NEAR EAST UNIVERSITY

Faculty of Engineering

Department of Computer Engineering

FIREWALLS AND TRUSTED SYSTEMS

Graduation Project

Com- 400

Student: Cem Baki (971480)

Supervisor: Prof. Dr Fakhraddin Mamedov

Nicosia 2003

ACKNWOLEDGEMENT

At the beginning, I would like to thank my supervisor Prof. Dr. Fakhreddin Mamedov for his guidance to overcome many difficulties require in this project.

Throughout the years of education, my friends and specially my family has given me an endless support and encouragement. I would like to thank all of them for everything.

Also, I wish to express my gratitude and appreciations to everybody who have contributed their efforts to make this project a successful one.

Finally, thank the committee members that are going to listen me.

ABSTRACT

The Internet and the highly growing network of computers in our everyday lives are making us more dependent on them. Such that, production of an ordinary product, traffic light systems, satellite controllers and even check-ups in hospitals are now mostly carried out by computers if not completely.

This increasing demand on computers and related technology brings along a huge problem with it – SECURITY. As computers play more and more important data between them such as financial, personal info and any other information regarded as 'confidential', the need for security arises along with the need to 'break into' them. All our information is under the threat of hackers or information thieves in other words, either for re-sale to third parties and/or for capturing sensitive data (e.g. credit card numbers) that would posses both personal and financial data for use of abuse.

This project discusses the ways of securing data and data transfer between computerized systems in order to protect the said information. The technology so called 'Firewall' and 'Trusted Systems' are currently the most popular way of securing data and is explained in detail within this report.

TABLE OF CONTENT

INTRODUCTION

Firewalls are currently the major way of securing computer and computerized systems. Educational institutions, small and large corporations, government organizations and many more organizations are using them as to secure their network of computers to protect themselves from threats such as hackers, viruses etc.

They are also used to limit the way one can use a computer such as preventing porn and illegal web sites to be viewed, to control and block unwanted (spam) emails etc.

Firewalls however may not be sufficient enough to completely secure a single or a network of computers. The way they work are explained in this project in detail including TCP/IP and other protocols used by networks and operators. According to this, chapters included, as also mentioned under the topic 'OVERVIEW' are;

Chapter 1: presents an Introduction to Internet Security, and presents firewalls as the primary means by which organizations can manage the risks associated with connecting their network to the Internet.
Chapter 2 provides a brief introduction to the TCP/IP Protocols. The TCP/IP five layer model is described and used as a framework to discuss security weaknesses in the protocols.
Chapter 3 describes generic Computer Security Risks and specific Internet Attacks.
Chapter 4 looks at the nature and role of a Network Security Policy.
Chapter 5 provides an introduction to Firewalls, and presents Firewall Theory & Architectures.
Chapter 6 presents two Case Studies. The first investigates the decision making process when selecting a firewall. The second discusses the management issues that are raised when one is installed.
Chapter 7 discusses Future Developments.
Chapter 8 presents the Summary and Conclusions

Firewalls and Internet security - Overview

On December 18, 1995 the Computer Emergency Response Team (1) (CERT) issued an advisory [CA-95:18] entitled "Widespread Attacks on Internet Sites". The advisory stated :

"Over the last several weeks, the CERT Co-ordination Centre has been working on a set of incidents in which the intruders have launched widespread attacks against Internet sites. Hundreds of sites have been attacked, and many of the attacks have been successful, resulting in root compromises at the targeted sites. We continue to receive reports, and we believe that more attacks are going undetected."

CERT handled a total of 2,412 computer security incidents during 1995 [CERT95]. More than 12,000 sites were affected by these incidents, which involved 732 break-ins and a similar number of probes and pranks. CERT reported that the most serious intruder activities during 1995 included:-

- IP spoofing ;
- Network File Service (NFS) attacks ;
- network scanning ;
- packet sniffers, and
- Send mail (2) attacks.

It is against this backdrop that the debate on Internet Security rages on. Some organizations believe that the Internet is too unruly to be used for business. Others believe that the Internet has too much potential for them to be dissuaded from using it by security breaches, preferring instead to seek ways to minimize their exposure to attacks and to manage the security issue.

However it becomes increasingly difficult to maintain an adequate level of security as the number of hosts on a network increases. This is because host based security does not scale well [Wack95]. Internet Firewalls avoid this problem because they are generally installed between an organisation's network and the Internet, thus providing a central point at which security measures can be concentrated. Internet firewalls maintain a level of segregation

between an organisation's network and the Internet that is conducive to good security whilst permitting the requisite level of connectivity.

This report discusses what is meant by Internet Security, and presents firewalls as the primary means by which organisations can manage the risks associated with connecting their network to the Internet.

Chapter 1 presents an Introduction to Internet Security and discusses the four constituents of it, Authentication, Access Control, Integrity and Confidentiality.

Chapter 2 provides a brief introduction to the TCP/IP Protocols. The TCP/IP five layer model is described and used as a framework to discuss security weaknesses in the protocols.

Chapter 3 describes generic Computer Security Risks and specific Internet Attacks.

Chapter 4 looks at the nature and role of a Network Security Policy.

Chapter 5 provides an introduction to Firewalls, and presents Firewall Theory & Architectures.

Chapter 6 presents two Case Studies. The first investigates the decision making process when selecting a firewall. The second discusses the management issues that are raised when one is installed.

Chapter 7 discusses Future Developments.

