.

These figures are more problematic than those for location update since service provision is likely to be more frequent. Simultaneous service of tens of thousands may become difficult, however it is our intention that these protocols be distributed amongst an increased number of service providers thereby distributing the load more evenly. Also note that the storage requirements for the mobile user are low (approximately 4k bytes), this is in line with the common assumption that resources are limited in typical mobile terminals.

The encryption overhead for the local network and mobile host is one public key and one secret key operation per message. Based on the discussion in the previous subsection we claim this to be a reasonable overhead. During the settling phase the local network performs an additional single public and secret key encryptions. The home network must evaluate the account that requires it to perform approximately two public key operations per service unit. If a dispute occurs then the local network and the PTIA are required to perform significant encryption overhead, namely two public key operations per service unit of data (plus the extra public key operations performed during setup and setting).

Whilst the evaluation phase of the billing protocol by the home network (and PTIA in the case of disputes) impose this overhead we must note that this phase is primarily designed to occur offline at times suitable to the parties involved. We feel that this justifies what is otherwise a potentially expensive task.

6.3.3 Prototype Evaluation

In 6.1 we posed a series of question that we would like to ask of the prototype implementation. In this section we repeat these questions and attempt to draw answer for them.

- *What can be said about transferring the theory of the MNPA into practice?*

  The MNPA theory did not have to be altered during the parts of the implementation we performed. The protocols were sufficiently general to

allow us to make simple coding decisions. The major difficulty we observed was making decisions about the size of data elements. Other implementation decisions that may need to be made as the prototype scales would be to put limits on the capacity of the various elements. The theory of the MNPA did not, in the case of our experiments, appear to obstruct decisions about non-MNPA functionality.

- *What can be said about the complexity of the system?*

  Complexity in any system is problematic but this is especially so where security is involved. In the MNPA experimental prototype we were able to keep the system relatively simple, though a fully functional implementation would be considerably more complex. The security aspects of the system do not appear to add significant complexity. Each component has clearly defined roles to play in the architecture and the distributed (and loosely coupled) nature of the MNPA enable this relative simplicity. Of course all mobile systems are complex, this is unavoidable, but the point remains that we feel our solution is reasonable in this respect.

- *What can be said about the scaling of the architecture?*

  A dramatic increase in the number of participants of a communications system may bring about unanticipated problems. These problems are caused by the effects of bottlenecks in system resources such as bandwidth and storage space. Scaling issues specific to the MNPA are likely to be focused on the storage aspects and the encryption overhead of the protocols. We have assessed these on a per component basis and we feel that under reasonable loading and hardware assumptions the architecture will scale to accommodate potentially very large numbers of users (tens of thousands per mobile switch).

- *What can be said about the effect of security mechanisms on performance?*

This question turns out to be answered somewhat by the previous question. Poor performance in encryption operations will cause delay in the MNPA protocols. Whilst we have taken this into consideration in the design we still rely on capable hardware implementations of encryption. Public key operations are typically very slow, however the mobile station usually only has to make a small number in any one protocol. Network encryption may be problematic when the system begins to experience heavy load, particularly during the billing protocol. It is our suggestion here that dedicated encryption hardware be specified for network servers.

- *What can be said about the overall security of the system?*
  Software Engineers are guided by the principle for testing that 'a successful test is one that finds bugs'. Security analysis is very similar, there is no serious expectation that you will create 100% security, however security should be viewed as a process that is ongoing throughout the system lifetime. Our implementation experiments with the MNPA did not attempt to investigate the best practice for secure programming.

Indeed our lack of emphasis in this direction has made us realize the importance of running code in secure operating system environments combined with robust code. During tests to capture and examine communications data we were able to determine little about the content but were able to cause components to behave unexpectedly by pushing erroneous data back at them.

6.4 Overall Project Evaluation

This section is divided into two parts, first a comparison of our work with existing work in the area and then a discussion of the shortcomings of the MNPA. The aim here is to take a very broad view of the research and look at the overall achievements and the problems remaining.

6.4.1 Comparison with Related Work

In this section we outline how our architecture, the MNPA, improves on previous research results and achieves our requirements (see chapter 3). Unlike the existing solutions the MNPA combines the notion of network anonymity with privacy enhancing protocols to create a system-wide protection of privacy. The Privacy Routing Capability (PRC) provides network anonymity in the MNPA whilst the Privacy Token Issuing Authority (PTIA) and the new protocols provide the privacy enhancements.

The existing work in network anonymity suffers from being largely impractical. The efforts that address this make their compromises in ways unsuitable for use in the MNPA. Other work tends to address the security model and the theoretical issues, leaving aside the practicality. The most important works in network anonymity, related to ours, are the Onion Routing project [Reed, et al., 1998] and Real-Time Mixes [Jerichow, et al., 1998].

The PRC is an efficient solution to network anonymity since it lowers the overhead of public key encryption by making this an offline process for the user. Instead only symmetric encryption is used online (by the user). The PRC is also a relatively simple solution; this means that analysis is relatively straightforward. This is significant since it would more likely lead to a secure implementation. The PRC solves the problems of Onion Routing by removing the trusted proxy component and does not suffer from the network dependencies that Real-Time Mixes do.

In chapter 3 we reviewed various schemes for privacy in mobile communications. Our conclusions were that although many useful solutions have been presented for privacy none of these attempt to achieve all of the requirements we set out earlier that chapter, particularly those of privacy protection from internal parties. Existing commercial systems such as GSM [Rahnema, 1993] and MobileIP [Perkins, 1997] both offer low levels of security but privacy of location information and protection from internal attacks is not a priority. It is important that future mobile data networks have this as a priority.

The major research efforts such as [Molva, et al., 1994] and [Horn and Preneel, 1998] provide higher levels of privacy, the latter being one of only a few to address billing in mobile systems. However, the highest levels of privacy are again not a priority. Whilst it may be argued that the highest levels are not widely necessary we feel that a system should offer them in the instances where it is required. The exact reasons for wanting privacy are often implied to be negative to society but we prefer to leave this concern for others to tackle.

By using the PRC within the MNPA to create network anonymity we demonstrated new protocols that allow privacy-enhanced activity to take place within a mobile network. As discussed, this is by incorporating the PTIA tokens into the protocols to allow anonymous authentication to occur. The protocols we describe are relatively compact with modest computation required from the interacting parties. An additional property of these protocols when combined with the PTIA is the ability to provide non-repudiation for service provision.

The protocols and supporting components of the MNPA introduce overhead that is not present in existing proposals. This is unsurprising given that increased security typically equates to decreased performance. The importance of this degradation of performance is perhaps not great since it is expected that the highest levels of privacy will be used in niche applications (most likely ones with less strict performance constraints) and for more sensitive users. We suggest therefore it may be important to provide configurable privacy, for example by using existing protocols for less sensitive applications. This is discussed further in the next chapter.

6.4.2 Shortcomings of the MNPA

Whilst the MNPA solves an interesting problem with some novel aspects there remains several shortcomings. This section outlines these problems.

- **Trust.** In the analysis of chapter 5 we discussed the trust requirements within the MNPA. While a considerable amount of control over personal

data is shifted towards the user it remains true that some data can be compromised by elements of the network operating in an incorrect manner. This may involve external attacks on hosts that are not properly secured or indeed dishonest behaviour by system operators. We have tried to make the implications as low as possible, but it seems inevitable that at least some trust will always be required. This shortcoming can be reduced by education of users and open information sharing concerning the performance of operators.

- **Efficiency.** Security mechanisms, e.g. encryption algorithms and protocols, impose performance overhead onto a system. These overheads are typically in some way proportional to the level of security required. Users should expect this to be the case, although clearly effort should always be made to reduce these as far as possible.

  In the case of the MNPA the main additional overheads (above other systems) are in the use of the PRC and the more complex protocols needed to achieve privacy. Although the components of the MNPA were designed to be practical it seems likely that for more performance sensitive applications, such as broadband multimedia, these overheads may be too costly at the present time (this is not to say that in the future this will be such a problem, indeed it was until recently considered impractical to use public-key operations in mobile devices). As we discussed in the previous section, configuration of privacy will be necessary in order to cope with environments where the highest levels of security cannot be provided. Note that it is important the user retains knowledge and control of such choices.