Chapter 8 presents the Summary and Conclusions

(1) The CERT Co-ordination Centre was formed by the Advanced Research Projects Agency (ARPA) in November 1988 in response to the need for central security co-ordination demonstrated by the Internet Worm Virus. CERT's charter is to work with the Internet community to detect and resolve computer security incidents and to take steps to prevent future incidents

(2) Sendmail is the program that UNIX systems use to handle electronic mail.

CHAPTER 1

Firewalls and Internet security - Introduction to Internet Security

Any one responsible for the security of a trusted network will be concerned when connecting it to an untrusted network. In the case of connections to the Internet this concern may be based largely on anecdotal evidence gleaned from widespread media coverage of security breaches. A closer inspection of the facts and statistics behind some of the media coverage will, however, only serve to deepen that concern. For example, the US National Computer Security Agency (NCSA) asserts that most attacks to computer systems go undetected and unreported, citing attacks made against 9000 Department of Defence computers by the US Defence Information Systems Agency (DISA). These attacks had an 88 per cent success rate and went undetected by more than 95 per cent of the target organisations. Only 5 percent of the 5 per cent that detected an attack, a mere 22 sites, reacted to it [Cobb95].

It is noteworthy that these sites belong to the US Department of Defence (DoD) and were not commercial sites, which may give security less priority than the DoD.

NCSA also quote the FBI as reporting that in more than 80 percent of FBI investigated computer crimes, unauthorised access was gained through the Internet [Cobb95].

Putting a value on the damage done by such attacks is difficult but a 1995 survey conducted by Ernst & Young, a New York based accounting firm, reported that one third of businesses connected to the Internet reported up to 100 000 USD in financial loss over a two year period due to malicious acts by computer users outside the firm. A little more than two percent of connected companies reported losses of more than 1M USD [McGa95].

There is amazement in the computer security industry at the level of ignorance to the problem. To understand the risks often involves a steep learning curve and they have few real parallels in everyday life, for example nobody worries that a burglar will be able to trick their front door into opening by posting cryptic messages through the letterbox. When there is a good "hacker" story to report the press goes into frenzy, but the general level of awareness is still surprisingly low. For example the Sunday Times which prides itself on providing

accurate coverage of IT issues published an article recently that claimed that most businesses worry too much about Internet security. The article goes on to explain that encryption is all that is needed to be completely secure. The article focuses purely on privacy of communication and completely misses the possibility of an attack originating from the Internet [Bray96].

Despite fears about security, organisations are increasingly coming to regard a presence on the Internet as an important part of their strategic planning. Security concerns will not be allowed to prevent organisations from exploiting the commercial opportunities the Internet is perceived to offer. As a result organisations have to find ways to manage the security issue. This ties growth in the Internet security market directly to growth in the Internet. The compound annual growth rate (CAGR) of the Internet firewall market between 1995 and 2000 is projected to be 174% [IDC96] driven by rapid growth of both the Internet (see table 1), and Intranets [Nadi96]. The most significant trend driving this growth is the rapid and aggressive deployment of World Wide Web servers for both Internet and Intranet use. Unit shipments of web server software are expected to grow from 127 000 units in 1995 to just more than 5 million units in 2000 [IDC96]. Although the IT industry has traditionally enjoyed rapid development this level of growth is unprecedented.

It is difficult to separate figures for the European or UK firewall markets from the world wide statistics quoted in the literature. 1996 may see similar levels of activity in Europe and the UK to those seen in the USA in 1995(1). A 1995 survey of government agencies and fortune 500 companies conducted by the Computer Security Institute [CSI95b] found that while 78% of respondents used the Internet, 39% did not have a firewall. Similarly 40% of the audience at a February 1996 NSCA conference devoted to firewalls and Internet security did not have a firewall [Book96].

	Hosts	**Domains**	**Network Class**		
	(000s)	(000s)	A	B	C
Jan 93	1313	21	54	3206	4998
Apr 93	1486	22	58	3409	6255
Jul 93	1776	26	67	3728	9972
Oct 93	2056	28	69	3849	12615

Jan 94	2217	30	74	4043	16422
Jul 94	3212	46	89	4493	20628
Oct 94	3864	56	93	4831	32098
Jan 95	4852	71	91	4979	34340
Jul 95	6642	120	91	5390	56057
Jan 96	9472	240	92	5655	87924

Table 1 : Growth of the Internet

Source : Network Wizards Internet Domain Survey, January 1996, available from HTTP://www.nw.com/

Given that approximately 40% of the fortune 500 companies using the Internet have still to install a firewall and that the Internet continues to double annually, it is little surprise that the security auditing business is booming [Book96]. Organisations are finding that they do not have the in-house skills or knowledge necessary to assess either the current situation or the potential risks, and are wrestling with what level of security they require. The rest of this chapter investigates what is meant by the term Internet security - often the starting point when an organisation calls in an external consultant [Hews96].

1.1. What is Internet Security?

The hardware, software and information that constitute computer systems is increasingly mission-critical. Protecting them can be as important as protecting other valuable resources, such as money, buildings, or employees. The purpose of computer security is to protect computer resources through the selection and application of appropriate safeguards.

Internet security protects computer resources against the risks and threats that arise as a result of a connection to the Internet.

Computer security supports the organisation's mission by protecting its physical and financial resources, reputation, legal position, employees, and other tangible and intangible assets.

A networked computer system may not be constrained to a single organisation(2). In an inter-organisational system, computer security benefits each organisation. Electronic commerce by definition is inter-organisational and effective security is essential to its success(3). Security on the buyer's system also benefits the seller as the buyer's system is less likely to be used for fraud, or otherwise negatively affect the seller, and vice-versa.

If a system has external users then its owners have a responsibility [NIST95] to share appropriate knowledge about the existence and general extent of security measures so that other users can be confident that the system is adequately secure. In addition to sharing information about security, organisation managers "should act in a timely, co-ordinated manner to prevent and to respond to breaches of security" [NIST95] to help prevent damage to others.

Computers and the environments they operate in are extremely dynamic. Changes in the system or the environment can create new vulnerabilities and it is almost inevitable that a system's users and operators will discover new ways to intentionally or unintentionally bypass or subvert security. It is therefore necessary to reassess the security of computer systems regularly to provide effective computer security.

Providing effective computer security requires a comprehensive approach that considers a variety of areas both within and outside of the computer security field and that extends throughout the entire information life cycle.

There are three general areas of concern when a trusted network is attached to an untrusted network:-

1. that inappropriate material will deliberately, or inadvertently, be passed to and from the untrusted network;
2. that unauthorised users will be able to gain access to the trusted network from the untrusted network;
3. that the operations of the trusted network may be disrupted as a result of attack from the untrusted network.

The computer and network security measures that are taken by an organisation are intended to minimise the potential for these to occur by means of the four fundamental components that make up computer network security :-

1. Identification and Authentication
2. Access Control
3. Integrity
4. Confidentiality

1.2. Identification and Authentication

The first component of computer security is authentication, or ensuring that users and computers are who they claim to be by establishing proof of identity. This is usually accomplished based on one, or a combination, of something you are (a biometric e.g., such characteristics as a voice pattern, handwriting or a fingerprint), something you know (a secret e.g., a password, Personal Identification Number (PIN), or cryptographic key) or something you have (a token e.g., an credit card or a smart card).

For example acquaintances can authenticate your identity (to a point) based on your physical features(4). Banks authenticate you based on something you have such as your credit card, and something you know, often your mother's maiden name. One-time passwords, or passwords that can only be used once then expire, are generally based on something you have. An example of this type of authentication are the one-time pads(5) used by the intelligence services during the second world war.

The lack of strong authentication has inhibited the development of electronic commerce. It is still necessary for contracts, legal documents and official letters to be produced on paper. Strong authentication is then, a key requirement if the Internet is to be used for electronic commerce [Ranu95e]. Strong authentication is generally based on modern equivalents of the one time pad. For example tokens are used in place of one-time pads and are stored on smart cards or disks, or in some cases the authenticating computer will generate a challenge which the user enters into a small device similar to a calculator to generate the correct response.

Authentication is an important part of everyday life. Letters are printed on headed paper and signed by the author. Digital signatures fulfil a similar requirement, although they are much more trustworthy as they are based on mathematical encryption algorithms and attest to the contents of a message as well as its author. Digital signatures are based on public key, or asymmetric encryption. The concept of public key encryption was introduced in 1976 by Whitfield Diffie and Martin Hellman [Diff76] in order to solve the key management problem that exists with secret key or symmetric encryption(6). Asymmetric cryptography uses key pairs, one key in the key pair is called the public key and the other is called the private key. Either key can be used to encrypt the message, but once encrypted only the other key in the pair can be used to decrypt it. It is immediately apparent that two scenarios are possible, one where the private key is used to encrypt the message and hence the public key is used to decrypt it, and vice-versa. By encrypting the message using the receiver's public key, the sender is assured that only the receiver can decrypt it confidentially. To digitally sign a message the sender passes the message through a hashing algorithm(7) to produce the message digest which he then encrypts with his private key. The output is called a digital signature and is attached to, and sent with, the message. In order to verify the signature the receiver also passes the message through the same hashing algorithm to re-create the message digest, and then decrypts the sender's digital signature using the sender's public key. If the message did not originate from the sender, or if its contents were altered, then the two digests will not match.

Under normal circumstances the private key is kept secret by the individual, but the public key is distributed as required. There is no need for the sender and receiver to share a secret key, however, asymmetric encryption key management still requires public keys to be distributed in an authenticated or trustworthy manner.

One means of achieving this is to use a certification authority. The main attribute of a certification authority is that it is trusted by a group of users to create certificates on their behalf [Chad94]. The certification authority verifies a user's public key by digitally signing it. This creates a certified public key, referred to as a certificate(8). The certification authority's digital signature attests that the public key is valid, and guarantees that it cannot be altered in any way.

One such certification authority is VeriSign Incorporated who began issuing key pairs and certificates in late April 1996 [Clar96] and have trademarked the term "Digital ID".

Security aware applications are required to make use of certificates and secure e-mail tools and browsers for the World Wide Web are now becoming available. When the certificate details have been installed in the client software (i.e. browser) they are automatically provided along with the client's requests allowing the server to authenticate you.

Identification and Authentication is a critical building block of computer security since it is the basis for most types of access control and for establishing user accountability.

1.3. Access Control

Access is the ability to do something with a computer resource (e.g., use, change, or view). Access control is the means by which the ability is explicitly enabled or restricted in some way (usually through physical and system-based controls).

Access control often requires that the system be able to identify and differentiate users. For example, access control is often based on least privilege, which refers to the granting to users of only those accesses required to perform their duties. User accountability requires the linking of activities on a computer system to specific individuals and, therefore, requires the system to identify users.

Access controls provide a technical means of controlling what information users can utilise, the programs they can run, and the modifications they can make.