- **Implementation.** The experimental work on the MNPA prototype was pleasing in terms of the series of question we asked ourselves beforehand (see 6.3.3). The main limitations we identified were the requirement for a secure operating environment for the MNPA components and the possible scaling problem of the encryption overhead.

During the design of the architecture we chose to assume such an environment to allow us to focus on the network aspects. This assumption may well choose to be a dangerous one, but we certainly hope that improvements are made in both future operating systems and education for systems administration.

The problem of encryption overhead is pertinent for two reasons. First, it is still a widely held assumption that mobile terminals have very limited resources. Second, that public key operations are prohibitive except in offline environments. We counter the first of these by predicting that as third generation multimedia capable terminals appear then the resource issue will decline in importance. The latter problem can be similarly addressed by ensuring that network providers keep up to date with upgrades to dedicated encryption hardware.

- **Legal Issues.** Perhaps the major problem with the MNPA is the requirements it aims to achieve – full privacy of user behaviour. Most western governments are under intense pressure from law enforcement agencies and an increasingly hysterical media to rid the world of the so-called 'four horsemen of the infocalypse', namely paedophiles, drug cartels, terrorists, and money-launderers.

  The result of this pressure looks likely to be new legislation to reduce rights to protect ones online privacy and increase investigatory powers. We note that few laws are currently in place, though many governments are proceeding with legislation in this area. Despite this unknown future we feel it is still extremely valuable to be able to demonstrate the highest levels of privacy. We discuss potential resolve of this problem in chapter 7.

6.5 Summary

This chapter has provided an evaluation of the MNPA. The evaluation consisted of presentation of an experimental prototype and the associated results followed by an overall evaluation of the project focusing on comparison with previous research results and the shortcomings of our work. We posed a series of questions that our prototype experiments set out to answer and we were indeed able to answer these favourably. Comparison with previous work and discussion of shortcomings of our work showed that the major aim of achieving user privacy in mobile networks has been fulfilled, but that there are some constraints on this achievement, most notably concerns about performance. These concerns are briefly addressed.

Chapter 7: Conclusions and Future Work

This thesis has presented a novel solution to what is a major obstacle to protecting the online behaviour of users of mobile networks. This comes at a time when there is great expectation of ubiquitous use of a new generation of mobile networks capable of high bandwidth multimedia. At the same time online privacy is a major concern amongst all levels of user. Privacy concerns are believed by many to be a significant inhibitor to the expansion of our online lives.

This final chapter presents a summary of the thesis, its findings and offerings. From the work we have done we also identify future directions to improve our architecture and associated issues. This is followed by concluding remarks on the research.

7.1 Thesis Summary

Chapter 1 of this thesis introduced the context of the work, namely distributed systems and the importance of security and privacy within them. It was identified that mobile networks do not offer a high degree of privacy and yet privacy is viewed as an inhibitor to online activity. Our aims for the thesis in achieving high levels of privacy mobile communications we presented next followed by a summary of the novelty of the results.

In chapter 2 and 3 we presented the background to the work. First, in chapter 2 we presented an introduction to communications networks, including mobile networks. This was followed by a discussion of computer and communications security. The last part of chapter 2 discussed privacy in detail including the legal implications. Chapter 3 introduced the requirements for security and privacy in mobile networks according to our aims [Askwith, et al., 1997]. The user must be able to achieve privacy against both parties external and internal to the network, and yet be able to satisfy the service provider requirements for fraud prevention. The main operations identified were as follows:

- Authentication

- Users must be able to register location updates with the local network without revealing identity.
- Users must be able to register location updates with the home network without revealing the actual location (the location is said to be pseudonymous).

- Accountability
  - Service providers must be able to collect recompense for any service provided to users without requiring the user to compromise their privacy.

- Communications
  - Users must be able to communicate with other hosts in the network and maintain the desired level of privacy.

The remainder of chapter 3 examined existing related work on privacy, focusing on mobile network solutions. We concluded from the literature that no existing proposal satisfies the requirements. In particular, solutions fall into two broad categories; those that cover mobile communications in a broad fashion but have a lower set of requirements for privacy and those that look at a less general scenario but have a high privacy requirement.

We then presented our solution to the requirements for user privacy in mobile networks in chapter 4. We called this the Mobile Network Privacy Architecture (MNPA) [Askwith, et al., 1998, Askwith, et al., 2000a, Askwith, et al., 2000b]. It extends the existing de facto topology by placing two new physical components into the network and supplying protocols, that utilize these components, to achieve registration and service provision accountability. The two new components are called the Privacy Routing Capability (PRC) and the Privacy Token Issuing Authority (PTIA).

The purpose of the PRC is to enable untraceable communications between two hosts in the network. It is an extension of an idea first presented in [Chaum,

1981], called digital-mixes. Our method reduces the real-time overhead for users [Askwith, et al., 2000b] by pre-computing the least efficient operations and removing the strongest security requirements found in some related solutions.

The PTIA is a third party distributed application resident in the MNPA, whose purpose is to manage privacy tokens. A privacy token is based on an anonymous public-key certificate (based on two public-keys; the user and the home network of the subscriber). Each home network collects an issuing certificate from the PTIA and then distributes complete privacy tokens to its users. These tokens then allow a user to engage in privacy-enhanced protocols with other parts of the network.

In addition to issuing privacy tokens the PTIA may also be required to act in dispute involving privacy tokens. If a network or a user is not satisfied by the result of part of a protocol then it may submit the data to the PTIA for evaluation. Although we do not discuss the precise legal operations required the technical solution is certainly novel.

Following the presentation of the MNPA we provide analysis in chapter 5. This consists of examining the type of attacks considered relevant to the architecture. Many attack scenarios are considered but the most problematic area for security is shown to be in system trust and in collusion within the MNPA. Trust is not a tangible quality of networks but we feel our solution lowers the requirements to a reasonable level. Collusion between the various elements of the MNPA is considered and we conclude that although such collusion may be useful for future requirements for law enforcement to utilize the risks are reasonably low currently.

A prototype implementation of the MNPA was presented in chapter 6. This implementation focuses on trying to determine the feasibility of the MNPA. The conclusions from the implementation were essentially favourable. Although we were able to determine little about the security effects, performance estimates seem within acceptable levels. The second part of chapter 6 provides an evaluation of the project. We discussed shortcomings, of which there are several,

and the positive aspects. On the positive side we have presented a novel framework into which flexible privacy policies may be developed for user privacy in mobile environments. In addition our results improve on previous published results in that they meet our requirements, as set out in chapter 3. However, the major problems affecting this framework are the less technical ones, people. Firstly political influence tends to be unfavourable towards privacy, and second it seems at this stage at least that problems of trust are both difficult to reason about and difficult to design into the framework.

7.2 Contributions

The MNPA takes a global approach to security and privacy in mobile networks, with particular emphasis on achieving high levels of privacy. The contribution of this work to the field is summarized as follows:

- We have identified a set of requirements for achieving security and privacy in mobile networks [Askwith, et al., 1997]. These requirements take into account both the system providers perspective and the user perspective. We believe our requirements for privacy to be more thorough than previous attempts. A critical analysis of previous research literature demonstrates this to be the case.

- We have developed a new architecture, the Mobile Network Privacy Architecture (MNPA) [Askwith, et al., 1998, Askwith, et al., 2000a], to meet our requirements. The MNPA has the following major contributions:
  - A new more efficient method for untraceable communications between two network hosts called the Privacy Routing Capability (PRC) [Askwith, et al., 2000b].
  - A novel third party application for enabling privacy enhanced communications based on anonymous certificates, called the Privacy Token Issuing Authority (PTIA). The PTIA is also capable of acting in disputes between users and service providers (where tokens are in use).

- A new location registration protocol that allows the update of user location information both locally and with the registered home network. Local update is provided with both mutual authentication and privacy of identification. Remote location update, that of updating the home network register, is provided without revealing the physical location, i.e. the location is pseudonymous.
- A new billing protocol that allows post-payment of services within a network. The protocol offers anonymity to the user and non-repudiation to both parties. Disputes can be solved by the PTIA.

- We have shown through analysis that the MNPA, and therefore user privacy in a mobile environment, is possible given certain constraints, such as well engineered software and robust cryptosystems. The main constraint is that the network itself must operate within certain user-network trust parameters. We have outlined these parameters informally.
- Our final contribution is to pose questions for further research. Details of these questions are given in the following section.

7.3 Future Work

So far in this chapter we have reiterated the project aims, findings and main results and considered the novel contributions of our work. While the contributions of this research are valuable it raises, as research should, some interesting questions. This section deals with, in our view, the more significant of these questions.

7.3.1 Integration with Existing Technologies

To increase the applicability of the MNPA theory we need to ensure it is flexible enough to cope with implementation in a variety of environments. We can narrow this down to two important scenarios, the Internet and 3rd Generation Systems (3Gs). Therefore it needs to be determined whether the MNPA could be adapted to work in these two important settings.

The Internet is currently based on the Internet Protocol (IP) though an updated protocol (IPv6) [Stallings, 1996] is being deployed in many networks. IPv6 will contain support for mobility. In addition a modified version of IP called MobileIP [Perkins, 1997] has been developed to combat the lack of mobility in IP. In addition IPSec [Kent and Atkinson, 1998] provides for confidentiality and authentication services in IP, and will be a part of IPv6 and Mobile IP.

Integrating the MNPA into these would require either modification of these protocols to include the ability to cope with our protocols of some method of encapsulating our protocols into them. By using encapsulation it may be possible to allow a mobile device to connect to non-MNPA compliant networks by using a PRC node to act as a home agent. This PRC node would not provide the authentication but rather would forward it as requested. We note however that this may require modification of the PRC to cope with this functionality.

3Gs are envisioned to be a heterogeneous set of various network types internetworked via a 3G core network. This is beneficial to any potential integration of the MNPA into 3Gs since compliant networks could operate our protocols on top of the core network. Those networks not part of the MNPA may be able to accept devices if they encapsulate MNPA protocols into the core access network protocols in a similar fashion described above.

7.3.2 Application of MNPA to Fixed Networking

Although mobility is going to be a major part of all future communications networks it is certain that fixed access will be prevalent. We consider that there are possibilities for adapting the MNPA for use with fixed terminals. First, the PRC can be used separately to achieve network anonymity and second the privacy tokens issued by the PTIA (and then by the home network to the user) may be used in conjunction with the PRC to allow users to connect to service providers anonymously and retain accountability.

Increasingly both business and consumers are looking to perform commercial transactions online, so called E-Commerce. The value of data in E-Commerce transactions will presumably be moderately high, in many cases deserving of high privacy protection. The MNPA could be ideal in such a situation, as it would allow a user to subscribe to a service provider who is capable of issuing privacy tokens. Note also, that the service provider does not need to be the same service provider that provides network level access. If E-Commerce service providers sign up to the PTIA network then they will be able to allow anonymous, yet accountable, subscription from users. The issue for further work is to examine what changes, if any, are required to the protocols to take account of the different nature of access.

7.3.3 Routing decisions within the PRC

In our discussion of the PRC we deliberately left out any decision as to how a mobile might choose a route through the PRC. This deliberate decision was made principally so we could give a clearer picture of the operation of the PRC but also since this problem is not a security issue. Clearly it remains a problem since the efficiency of the PRC is vital to the feasibility of the MNPA. The actual routing algorithm is not necessarily a problem since much work has already been done in this area[2], e.g. [Baker and Atkinson, 1997, Moy, 1998, Rehkter and Li, 1998]. The problem for the PRC is that it relies on Source Routing which has tended not to be used on the Internet due to it's problems with scalability (nodes need to be aware of routes, instead of allowing the network to handle the problem).

The future work for PRC routing lies in how the mobile terminal might achieve this operation well. Presumably the mobile terminal must be able to gain sufficient information from the access point to determine the global position and any relevant local and intermediate conditions affecting routing. A significant concern will be to balance the computation between terminal and network, it would not be reasonable to require the terminal to dynamically collect and store large routing tables.

[2] A search at http://liinwww.ira.uka.de/bibliography/Distributed/rfc.html reveals over 50 different RFCs relating to Internet routing.

In our evaluation of the PRC we noted that when a user moves location between sending a message through the PRC and receiving a reply the communication would fail. One suggestion we offer to combat this problem is to deposit with the local network a short-term PRC address at the point of handover. This may require a new protocol to be created for handover since the mobile user must be able to prepare a PRC route from the new location to the old one.

7.3.4 Access to Data for Law Enforcement

Apart from some discussion during the background chapters we have not considered the problem of law enforcement access to data within the MNPA. The main reasons for this are first that we wanted to focus on creating an architecture that was a simple as possible in meeting our requirements. Second, it is not clear what requirements law enforcement has, or whether these are justified (even without the complication of differing requirements each nation may propose). The civil rights issues of strong capabilities should be closely scrutinized. Third, as noted by [Abelson, et al., 1997], incorporating law enforcement access into communications system would almost certainly lead to the introduction of new vulnerabilities.

However, legislation is likely to dictate some form of access [Koops, 1997a] and service providers and network operators would clearly require solutions for compliance. Access to data is likely to be one of two forms, access to content and access to identification/location information. The simplest solution to the former would be some form of Key Recovery, despite the general inadvisability of these schemes. Key Recovery could be operated independently of the MNPA, perhaps through external Public Key Infrastructure (PKI) [Chokhani, 1994].

Access to identification/location information to allow law enforcement to identify traffic is more complicated since no one party, apart from the mobile user, holds both data. The home network needs to provide law enforcement with the identity associated to the appropriate privacy tokens. Given this, a protocol needs to be implemented to allow law enforcement to query the PRC to reveal the location of

the local network where the mobile user is currently registered. Monitoring of the local network for traffic associated with that user may still not reveal enough details since it one of the most important pieces of data for law enforcement is likely to be the identities of the suspects communications partners. Again if these are protected a protocol is required to query the PRC for this data.

7.3.5 Usability of Privacy Enhancing Technologies

The remaining questions for further work are all related to usability. First we look at general usability issues before discussing trust modeling and finally at configurability of privacy. Security is widely understood to be very difficult to achieve, one major impediment to which is the human element. From a rather unscientific viewpoint if a mistake can be made to compromise security then a user will almost certainly make that mistake at some point. This is backed up by research into the Human Computer Interface (HCI) of PGP [Whitten and Tygar, 1999]. Users struggle with security because the concepts are often quite complex and the goals are usually at odds with the goals of the task in hand, i.e. security can been seen as trying to stop certain things happening rather than making them happen.

The MNPA is a complex architecture potentially requiring knowledge of many different parameters to achieve the users' privacy goals. It is unreasonable to expect non-expert users to be able to navigate all these decisions safely without a very carefully thought out user interface. An interface needs to be able to make as much of the underlying technology transparent to the user yet also inform the user of the privacy implications of their decisions, including trust requirements which we examine next.

A further consideration would be to aid the developer community through, for example, a Privacy Middleware. Such software would abstract the privacy and security mechanisms in such a way that the developer could more easily incorporate them into their applications.

7.3.6 Reasoning about Trust and Privacy

During our analysis we discussed the trust issues the user needs to consider when using the MNPA. These decisions are largely subjective and thus essentially a risk management exercise. This is obviously not an ideal situation and it would be preferable if there existed some way of formally reasoning about trust scenarios. We are not currently aware of any such capabilities with regard to privacy.

Also, privacy itself is a complex concept that requires more formal analysis. When a user interacts with an element of a system it would be useful to determine what information is at risk and what information may be provided without concern. The parameters for privacy analysis within the MNPA are complex due to the number of parties involved.

In common with the last subsection any reasoning about trust and privacy is also a usability issue since a user needs to be able to make decisions about their privacy requirements on the spot. Therefore any results of trust modeling need to be useful to the user as well as developers and researchers. Tools provided to researchers and developers would be very different from those needed by users since the user is only interested in themselves whereas the planning of entire systems, including interactions between elements is the concern in research and development.