Computer-based access controls are called logical access controls [NIST95]. Logical access controls can prescribe not only who or what (e.g., in the case of a process) is to have access to a specific system resource but also the type of access that is permitted. These controls may be built into the operating system, may be incorporated into applications programs or major utilities (e.g., database management systems or communications systems), or may be implemented through add-on security packages. Logical access controls may be implemented internally to the computer system being protected or may be implemented in external devices.

Logical access controls can help to protect :-

- operating systems and other system software from unauthorised modification or manipulation (and thereby help ensure the system's integrity and availability)
- the integrity and availability of information by restricting the number of users and processes with access
- confidential information from being disclosed to unauthorised individuals.

The concept of access modes is fundamental to access control. Common access modes, which can be used in both operating systems and applications, include the following :-

· Read

- Write
- Delete
- Create
- Execute

In deciding whether to permit someone to use a system resource logical access controls examine whether the user is authorised for the type of access requested based on access criteria such as :-

- Identity
- Roles
- Location

· Time

- Transaction
- Service Constraints

Identity - It is probably fair to say that the majority of access controls are based upon the identity of the user (either human or process), which is usually established through identification and authentication.

Roles - Access to information may also be controlled by the job assignment or function (i.e., the role) of the user who is seeking access. Examples of roles include data entry clerk, purchase officer, project leader and programmer. Access rights are grouped by role name, and the use of resources is restricted to individuals authorised to assume the associated role. An individual may be authorised for more than one role, but may be required to act in only a single role at a time. Changing roles may require logging out and then in again, or entering a role-changing command. The use of roles can be a very effective means of providing access control.

Location - Access to particular system resources may also be based upon physical or logical location for example, users can be restricted based upon network addresses (e.g., users from sites within a given organisation may be permitted greater access than those from outside).

Time - Time-of-day or day-of-week restrictions are common limitations on access. For example, use of confidential personnel files may be allowed only during normal working hours and denied at all other times.

Transaction - Another approach to access control can be used by organisations handling transactions (e.g., account inquiries). Phone calls may first be answered by a computer that requests that callers key in their account number and perhaps a PIN. Some routine transactions can then be made directly, but more complex ones may require human intervention. In such cases, the computer, which already knows the account number, can grant a clerk, for example, access to a particular account for the duration of the transaction. When completed, the access authorisation is terminated. This means that users have no choice in which accounts they have access to, which can reduce the potential for mischief. It also prevents users from casually browsing through accounts thereby improving confidentiality.

Service Constraints - Service constraints refer to those restrictions that depend upon the parameters that may arise during use of the application or that are pre-established by the resource owner/manager. For example, a particular software package may only be licensed by the organisation for five users at a time. Access would be denied for a sixth user, even if the user were otherwise authorised to use the application. Access may also be selectively permitted based on the type of service requested. For example, users of computers on a

network may be permitted to exchange electronic mail but may not be allowed to log in to each others' computers.

External Access Controls - External access controls are a means of controlling interactions between the system and outside people, systems, and services. External access controls use a wide variety of methods, often including firewalls as will be discussed in later chapters.

1.4. Integrity

Integrity is the degree to which something is free from corruption, i.e. whether or not something has been damaged, altered, added or removed. In addition to improving authentication, digital signatures also improve the level of confidence in the integrity of a message as discussed earlier in this chapter.

Integrity does not apply only to messages however. The integrity of files and applications is also very important. One of the most common means of gaining unauthorised access to a computer system is to install altered copies of operating system programs that provide access to the intruder when they are executed(9). It is important therefore that the integrity of operating system components can be verified. Attackers themselves understand this well, as is illustrated by [Shim95] which describes how an attacker who, whilst being monitoring began, immediately upon discovering that one of his back door programs had been removed, to compare copies of other files he had replaced with the originals that he had stored elsewhere.

The integrity of anti-virus software should also be verified regularly. Most packages on the market perform a self-verification of their integrity. The problem with this is that rogue software would presumably not be designed to point out that it differed from the original. In cases such as this verification of integrity should be independent in order to be trustworthy.

In some cases the integrity of data files is also often assumed to be verified by the application software. Whilst the application software will generally notify the user of damage or corruption to the file it will not generally report that Company A has been removed from a list of companies tendering for a major contract for example. Again integrity needs to be verified independently.

Both message digests and digital signatures can attest to the integrity of files in all of these cases. The point is that in order to be trusted independent verification is required.

1.5. Confidentiality

Confidentiality is the degree to which the privacy or secrecy of something can be trusted. The confidentiality of most paper based communication is entrusted to envelopes. Most messages transmitted over the Internet cannot claim even this level of confidentiality, being more akin to postcards. The lack of privacy (or confidentiality) on the Internet applies equally to files transferred over it, and information moving to and from World Wide Web clients and servers.

E-mail, File Transfer and World Wide Web applications accounted for approximately half of the bytes transferred on the Internet backbone in 1994(10) [MERI94]. Regardless of what the data was, the vast majority of this traffic was transmitted without any regard for its confidentiality.

Initiatives to correct this state of affairs have been underway for some time and are likely to come to fruition in 1996, for example, Web Browsers that are able to use certificates and therefore make use of the Privacy Enhanced Mail (PEM) standard and Secure Multipurpose Internet Mail Extensions (S/MIME) standards. These will be discussed in more detail in chapter 2.

The first step in protecting a computer or network of computers is to establish a security policy that addresses each of the components of computer security that have been described above. In the case of computer networks such a policy is generally referred to as a Network Security Policy.

1.6. Network Security Policy

The Network Security Policy identifies the threats against which protection is required, and defines the required level of protection. The Network Security Policy will itself contain several different policies, for example a Network Service Access Policy and System Specific Policies.

The Network Security Policy will be based on a security strategy such as Least Privilege, Defence In Depth, Choke Point, Weakest Link ,Fail Safe Stance etc. These and other strategies are discussed in chapter 4. The role of the security strategy can be illustrated with a small example :

Strategy 1 : Everything is forbidden unless explicitly permitted.

Strategy 2 : Everything is permitted unless explicitly forbidden.(11)

Implementations of both of these strategies can be found in organisations. They adopt philosophically opposing views of how to implement security.

Some understanding of the services available on the Internet, and the risks these present, is required before an effective network security policy can be developed.

The next chapter introduces the Internet protocols and services. Chapter 3 then introduces computer security risks and attacks, and chapter 4 addresses netwrok security policy. Once a security strategy and policy have been decided a means of implementing them is required. The generic term "Firewall" is increasingly being used to describe the combination of hardware, software and management activities that are used to effect the network security policy. The theory and architecture of firewalls is presented in chapter 5.

(1) There are some barriers to this however. As the Internet facilitates the trend towards increasing globalisation, issues such as export restrictions of , for example, cryptography technology, are presenting interesting problems for governments on both sides of the Atlantic. The size of the North American market tends to generate a critical mass for de-facto standards that are often based on technology that is subject to export restrictions in the USA. European governments are concerned about issues of national security that would arise from their reliance on foreign owned and developed security technology.

(2) [Ches94] identified joint ventures and mergers as posing particular problems in terms of computer security as security itself is generally constrained to a single organisation.

(3) In the same way that preventing the forgery of bank notes is essential to the success of commerce based on paper money.

(4) A slightly macabre example is when friends or family have to identify (i.e. authenticate) a corpse.

(5) A one-time pad, sometimes called the Vernam cipher [Vern26], is said to offer perfect secrecy as it is based on an entirely random string of bits that is the same length as the plaintext message. The plaintext message and the string of random bits are combined using a bitwise exclusive-or operation to produce the ciphertext. Because the string of bits is entirely random, an opponent with infinite computational resources can only guess the plaintext if he sees the ciphertext.

Key management issues render the one-time pad impractical as the secret key, since it can only be used once, and it is as long as the message itself. The one-time pad did see use in the second world war however, over diplomatic channels that required exceptionally high security.

Analysis of the one-time pad is one of the cornerstones of modern cryptography [Shan49].

For more information about one time pads (including a picture of one captured from the Russians by MI5 see [Ranu95b].

(6) Symmetric encryption is based on both the sender and receiver of the encrypted message knowing and using the same secret key. Unless the two parties are together and alone the possibility exists that whilst they try to agree upon a secret key a third party will discover it. The generation, storage and transmission of keys is called key management and is something that affects all cryptography systems. Because all keys in symmetric cryptography must remain secret the key management problem is particularly difficult.

(7) A hash function H is a mathematical transformation that takes the variable sized input m and returns a fixed-size string, which is called the hash value h (i.e. h = H(m)) Examples of well known hash functions are MD2 (Message Digest 2), MD4, MD5, SHA (Secure Hash Algorithm). As hash functions are generally faster than asymmetric encryption algorithms, the digital signature of a document is typically computed by computing the digital signature of its hash value, which is small (128 bits for MD5) compared to the document itself. It is not feasible for anyone to either find a message with a given hash value or to find two messages that have the same value as there are 2n possibilities, where n is the number of bits in the hash

value (128 bits for MD5). If either were possible it would be possible to attach a false message to a sender's signature.

Because hash functions are one-way functions, i.e. the function cannot be reversed, a document's hash value (also called its digest) can be made public without revealing the contents of the document itself. This is important in digital timestamping where, by using hash functions, one can have a document timestamped without revealing its contents to the timestamping service.

(8) The most widely accepted format for certificates is defined by the CCITT X.509 international standard [CCIT88]. Further refinements are found in the PKCS set of standards and the PEM standard (RFCs 1421-1424).

(9) Such a means of access to a computer system is often referred to as a "back door" to the system

(10) NSFNET performance statistics have been collected, processed, stored, and reported by the Merit Network since 1988, in the early stages of the NSFNET project. During December 1994, the numbers contained in Merit's statistical reports began to decrease, as NSFNET traffic began to migrate to the new NSF network architecture. In the new architecture, traffic is exchanged at interconnection points called NAPs (Network Access Points.) Each NAP provides a neutral interconnection point for U.S.-based and international network service providers. On April 30, 1995, the NSFNET Backbone Service successfully made the transition to the new network architecture. Although the reports are inclusive through to the end of the NSFNET service, the November 1994 reports were the last to reflect the nature of the NSFNET backbone traffic in its entirety.

(11) [Wack95] argues that strategy 1 is much harder to implement than strategy 2. Whilst this is true of routers and packet filters it is not necessarily true of application gateways. Furthermore, dual-homed hosts intrinsically deny everything unless it is permitted.

CHAPTER 2

Firewalls and Internet security - The TCP/IP Protocols

TCP/IP is a motley collection of standards written by professionals and amateurs alike. In the preface to "IPng and the TCP/IP Protocols" [Thom95] Stephen Thomas writes :-

"There are several characteristics of TCP/IP that have contributed to its immense success, perhaps none more important than its ruthless eradication of the impractical. The TCP/IP community discourages elaborate formal frameworks, unworkable schemes and senseless strategies. Instead TCP/IP's designers - a mix of academics, engineers, network administrators, and users - concentrate on solving real problems with real applications."