7.3.7 Configuring User Privacy

Differing levels of privacy are likely to be required depending on the situation a user find themselves in. Factors likely to affect a decision about a suitable privacy level will include; perceived value of data involved, performance requirements (e.g. provision of video might require low security overheads), local security policies, communications partners. Any one of these, amongst many others, could influence a degradation of privacy.

Privacy levels may range from no privacy, though this might not be advisable from a system security perspective, through to the highest level we have sought in this work. In chapter 3 we described six basic levels of privacy beginning and

ending with these two. The levels in between describe states that protect information from external attacks, local network attacks, and home network attacks respectively. Whilst this categorization has proved useful for analyzing existing results and setting a target for the highest privacy, it is probably not flexible enough to describe user policies.

Future work in this direction then, requires first that a way of describing privacy levels is developed that can be used to allow a user to describe their privacy policy. Following this we need a way to translate this policy so that the correct protocol parameters are used. Ideally no serious alterations to protocols should occur, but clearly a need for more careful examination of how parameters might be included is needed.

7.3.7 Denial of Service

Recently attacks on the availability of networks services have become more common and increasingly serious. These attacks are called Denial of Service attacks (DoS) [Needham, 1994b]. The reason that they are common is because they are often very easy to mount due to the fact that they abuse normal acceptable operations within a service. A typical example of a DoS attack is when an attacker floods a web server with bogus requests. If enough requests are made together then the act of processing all requests could cause the server to overload and crash.

In the MNPA we have not given priority to DoS attacks as these are lower in priority to a user than privacy. However, future development of the MNPA should include consideration of this class of attack. Places major targets for DoS attacks are the two protocols and the PRC. The two protocols need to be able to detect false requests ideally without performing the whole protocol. Although the protocols partly deal with this (via token checking), it could be improved. The PRC could be flooded with data by simple a DoS attack. An attacker could attempt to send very large messages that carry no payload apart from PRC hop information and have no destination. Some kind of access control to the PRC might solve this potential problem.

7.4 Summary

Communications networks are becoming an increasingly important part of everyday life throughout the world with the rapid expansion of both the Internet and mobile telephony. Third generations mobile systems that will carry a wide variety of data on behalf of users are soon to be introduced. Amidst all this information revolution users are becoming alarmed at the possibilities for privacy intrusions from both external sources (i.e. crackers) and internal sources (i.e. corporations and governments). The aim of this thesis was to investigate and provide solutions to strong privacy in mobile communications.

We set about defining the problem so as to gain a set of requirements. These requirements were threefold, first, to be able to perform mutual authentication with the local and home networks to achieve location registration whilst maintaining location and identification privacy. Second, to be able to receive service from a provider without privacy compromise whilst also being able to maintain accountability for and service received. Finally, we needed to be able to communicate with any host in the network in an anonymous way.

Existing solutions related to privacy in mobile communications were shown to fail to meet all of our requirements, though many are excellent solutions in terms of their intended goals. Two major problems emerged from the literature, first, many solutions aim for lower privacy goals than those we wished to achieve and secondly many related solutions concern themselves with a narrower target than the security and privacy of the whole network.

The main direction of our work has been to combine network anonymity with privacy enhancing technologies to achieve privacy for all user behaviour thereby satisfying our initial requirements [Askwith, et al., 1997]. The result is a new architecture entitled the Mobile Network Privacy Architecture (MNPA) [Askwith, et al., 1998, Askwith, et al., 2000a]. The MNPA describes two supporting physical components, the Privacy Routing Capability (PRC) [Askwith, et al., 2000b] and the Privacy Token Issuing Authority (PTIA) [Askwith, et al., 2000a].

The PRC enables network anonymity for individual messages whilst the PTIA assists in creating anonymous authorization tickets for users. In addition to these components are two new privacy-enhancing protocols, one for location registration and update, and one for service billing. These protocols protect against intrusions of privacy from external and internal attacks.

Our analysis of the MNPA showed that given awareness of certain operational assumptions the desired security of the system is achievable. Additionally, we believe that the architecture and its components are flexible enough to cope with the new developments in mobile networking. This presents exciting possibilities for the future of communications, possibilities that are currently being inhibited by user concerns over online privacy. It is our thesis therefore that the MNPA can contribute to the development of mobile networking in the 21st century by offering a realistic approach to satisfying strong user privacy requirements.

References

[Abadi, et al., 1997] Abadi, M., Lomas, T. M. A., and Needham, R., "Strengthening Passwords," DEC System Research Center, Technical Report DEC-SRC-1997-033, 16th December 1997.

[Abelson, et al., 1997] Abelson, H., Anderson, R., Bellovin, S. M., Benaloh, J., Blaze, M., *et al.*, "The Risks of Key Recovery, Key Escrow, and Trusted Third Party Encryption," available online: http://research.att.com/dist/mab/, 27th May 1997.

[AES, 2000] Advanced Encryption Standard Development Effort Web Site (NIST), http://csrc.nist.gov/encryption/aes/

[Agre and Rotenberg, 1998] Agre, P. and Rotenberg, M., *Technology and Privacy: The New Landscape*. Cambridge, Mass, USA: MIT Press, 1998.

[Akyildiz, et al., 1999] Akyildiz, I. F., McNair, J., Ho, J. S. M., Uzunalioglu, H., and Wang, W., "Mobility Management in Next-Generation Wireless Systems," *Proceedings of the IEEE*, vol. 87, no 8, pp. 1347-1384, 1999.

[Anderson, 1994] Anderson, R. J., "Why Cryptosystems Fail," *Communications of the ACM*, vol. 37, no 11, pp. 32-40, 1994.

[Anderson and Kuhn, 1996] Anderson, R. and Kuhn, M., "Tamper Resistance - A Cautionary Note," in Proceedings of USENIX Workshop on Electronic Commerce, Oakland, California, USA, 1996.

[Anderson and Needham, 1995] Anderson, R. and Needham, R., "Robustness Principles for Public Key Protocols," in Proceedings of CRYPTO' 95, Santa Barbra, CA, USA, 1995.

[Araki, et al., 1998] Araki, K., Satoh, T., and Miura, S., "Overview of Elliptic Curve Cryptography," in Proceedings of 1st International Workshop on Practice and Theory in Public Key Cryptography, 1998.

[Askwith, et al., 1997] Askwith, B., Merabti, M., Shi, Q., and Whiteley, K., "Achieving User Privacy in Mobile Networks," in Proceedings of Computer Security Applications Conference, San Diego, California, USA, December 1997.

[Askwith, et al., 1998] Askwith, B., Merabti, M., and Shi, Q., "Accountable Anonymity in Mobile Communications," in Proceedings of MoMuC'98, Berlin, Germany, October 1998.

[Askwith, et al., 2000a] Askwith, B., Merabti, M., and Shi, Q., "Privacy Routing Capability: An Efficient Anonymity Scheme for Low-Power Users," in Proceedings of 1st Annual Postgraduate Symposium on the Convergence of Telecommunications, Networking and Broadcasting (PGNet 2000), Liverpool John Moores University, Liverpool, June 2000.

[Askwith, et al., 2000b] Askwith, B., Merabti, M., and Shi, Q., "MNPA: A Mobile Network Privacy Architecture," *Computer Communications,* vol. 23, pp. 1777-1788, 2000.

[Asokan, 1994] Asokan, N., "Anonymity in a Mobile Computing Environment," in Proceedings of Workshop on Mobile Computing Systems, Santa Cruz, California, USA, 1994.

[ASPeCT, 1998] Advanced Security For Personal Communications Technologies (ASPeCT) Web Site, http://www.esat.kuleuven.ac.be/cosic/aspect/

[Aziz and Diffie, 1994] Aziz, A. and Diffie, W., "Privacy and Authentication for Wireless Local Area Networks," *IEEE Personal Communications*, vol. 1, no 1, pp. 25-31, 1994.