A useful place to begin looking at TCP/IP is to review its history.

2.1. History of TCP/IP

In 1969 the US Defence Advanced Research Projects Agency (DARPA) began funding a project to develop a high speed, packet switching communications network to link its research centres and laboratories. The system became known as ARPANET and was one of the first communications systems to utilise a layered architecture, preceding the ISO OSI reference model by almost a decade. Although the ARPANET was a general success, its first generation protocols were expensive to implement, slow and prone to crashes (both of the individual stacks and of the networks themselves). In 1974 a new set of core protocols was proposed by Vinton Cerf and Robert Kahn [Cerf74]. This proposal was the basis for the development of the Internet Protocol (IP) and the Transmission Control Protocols (TCP). Over a period of three years the ARPANET host systems migrated to use these protocols. TCP/IP was better suited to inter-networking than the Xerox Networking System (XNS) protocol stack, the other major protocol stack available at that time, for two reasons. Firstly it utilised a defined routing hierarchy that allowed large inter-networks to be managed in a structured way, and secondly, its addresses were centrally administered so duplicates could only be the result of error. DARPA funded the integration of the TCP/IP protocols into the University of California's Berkeley Software Distribution (BSD) version of UNIX. Version 4.2 of the BSD UNIX released in September 1983 was the first to include the TCP/IP

protocols in the generic operating system, and this was eventually carried over into commercial versions of UNIX. SUN Microsystems later published their Open Network Computing (ONC) Standards, better known as the Network Filing System (NFS). NFS is designed to utilise a TCP/IP stack and has since been widely licensed.

TCP/IP has two major shortcomings at present. The first is that address space is limited and will eventually run out. The second is that there are a number of security weaknesses. To explore the latter it is necessary to understand the layered architecture of TCP/IP.

2.1.1. TCP/IP Architecture

TCP/IP has a layered architecture characterised by increasing abstraction as we move up the layers. Entities within each layer provide services to those in the layer above it, and request services from those in the layer below it. The TCP/IP model consists of five layers(1) :-

1. The Application Layer provides the application program or process. Examples of Internet application layer protocols are File Transfer Protocol (FTP), Simple Mail Transfer Protocol (SMTP), Network News Transfer Protocol (NNTP), and Hypertext Transfer Protocol (HTTP). An application engages network services from the TCP or UDP transport layers through one of several APIs such as Berkeley Sockets or the Transport Layer Interface.
2. The Transport Layer provides two types of service to the Application Layer. The first is a connection-oriented, full duplex service provided by the Transmission Control Protocol (TCP). The second is a connectionless service provided by the User Datagram Protocol (UDP).
3. The Network Layer is responsible for moving data between communicating endpoints. Unless the sending and receiving hosts are on the same network this job involves routing - determining and delivering the data along the best inter-network path. The primary protocol in use at the network layer is the connectionless Internet Protocol (IP). The basic unit of communication at this layer is called the IP datagram (sometimes referred to as an IP Packet).

4. The Data Link Layer controls the link between 2 nodes in a network by controlling the flow of data to and from the physical layer, the detection and correction of errors in the data, and sequencing.
5. The Physical Layer is the lowest layer of the model. This layer sends and receives a stream of bits across the physical medium that connects the systems.

A message is transferred across the network by passing down the layers on the sending host and back up them on the receiving host. The message is packaged in a new envelope by each layer before it passes it down (see figure 1). After transmission across the network the message is passed up the layers with each layer removing its respective envelope. The original message is finally passed to the application layer on the receiving host.

Armed with an understanding of the layered nature of Internet communications we are now equipped to explore some of the security issues intrinsic to each of the layers.

2.1.2. Security Weaknesses in TCP/IP in context of TCP/IP Model

There are security weaknesses at each layer of the model and an attacker could exploit any one of them. The analysis presented here is not intended to be exhaustive. New weaknesses are discovered regularly and the reader should register with one of the Internet security advisory services such as CERT to stay fully abreast of current weaknesses.

Figure 1. Protocol Enveloping and the TCP/IP 5 Layer Model

2.2. Physical and Data Link Layers

Security at the Physical and Data Link layers is primarily concerned with access control to, and confidentiality of, the physical transmission medium.

2.2.1. Network Layer

Security issues at the Network Layer are related to end-to-end delivery of a datagram or IP Packet. The issues include Network Snooping, Message Replay, Message Alteration, Message Delay and Message Denial.

Network Snooping - An earlier analogy likened messages travelling over the Internet to messages written on a post card. Just as everyone who handles a postcard can choose whether or not to read the message, so any system on a network with a shared transmission medium can read every network datagram, whether addressed to it or not. Snooping or Sniffing is a passive attack, i.e. the attacker observes network traffic but does not disturb it. Sniffer software is readily available [Tard95], and is used legitimately as a network troubleshooting tool. However sniffer software is also widely used by attackers to collect account names and passwords. [Cert95] reported that in some cases a packet sniffer was found to have been running for months(2). Sniffing works by placing a system's network interface software into promiscuous mode. All packets are then passed to the sniffer software which can then display or record the information.

Message Replay - An attacker can sometimes replay traffic recorded with sniffer software to attack a computer.

Message Alteration - The importance of Integrity was discussed in Chapter 1. There is currently no widely implemented means of guaranteeing the integrity of an IP datagram. An attacker who modifies the contents of a datagram can also recalculate and update its header checksum, and the datagram recipient will be unable to detect the change.

Message Delay and Denial - An attacker can delay or deny IP messages by changing the screening and routing rules used by routers, or by overwhelming one of the end systems with large amounts of network traffic. Routers and hosts have to discard incoming packets if they have no remaining buffer space. If the packets contain UDP datagrams they are lost forever. If the packets contain TCP segments the reliable transport mechanism eventually recovers by requesting retransmission of the lost data.

Authentication at the network layer is concerned with identifying computer systems rather than computer users. The identity of a computer system on the Internet is its IP address.

There is no widespread means of authenticating an IP address and this results in attacks such as Address Masquerading and Address Spoofing.

Address Masquerading occurs when an attacker configures his network interface with the same address as another system. This can gain the attacker access to resources intended for the true owner of the IP address since access to some services, such as NFS, is only contingent upon the use of a "correct" network address(3). Address masquerading is limited to machines on the same network.

Address spoofing attacks a weakness of TCP but its net effect is at the network layer. It is a sophisticated attack on the three way handshake that is used to establish a reliable transport connection between two systems. An address spoofing attack exploits several weaknesses

- The trust relationship between two computers
- The predictability of TCP's Initial Sequence Number (ISN) which plays an important part in the ordering of all subsequent exchanges in the conversation.
- The weak (IP address based) authentication used by some commands.

The possibility of an address spoofing attack was first suggested by Robert Tappan Morris(4) in 1985 [Morr85] and expanded by Stephen Bellovin [Bell89] who noted that :-

"Some of these flaws [sequence number spoofing, routing attacks, source address spoofing, and authentication attacks] occur because hosts rely upon IP source address for authentication.....Others exist because network control mechanisms, and in particular routing protocols, have minimal or non-existent authentication"

The first Address Spoofing attack was reported by CERT on 23 January 1995 [CA-95:01](5) which stated :-

"The CERT Co-ordination Centre has received reports of attacks in which intruders create packets with spoofed source IP addresses. These attacks exploit applications that use authentication based on IP addresses. This exploitation leads to user and possibly root access on the targeted system"

Routing Attacks - Most routing protocols are susceptible to false route update messages as they do not use secure authentication mechanisms. IP also supports a source

routing option that allows an attacker to specify the routing path packets travel along to their destination.

Tunneling - IP can be encapsulated within TCP, a technique known as tunneling. Such tunnels can then be used to bypass the firewall. The extent of damage done depends upon how routing information is propagated [Ches94].

2.2.2. Transport layer

TCP segments and UDP datagrams are transmitted across the network as IP datagrams and so many of the security weaknesses already discussed apply, specifically the authenticity, integrity and confidentiality of TCP and UDP messages is not guaranteed. If a means of solving these problems at the network layer were to be found, then these issues in the Transport layer protocols would similarly be resolved.

There are however some weaknesses introduced in the transport layer itself that allow attacks such as UDP packet storm denial of service attacks [CA-96:01] and TCP Session Hijacking [CA-95:01].

UDP packet storm denial of service attacks. - [CA-96:01] reports of programs that launch denial-of-service attacks by creating a "UDP packet storm" either on a system or between two systems. An attack on one host causes that host to perform poorly. An attack between two hosts can cause extreme network congestion in addition to adversely affecting both hosts' performance. When a connection is established between two UDP services, each of which produces output, a very high number of packets can be produced. This can be used to effect a denial of service attack on the machine(s) providing the UDP services. Anyone with network connectivity can launch such an attack as no account access is needed.

Session Hijacking - Having gained root access to a system, an attacker can use a tool to dynamically modify the UNIX kernel. This modification allows the attacker to hijack existing terminal and login connections from any user on the system. An attacker can bypass one-time passwords and other strong authentication schemes by taking over the connection after the authentication is complete. An attacker can gain access to remote sites by hijacking the connection after the user has completed the authentication to the remote location.

2.2.3. Application Layer

Application layer security issues include all of the issues discussed so far. However even if a totally secure "pipe" between two systems could be provided, it would not assist in the authentication of a remote user, or in preventing attacks targeted on application layer protocols such as SMTP, FTP or HTTP. These are discussed in the next section.

2.3. Internet Services and Security

After receiving incoming data from IP, the transport protocol (TCP or UDP) passes it to the correct application process. Application processes (also called network services) are identified by 16 bit values called port numbers. TCP ports are defined by non-negative numbers in the range 0 to 65 635. "Well known" TCP port numbers(6) are the first 1 024 ports (0-1 023), which are managed and assigned by the Internet Assigned Numbers Authority (IANA). The first header word of each TCP segment or UDP packet contains the source port number identifying the application that sent the data and the destination port number identifying the application that is to receive the data.

2.3.1. Domain Name System (DNS)

DNS is a distributed database system used to match host names with IP addresses. A host normally requests the IP address of a given domain name by sending a UDP message to the DNS server which responds with the IP address or with information about another DNS server. [Ches94] discusses several weaknesses in DNS software, concluding that machines should use address based authentication which, although weak, is far better than name based authentication. [Ches94] also notes that DNS contains a wealth of information(7) about a site that can be useful to an attacker.

2.3.2. Simple Mail Transfer Protocol (SMTP)

[Ches94] notes that whilst SMPT is fairly innocuous(8), the most common implementation of it, the sendmail program, is a security nightmare. Sendmail violates the

principle of least privilege (discussed in chapter 3) as it runs with root privilege. [Ches94] notes that privileged programs should be as small and modular as possible. Sendmail consists of tens of thousands of lines of C code. The content of mail messages can also pose dangers. Automatic execution of messages, the ability to mail executable programs and the ability to mail postscript files are all dangers with MIME [Ches94].