[Bainbridge and Pearce, 2000] Bainbridge, D. and Pearce, G., "Titling at Windmills - Has the New Data Protection Law failed to make a Significant Contribution to Rights and Privacy," *Journal of Informnation, Law and Technology (JILT)*, vol. 5, no. 2, available online: http://elj.warwick.ac.uk/jilt/00-2/bainbridge.html, 2000.

[Baker and Atkinson, 1997] Baker, F. and Atkinson, R., "RFC 2082 - Routing Information Protocol v2 - MD5 Authentication," IETF Request for Comments, available online: http://www.ietf.org/rfc/rfc2082.txt, January 1997.

[Bellare, et al., 1995] Bellare, M., Garay, J. A., Hauser, R., Herzberg, A., Krawczyk, H., *et al.*, "*i*KP - A Family of Secure Electronic Payment Protocols," in Proceedings of First USENIX Workshop on Electronic Commerce, New York, NY, USA, 1995.

[Beller, et al., 1993] Beller, M. J., Chang, L., and Yacobi, Y., "Privacy and Authentication on a Portable Communications System," *IEEE Journal on Selected Areas In Communications*, vol. 11, no 6, pp. 821-829, 1993.

[Benassi, 1999] Benassi, P., "TRUSTe: An Online Privacy Seal Program," *Communications of the ACM*, vol. 42, no 2, pp. 56-59, 1999.

[Bharghavan, 1994] Bharghavan, V., "Secure Wireless LANs," in Proceedings of ACM Conference on Computer Security, Fairfax, Virginia, USA, 1994.

[Bharghavan, 1995] Bharghavan, V., "A Protocol for Authentication, Data and Location Privacy, and Accounting in Mobile Communications," University of Illinois Urbana Champaign, Tech. Report, available online: http://shiva.crhc.uuic.edu/Papers/Sec.ps.gz, 1995.

[Bharghavan and Ramamoorthy, 1995] Bharghavan, V. and Ramamoorthy, C. V., "Security Issues in Mobile Communications," in Proceedings of International Symposium on Autonomous Decentralised Systems, 1995.

[Bird, 1995] Bird, R., Gopal, I, Herzberg, A, Janson, P, Kutten, S, and Yung, M, "The KryptoKnight Family of Light-Weight Protocols for Authentication and Key Distribution," *IEEE Transactions on Networking*, vol. 3, no 1, pp. 31-41, 1995.

[Boneh, 1999] Boneh, D., "Twenty Years of Attacks on the RSA Cryptosystem," *Notices of the AMS*, no 2, pp. 203-213, 1999.

[Boyd and Mao, 1993] Boyd, C. and Mao, W., "On a Limitation of a BAN Logic," in Proceedings of EUROCRYPT '93, 1993.

[BP, 2000] Bletchley Park Web Site, http://www.bletchley-park.co.uk/

[Brands, 1993] Brands, S., "Untraceable Off-line Cash in Wallets with Observers," in Proceedings of CRYPTO 93, Santa Barbara, California, USA, 1993.

[Burrows, et al., 1990] Burrows, M., Abadi, M., and Needham, R., "A Logic of Authentication," DEC, Palo Alto, Tech. Report SRC 39, February 22 1990.

[Buschkes, et al., 1998] Buschkes, R., Kesdogan, D., and Reichl, P., "How to Increase Security in Mobile Networks by Anomaly Detection," in Proceedings of 14th Annual Computer Security Applications Conference, Phoenix, Arizona, USA, 1998.

[Campbell, 1999] Campbell, D., "Interception Capabilities 2000," European Parliament, STOA Technical Report, available online: http://www.iptvreports.mcmail.com/ic2kreport.htm, April 1999.

[Campbell, 1999] Campbell, D., "Special Investigation: ILETS and the ENFOPOL Affair," *Telepolis,* available online: http://www.heise.de/tp/english/special/enfo/6398/1.html, 1999.

[Carlsen, 1994] Carlsen, U., "Optimal Privacy and Authentication on a Portable Communications System," *Operating Systems Review*, vol. 28, no 3, pp. 11-23, 1994.

[CCITSE, 1996] Common Criteria Partners Website, http://www.radium.ncsc.mil/tpep/library/ccitse/index.html, 1996.

[CERT, 1999] CERT Coordination Center "CERT® Advisory CA-1999-04 Melissa Macro Virus", available online: http://www.cert.org/advisories/CA-1999-04.html, March 31st 1999.

[Chan, et al., 1998] Chan, A., Frankel, Y., and Tsiounis, Y., "Easy Come - Easy Go Divisible Cash," in Proceedings of EUROCRYPT '98, Espoo, Finland, 1998.

[Chaum, 1981] Chaum, D. L., "Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms," *Communications of the ACM*, vol. 24, no 2, pp. 84-88, 1981.

[Chaum, 1982] Chaum, D., "Blind Signatures for Untraceable Payments," in Proceedings of CYRPTO '82, Santa Barbra, CA, USA, 1982.

[Chaum, 1989] Chaum, D., "Online Cash Checks," in Proceedings of EUROCRYPT '89, 1989.

[Chaum, et al., 1988] Chaum, D., Fiat, A., and Naor, M., "Untraceable Electronic Cash," in Proceedings of CRYPTO '88, 1988.

[Chen and Mitchell, 1997] Chen, L. and Mitchell, C. J., "An Anonymous and Undeniable Payment Scheme," in Proceedings of First International Conference on Information and Communications Security, ICICS '97, Beijing, China, 1997.

[Chokhani, 1994] Chokhani, S., "Towards a National Public Key Infrastructure," *IEEE Communications Magazine*, vol. 27, no 9, pp. 70-74, 1994.

[Clarke, 1997] Clarke, R., "Data Surveillance: Theory, Practice & Policy," Australian National University, Canberra, description of thesis by submission of published work, available online: http://www.anu.edu.au/people/Roger.Clarke/DV/PhD.html, 17th January 1997.

[Cooper, 1998] Cooper, D. A., "A Closer Look at Revocation and Key Compromise in Public Key Infrastructures," in Proceedings of NISSC '98, Crystal City, VA, USA, 1998.

[Cooper and Birman, 1995] Cooper, D. A. and Birman, K. P., "Preserving Privacy in a Network of Mobile Computers," in Proceedings of IEEE Symposium on Research in Computing, Oakland, California, USA, 1995.

[Coulouris, et al., 1994] Coulouris, G., Dollimore, J., and Kindberg, T., *Distributed Systems Concepts and Design*, 2nd ed., Addison-Wesley, 1994.

[CRCL, 2000] Cyber-Rights and Cyber-Liberties Web Site, http://www.cyber-rights.org/

[CSI, 2000] Computer Security Institute (CSI), "Computer Security: Issues and Trends," San Fransisco, CA, USA, Survey VI/1, Spring 2000.

[Cullen and Lobley, 1996] Cullen, J. M. and Lobley, N. C., "The Universal Mobile Telecommunications System - a Mobile Network for the 21st Century," *BT Technology Journal*, vol. 14, no 3, pp. 123-131, 1996.

[Davida, et al., 1997] Davida, G., Frankel, Y., Tsiounis, Y., and Yung, M., "Anonymity Control in E-Cash Systems," in Proceedings of Financial Cryptography FC '97, Anguilla, British West Indies, 1997.

[Denning, 1994] Denning, D. E., "The US Key Escrow Encryption Technology," *Computer Communications*, vol. 17, no 7, pp. 453-457, 1994.

[Denning and Baugh, 1997] Denning, D. E. and Baugh, W. E., "Cases Involving Encryption in Crime and Terrorism," Georgetown University, available online: http://www.cs.georgetown.edu/~denning, 10th October 1997.

[Diffie and Hellman, 1976] Diffie, W. and Hellman, M. E., "New Directions in Cryptography," *IEEE Transactions on Information Theory*, vol. 22, no 6, pp. 644-654, 1976.

[DTI, 1999] DTI, "A Report Summarising Responses to: Building Confidence in Electronic Commerce - A Consultation Document," Department of Trade and Industry (DTI), UK Government, London, URN 99/891, July 1999.

[Dukach, 1992] Dukach, S., "SNPP: A Simple Network Payment Protocol," in Proceedings of Computer Security Applications Conference, 1992.

[Economist, 1999] Economist, "The end of privacy," in *The Economist*, 1999, available online: http://www.economist.com/.

[EFF, 1997] Electronic Freedom Foundation Web Site, http://www.eff.org

[Ellis, 1997] Ellis, J. H., "The Story of Non-Secret Encryption," CESG Report 16th December 1997.

[Entrust, 2000] Entrust Web Site, http://www.entrust.com/

[EPIC, 1997] Electronic Privacy Information Center Web Site, http://www.epic.org

[Fasbender, et al., 1996] Fasbender, A., Kesdogan, D., and Kubitz, O., "Analysis of Security and Privacy in Mobile IP," in Proceedings of 4th International Conference on Telecommunications Systems Modelling and Analysis, Nashville, Tennessee, USA, 1996.

[Fasbender, et al., 1996] Fasbender, A., Kesdogan, D., and Kubitz, O., "Variable and Scalable Security: Protection of Location Information in Mobile IP," in Proceedings of Vehicular Transport Conference (VTC 96), Atlanta, Georgia, USA, 1996.

[Federrath, et al., 1995] Federrath, H., Jerichow, A., Kesdogan, D., and Pfitzman, A., "Security in Public Mobile Communication Networks," in Proceedings of IFIP/TC6 Personal Wireless Communications, Prague, Czech Republic, 1995.

[Federrath, et al., 1996] Federrath, H., Jerichow, A., and Pfitzman, A., "MIXes in Mobile Communication Systems: Location Management with Privacy," in Proceedings of Information Hiding, Cambridge, UK, 1996.

[FIPR, 2000] Foundation for Information Policy Research Web Site, http://www.fipr.org/

[Fox and Gribble, 1996] Fox, A. and Gribble, S., "Security on the Move: Indirect Authentication using Kerberos," in Proceedings of MOBICOM '96, Rye, NY, USA, 1996.

[France, 1999] France, E., "Data Protection Act 1998: Preparing for the New Law," in Proceedings of Compters, Freedom and Privacy (CFP99), Washington, DC, USA, 1999.

[Franklin and Reiter, 1996] Franklin, M. K. and Reiter, M. K., "The Design and Implementation of a Secure Auction Service," *IEEE Transactions on Software Engineering*, vol. 22, no 5, pp. 302-312, 1996.

[Franz and Jerichow, 1998] Franz, E. and Jerichow, A., "A Mix-Mediated Anonymity Service and its Payment," in Proceedings of ESORICS '98, Louvain-la-Neuve, Belgium, 1998.

[Froomkin, 2000] Froomkin, A. M., "The Death of Privacy," *Stanford Law Review*, vol. 52, pp. 1461-1543, 2000.

[Fujisaki and Okamoto, 1998] Fujisaki, E. and Okamoto, T., "Practical Escrow Cash Schemes," *IEICE Trans. Fundamentals*, vol. E81-A, no 1, pp. 11-19, 1998.

[Gladman, et al., 1999] Gladman, B., Ellison, C., and Bohm, N., "Digital Signatures, Certificates and Electronic Commerce," avaialble online: http://www.btinternet.com/~brian.gladman/, 8th June 1999.

[Gladman, 2000] Gladman, B., "The Regulation of Investigatory Powers Bill - The Provision for Government Access to Keys," FIPR Report, February 2000.

[Goldreich, 1997] Goldreich, O., "On the Foundations of Modern Cryptography," in Proceedings of CRYPTO '97, Santa Barbara, CA, USA, 1997.

[Goldschlag, et al., 1999] Goldschlag, D., Reed, M., and Syverson, P., "Onion Routing for Anonymous Internet Connections," *Communications of the ACM*, vol. 42, no 2, pp. 39-41, 1999.

[Graham, 1998] Graham, R., "FAQ: Network Intrusion Detection Systems," available from robert_david_graham@yahoo.com, 1998.

[Gregory, et al., 1998] Gregory, D., Shi, Q., and Merabti, M., "An Intrusion Detection System based upon Autonomous Mobile Agents," in

Proceedings of IFIP TC 11 14th International Conference in Information Security (SEC 98), Vienna, Austria and Budapest, Hungary, 1998.

[Groves and Clapton, 1996] Groves, I. S. and Clapton, A. J., "Third Generation Mobile Systems," *BT Technology Journal*, vol. 14, no 3, pp. 115-121, 1996.

[GSMWorld, 2000] GSMWorld Web Site, http://www.gsmworld.com/

[Guillou and Ugon, 1986] Guillou, L. C. and Ugon, M., "Smart Card: A Highly Reliable and Portable Security Device," in Proceedings of CRYPTO '86, Santa Barbara, 1986.

[Gulcu and Tsudik, 1996] Gulcu, C. and Tsudik, G., "Mixing E-mail with BABEL," in Proceedings of SNDSS, San Diego, California, USA, 1996.

[Halsall, 1996] Halsall, F., *Data Communications, Computer Networks and Open Systems*, 4th ed: Addison Wesley, 1996.

[Hardjono and Seberry, 1996] Hardjono, T. and Seberry, J., "Security Issues in Mobile Information Networks," *IEICE Trans. Fundamentals.*, vol. E79-A, no 7, pp. 1021-1026, 1996.

[Herzberg, et al., 1994] Herzberg, A., Krawczyk, H., and Tsudik, G., "On Travelling Incognito," in Proceedings of IEEE Workshop on Mobile Systems and Applications, 1994.

[Hoff, et al., 1996] Hoff, S., Jakobs, K., and Kesdogan, D., "Anonymous Mobility Management for Third Generation Mobile Networks," in Proceedings of IFIP/TC6&TC11 Communications and Multimedia Security, Essen, Germany, 1996.

[Horn and Preneel, 1998] Horn, G. and Preneel, B., "Authentication and Payment in Future Mobile Systems," in Proceedings of ESORICS '98, Louvain-la-Neuve, Belgium, 1998.

[ITSEC, 2000] The UK Government ITSEC Web Site, http://www.itsec.gov.uk/

[Jain, et al., 2000] Jain, A., Hong, L., and Pankanti, S., "Biometrics Identification," *Communications of the ACM*, vol. 43, no 2, pp. 91-98, 2000.

[Jakobsson, 1998] Jakobsson, M., "A Practical Mix," in Proceedings of EUROCRYPT '98, Espoo, Finland, 1998.

[Jakobsson and Juels, 1998] Jakobsson, M. and Juels, A., "X-Cash: Executable Digital Cash," in Proceedings of Financial Cryptography (FC'98), Anguilla, British West Indies, 1998.

[Jakobsson and M'Raihi, 1998] Jakobsson, M. and M'Raihi, D., "Mix-based Electronic Payments," in Proceedings of SAC'98, 1998.

[Jerichow, et al., 1998] Jerichow, A., Muller, J., Pfitzmann, A., Pfitzmann, B., and Waidner, M., "Real-Time Mixes: A Bandwidth-Efficient Anonymity Protocol," *IEEE Journal on Selected Areas in Communications*, vol. 16, no 4 pp. 495-509, 1998.

[Jones, 1998] Jones, C., "Bad Days for Software," *IEEE Spectrum*, no 9, pp. 47-52, 1998.

[Kazovsky, et al., 1998] Kazovsky, L. G., Khoe, G. D., and van Deventer, M. O., "Future Telecommunication Networks: Major Trend Projections," *IEEE Communications Magazine*, no 11, pp. 122-127, 1998.

[Kent and Atkinson, 1998] Kent, S., and Atkinson, R., "Security Architecture for the Internet Protocol," IETF Request for Comments 2401, available online: http://www.ietf.org/rfc/rfc2401.txt, November 1998.

[Kesdogan, et al., 1998] Kesdogan, D., Egner, J., and Buschkes, R., "Stop-And-Go-MIXes Providing Probabilistic Anonymity in an Open System," in Proceedings of Workshop on Information Hiding, Portland Oregon, USA, 1998.

[Koops, 1997] Koops, B., "Crypto Regulation in Europe: Some Key Trends and Issues," *Computer Networks and ISDN Systems*, vol. 29 pp. 1823-1831, 1997.

[Lampson, 1983] B. W. Lampson, "Hints for Computer System Design," *ACM Operating Systems Review*, vol. 15, no. 5, pp. 33-48, 1983.

[Laudon, 1996] Laudon, K. C., "Markets and Privacy," *Communications of the ACM*, vol. 39, no 9, pp. 92-104, 1996.

[Leiner, et al., 1997] Leiner, B. M., Cerf, V. G., Clark, D. D., Kahn, R. E., Kleinrock, L., *et al.*, "A Brief History of the Internet," *Communications of the ACM*, no 2, 1997.

[Li and Liao, 1997] Li, V. O. K. and Liao, W., "Personal Communication Systems," *Proceedings of the IEEE*, vol. 85, no 7, pp. 1063-1108, 1997.

[Litman, 2000] Litman, J., "Information Privacy / Information Property," *Stanford Law Review*, vol. 52, pp. 1283-1313, 2000.

[Loscocco, et al., 1998] Loscocco, P. L., Smalley, S. D., Muckelbauer, P. A., Taylor, R. C., Turner, S. J., *et al.*, "The Inevitability of Failure: The Flawed Assumption of Security in Modern Computing Environments," in Proceedings of NISSC '98, Baltimore, MD, USA, 1998.

[MacDonald, 1979] MacDonald, V. H., "Advanced Mobile Phone Service: The Cellular Concept," *The Bell System Technical Journal*, vol. 58, no 1, pp. 15-41, 1979.

[McGraw and Felten, 1998] McGraw, G. and Felten, E. W., "Mobile Code and Security," *IEEE Internet Computing*, no 6, pp. 26-29, 1998.

[Medvinsky and Neuman, 1993] Medvinsky, G. and Neuman, B. C., "NetCash: A Design for Practical Electronic Currency on the Internet," in Proceedings of ACM Computer and Communications Security, 1993.

[Menezes and Vanstone, 1993] Menezes, A. J. and Vanstone, S. A., "Elliptic Curve Cryptosystems and Their Implementations," *Journal of Cryptology*, vol. 6, no 4, pp. 209-224, 1993.

[MIDS, 1998] MIDS, "More than 100 Million Internet Users," Matrix Information and Directory Services, Austin, Texas, Press Release, available online: http://www.mids.org/press/pr199801.html, June 29 1998.

[Mohan, 1996] Mohan, S., "Privacy and Authentication Protocols for PCS," *IEEE Personal Communications*, vol. 3, no 5, pp. 34-38, 1996.

[Molva, et al., 1992] Molva, R., Tsudik, G., Herreweghen, E. v., and Zatti, S., "KryptoKnight Authentication and Key Distribution System," in Proceedings of ESORICS '92, Toulouse, France, 1992.

[Molva, et al., 1994] Molva, R., Samfat, D., and Tsudik, G., "Authentication of Mobile Users," *IEEE Network*, vol. 8, no 2, pp. 26-34, 1994.

[Morris and Thompson, 1979] Morris, R. and Thompson, K., "Password Security: A Case History," *Communications of the ACM*, vol. 22, no 11, pp. 594-597, 1979.

[Morton, 1999] Morton, D., "The Electrical Century: Radio Boradcasting in the Electrical Century," *Proceedings of the IEEE*, vol. 87, no 5, pp. 929-932, 1999.

[Moy, 1998] Moy, J., "RFC 2328 - Open Shortest Path First v2," IETF Request for Comments, available online: http://www.ietf.org/rfc/rfc2328.txt, April 1998.

[Mu and Varadharajan, 1996] Mu, Y. and Varadharajan, V., "On the Design of Security Protocols for Mobile Communications," in Proceedings of ACISP 96, Wollongong, NSW, Australia, 1996.

[Mukherjee, et al., 1994] Mukherjee, B., Heberlein, L. T., and Levitt, K. N., "Network Intrusion Detection," *IEEE Network*, vol. 8, no 3, pp. 26-41, 1994.

[Naor and Nissim, 2000] Naor, M. and Nissim, K., "Certificate Revocation and Certificate Update," *IEEE Journal on Selected Areas in Communications*, vol. 18, no 4, pp. 561-570, 2000.

[NBS, 1977] NBS, "Data Encryption Standard," NBS, Washington D.C. USA, FIPS Pub. 46, Jan 1977.

[Needham, 1994] Needham, R. M., "Denial of Service: An Example," *Communications of the ACM*, vol. 37, no 11, pp. 42-46, 1994.

[Netscape, 1996] Netscape, "Secure Sockets Layer," available online: http://www.netscape.com/, 1996.

[Neuman and Ts'o, 1994] Neuman, B. C. and Ts'o, T., "Kerberos: An Authentication Service for Computer Networks," *IEEE Communications Magazine*, vol. 27, no 9, pp. 33-38, 1994.

[NIST, 2000] NIST Trusted Computer System Evaluation Criteria Web Site, http://www.radium.ncsc.mil/tpep/library/tcsec/index.html

[NUA, 2000] Nua Ltd Web Site, http://www.nua.ie/

[Okamoto and Ohta, 1991] Okamoto, T. and Ohta, K., "Universal Electronic Cash," in Proceedings of CRYPTO '91, Santa Barbra, CA, USA, 1991.

[O'Mahony, 1998] O'Mahony, D., "UMTS: The Fusion of Fixed and Mobile Networking," *IEEE Internet Computing*, no 1, pp. 49-55, 1998.

[Orwell, 1949] Orwell, G., *Nineteen Eighty-Four*: Penguin (1954), 1949.

[Pandya, et al., 1997] Pandya, R., Grillo, D., Lycksell, E., Mieybegue, P., Okinaka, H., *et al.*, "IMT-2000: Network Aspects," *IEEE Personal Communications*, vol. 4, no 4, pp. 20-29, 1997.

[Patel and Crowcroft, 1997] Patel, B. and Crowcroft, J., "Ticket Based Service Access for the Mobile User," in Proceedings of MobiCom 97, Budapest, Hungary, 1997.

[Pedersen, 1997] Pedersen, T. P., "Electronic Payments of Small Amounts," in Proceedings of Security Protocols '97, France, 1997.

[Perkins, 1997] Perkins, C. E., "Mobile IP," *IEEE Communications Magazine*, vol. 35, no 5, pp. 84-99, 1997.

[Perlman, 1999] Perlman, R., "An Overview of PKI Trust Models," *IEEE Network*, no 6, pp. 38-43, 1999.

[Petersen and Poupard, 1997] Petersen, H. and Poupard, G., "Efficient Scalable Fair Cash with Off-line Extortion Prevention," in Proceedings of First International Conference on Information and Communications Security, ICICS '97, Beijing, China, 1997.

[Pfleeger, 1996] Pfleeger, C. P., *Security In Computing*, 2nd ed: Prentice Hall International, 1996.

[PGP, 1999] PGP International Web Site, http://www.pgpi.com/

[Privacy International, 1997] Privacy International Web Site, http://www.privacy.org

[Radu, et al., 1997] Radu, C., Govaerts, R., and Vandewalle, J., "Efficient Electronic Cash with Restricted Privacy," in Proceedings of Financial Cryptography '97, Anguilla, BWI, 1997.

[Rahnema, 1993] Rahnema, M., "Overview of the GSM System and Protocol Architecture," *IEEE Communications Magazine*, vol. 26, no 4, pp. 92-100, 1993.

[Rantos and Mitchell, 1999] Rantos, K. and Mitchell, C., "Key Recovery in ASPeCT Authentication and Initialisation of Payment Protocol," in Proceedings of 4th ACTS Mobile Communications Summit, Sorrento, Italy, 1999.

[Raychaudhuri, 1999] Raychaudhuri, D., "Wireless ATM Networks: Technology Status and Future Directions," *Proceedings of the IEEE*, vol. 87, no 10, pp. 1790-1806, 1999.

[Reagle and Cranor, 1999] Reagle, J. and Cranor, L. F., "The Platform for Privacy Preferences," *Communications of the ACM*, vol. 42, no 2, pp. 48-55, 1999.

[Reed, et al., 1996] Reed, M., Syverson, P., and Goldschlag, D., "Proxies for Anonymous Routing," in Proceedings of 12th Annual Computer Security Applications Conference, San Diego, 1996.

[Reed, et al., 1998] Reed, M. G., Syverson, P. F., and Goldschlag, D. M., "Anonymous Connections and Onion Routing," *IEEE Journal of Selected Areas in Communications*, vol. 16, no 6, 1998.

[Rehkter and Li, 1998] Rehkter, Y. and Li, T., "Border Gateway Protocol v4 BGP-4," IETF, Internet Draft draft-ietf-idr-bgp4-08.txt, April 1998.

[Reichenbach, et al., 1997] Reichenbach, M., Damker, H., Federrath, H., and Rannenburg, K., "Individual Management of Personal Reachability in Mobile Communication," in Proceedings of SEC '97, 1997.

[Reiter and Rubin, 1998] Reiter, M. K. and Rubin, A. V., "Crowds: Anonymity for Web Transactions," *ACM Transactions on Information and System Security*, vol. 1, no 1, pp. 66-92, 1998.

[Reiter and Rubin, 1999] Reiter, M. K. and Rubin, A. V., "Anonymous Web Transactions with Crowds," *Communications of the ACM*, vol. 42, no 2, pp. 32-38, 1999.

[Risks, 2000] Risk List Archive, available online: http://catless.ncl.ac.uk/

[Rivest, et al., 1978] Rivest, R. L., Shamir, A., and Adleman, L., "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems," *Communications of the ACM*, vol. 21, no 2, pp. 120-126, 1978.

[Rivest, 1990] Rivest, R. L., "Cryptography," in *Handbook of Theoretical Comouter Science*, van Leeuwen, J. (Editor).: Elsevier Science, 1990, pp. 718-755.

[Rivest, 1992] Rivest, D. L., "The MD5 Message Digest Algorithm," Internet Request for Comments 1321, available online: http://www.ietf.org/rfc/rfc1321.txt, April 1992.

[Rubin and Geer, 1998] Rubin, A. D. and Geer, D. E., "Mobile Code Security," *IEEE Internet Computing*, no 6, pp. 30-34, 1998.

[Samfat and Molva, 1994] Samfat, D. and Molva, R., "A Method Providing Identity Privacy to Mobile Users during Authentication," in Proceedings of Workshop on Mobile Computing Systems, Santa Cruz, California, USA, 1994.

[Samfat and Molva, 1997] Samfat, D. and Molva, R., "IDAMN: An Intrusion Detection Architecture for Mobile Networks," *IEEE Journal on Selected Areas in Communications*, vol. 15, no 7, pp. 1373-1380, 1997.

[Samfat, et al., 1995] Samfat, D., Molva, R., and Asokan, N., "Untraceability in Mobile Networks," in Proceedings of MobiCom 95, Berkeley, California, USA, 1995.

[Sandhu and Samarati, 1994] Sandhu, R. S. and Samarati, P., "Access Control: Principles and Practice," *IEEE Communications Magazine*, vol. 27, no 9, pp. 40-48, 1994.

[Sandhu and Samarati, 1996] Sandhu, R. and Samarati, P., "Authentication, Access Control, and Audit," *ACM Computing Surveys*, vol. 28, no 1, pp. 241-243, 1996.

[Schneier, 1996] Schneier, B., *Applied Cryptography*, 2nd ed. New York: John Wiley & Sons, 1996.

[Schneier, 1998] Schneier, B., "Cryptographic Design Vulnerabilities," *IEEE Computer*, no 9, pp. 29-33, 1998.

[Schneier, 2000] Schneier, B., "Cryptogram," http://www.counterpane.com/crypto-gram.html, Counterpane Internet Security Inc., 2000.

[Schneier and Schostack, 1999] Schneier, B. and Schostack, A., "Breaking up is Hard to do: Modelling Security Threats for Smart Cards," in Proceedings of First USENIX Symposium on Smart Cards, 1999.

[Schwartau, 1999] Schwartau, W., "Surviving Denial of Service," *Computers and Security*, vol. 18, no 2, pp. 124-133, 1999.

[Simon, 1996] Simon, D. R., "Anonymous Communication and Anonymous Cash," in Proceedings of Information Hiding, Cambridge, UK, 1996.

[Sirbu and Tygar, 1995] Sirbu, M. and Tygar, J. D., "NetBill: An Internet Commerce System Optimized for Network-Delivered Services.," *IEEE Personal Communications*, vol. 2, no 3, pp. 34-39, 1995.

[Socolofsky and Kale, 1991] Socolofsky, T. and Kale, C., "RFC 1180 - A TCP/IP Tutotrial," IETF, January 1991.

[Stallings, 1996] Stallings, W., "IPv6: The New Internet Protocol," *IEEE Communications Magazine*, vol. 29, no 7, pp. 96-108, 1996.

[Sumari, 2000] Sumari, P., "A Study of Storage Architecutres for Video on Demand," PhD Thesis, School of Computing and Mathematical Sciences, Liverpool John Moores University, Liverpool, 2000.

[Syverson, et al., 1997] Syverson, P. F., Stubblebine, S. G., and Goldschlag, D. M., "Unlinkable Serial Transactions," in Proceedings of Financial Cryptography FC '97, Anguilla, British West Indies, 1997.

[Szabo, 1995] Szabo, N., "Privacy Marketing," in *TBTF*, 1995, available online: http://www.tbtf.com/resource/priv-marketing.html.

[Tannenbaum, 1996] Tannenbaum, A. S., *Computer Networks*, 3rd ed: Prentice Hall, 1996.

[Tari and Chan, 1998] Tari, Z. and Chan, S.-W., "A Role-Based Access Control for Intranet Security," *IEEE Internet Computing*, no 5, pp. 24-34, 1998.

[UK-Crypto, 1999] UK-Crypto Internet Mailing List, 1999.

[UN, 1948] "United Nations - Universal Declaration of Human Rights," United Nations, available online: http://www.un.org/Overview/rights.html, 1948.

[Varadharajan and Mu, 1996] Varadharajan, V. and Mu, Y., "Design of Secure End-to-End Protocols for Mobile Systems," in Proceedings of Mobile Communications, Canberra, Australia, 1996.

[ViaCode, 1999] ViaCode, "ViaCode - Enabling Secure Electronic Commerce," Royal Mail, 1999.

[Voydock and Kent, 1983] Voydock, V. L. and Kent, S. T., "Security Mechanisms in High-Level Network Protocols," *ACM Computing Surveys*, vol. 15, no 2, pp. 135-171, 1983.

[Wagner, et al., 1997] Wagner, D., Schneier, B., and Kelsey, J., "Cryptanalysis of the Cellular Encryption Algorithm," in Proceedings of CRYPTO '97, Santa Barbra, CA, USA, 1997.

[Wang, et al., 1998] Wang, H., Lee, M. K. O., and Wang, C., "Consumer Privacy Concerns about Internet Marketing," *Communications of the ACM*, vol. 41, no 3, pp. 63-70, 1998.

[WAPForum, 2000] WAP Forum Web Site, http://www.wapforum.org/

[Warren and Brandeis, 1890] Warren, S. and Brandeis, L. D., "The Right to Privacy," *Harvard Law Review*, vol. 4, no 193, available online: http://www.louisville.edu/library/law/brandeis/privacy.html, 1890.

[Whitten and Tygar, 1999] Whitten, A. and Tygar, J. D., "Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0," in Proceedings of Eighth USENIX Security Symposium (Security '99), Washington, D.C., USA, 1999.

[Young, 1998] Young, J., "Crack A5," JYA News Archive, available online: http://jya.com/crack-a5.htm, 18th April 1998.

[Zheng, 1996] Zheng, Y., "An Authentication and Security Protocol for Mobile Computing," in Proceedings of Mobile Communications, Canberra, Australia, 1996.