2.3.3. Telnet

Telnet provides simple terminal access to a host computer. The user is normally authenticated based on user name and password. Both of these are transmitted in plain text over the network however, and is therefore susceptible to capture. [Ches94] recommends that strong authentication be used when establishing a telnet session.

Strong authentication does not protect the rest of the session from attackers however, who can hijack it or capture packets using techniques already discussed. Symmetric encryption of telnet sessions presents the problem of an untrusted computer being provided with the key. One solution is to use asymmetric encryption for key management of telnet sessions.

2.3.4. The Network Time Protocol (NTP)

NTP is used to synchronise a host's clock with reference clocks to within 10ms or less [Ches94]. This is extremely useful when analysing an attack through the log files of different computers involved. [Shim95] describes NTP as "an absolute, non-negotiable requirement" because of the importance of synchronised log files.

NTP itself can be the target of various attacks aimed at altering the hosts time [Ches94]. More recent versions of NTP support cryptographic authentication, but as this is provided on a hop-by-hop basis, a host upstream may still be attacked.

2.3.5. Finger and Whois

Two standard protocols, finger and whois, are commonly used to look up information about individuals [Ches94].

The finger protocol in particular presents security problems, not in its implementation, but in the information it provides. Farmer and Venema [Farm94] call finger "one of the most dangerous services, because it is so useful for investigating a potential target".

As firewalls do not generally allow users to login there is little point in providing the finger service. However the wider point is still noteworthy, i.e. that services that provide information about a network's users are of immense value to an attacker.

2.3.6. Remote Procedure Call Based Protocols

There are several remote procedure call protocols known as RPCs. The most popular, Sun RPC, was developed by Sun Microsystems [Sun88, Sun90] and is generally referred to as simply "RPC".

RPC is layered on top of TCP or UDP. However it is used as a general purpose transport protocol in the same way as TCP and UDP by a variety of application protocols such as Network File System (NFS) and Network Information Service (NIS).

NFS and NIS are vulnerable programs from a security point of view, for example an attacker with access to an NFS server can probably read any file on the system [Chap95].

When an RPC based service such as an NFS server starts it allocates itself a random TCP or UDP port (some use TCP, some use UDP, and some use both). It then contacts an RPC service called the portmapper. The portmapper, which is allocated to well known port 111, registers its RPC service number(9) and the particular port(s) it is using at present.

An RPC based client that wishes to contact a particular RPC based server on a host first contacts the host's portmapper server. The client passes the server the RPC service number of the server it wishes to access. If the server is running, the portmapper responds with the TCP or UDP port number it registered. The client is then able to communicate directly with the server. RPC servers are susceptible to many different attacks(10), for example :-

- RPC based services are vulnerable to denial of service attacks as the call to de-register a service is not well authenticated [Ches94].
- The portmapper does not authenticate requests to provide information about servers and is therefore a security risk.
- If access to the portmapper is blocked by a screening router, an attacker can bypass the portmapper and simply probe each of the TCP and UDP ports until he receives a response indicating an RPC based server is active.
- The most serious problem with the portmapper though, is its ability to issue indirect calls [Ches94]. To avoid the overhead associated with the additional round trip necessary to identify a server's port number, the remote client can ask the portmapper to forward the call to the server. However the server is not able to tell that the request did not originate locally, and is therefore unable to assess the level of trust that should be afforded to the call.

Network Information Service(11) - One of the most dangerous RPC applications according to [Ches94] is the Network Information Service (NIS). NIS is used to distribute a variety of important databases, including the password file, the host address table, and the public and private key databases used for secure RPC [Ches94]. The security implications of this should be obvious, [Ches94] comments:-

" If you are suitably cautious (read: 'suitably paranoid'), your hackles should be rising by now. Many of the risks are obvious. An intruder who obtains your password file has a precious thing indeed. The key database can be almost as good; private keys for individual users are generally encrypted with their login passwords."

The Network File System - The Network File System (NFS) [Sun88, Sun90] allows computers to mount file systems that are physically attached to other computers(12). This is desirable because it allows users to access files without having to transfer them across the network, saving time and removing the need to keep different versions in synch. However there are some serious security problems associated with NFS, for example :-

- NFS clients are allowed to read, change or delete files without having to log onto the server or enter a password.
- NFS has very weak client authentication.

- If not properly configured NFS can allow *any* other host to simply mount its file system.

Given the scope for abuse, it is not surprising that most of the literature advises extreme caution before allowing RPC based services through a firewall [Ches94, Chap95, Siya95, Stal95].

2.3.7. File Transfer Protocols

TFTP - The trivial file transfer protocol (TFTP) is a simple UDP based file transfer mechanism. TFTP is often used to boot diskless workstations. The protocol has no authentication. [Ches94] recommends that TFTP be disabled if it is not absolutely needed because its simplicity makes it very useful to attackers.

FTP - The file transfer protocol [Post85] is one of the most widely and heavily used Internet applications [MERI94]. FTP can be used to transfer both ASCII and binary files. Separate channels are used for commands and data transfer. Anonymous FTP allows external users to retrieve files from a restricted area without prior arrangement or authorisation. By convention users log in with the userid "anonymous" to use this service. Some sites request that the user's electronic mail address be used as the password.

The FTP daemon runs with extremely high privilege levels. [Ches94] reports that historically, there have been several bugs in the daemon, which have opened disastrous security holes.

Most organisations will require an anonymous FTP repository somewhere. [Ches94] provides some useful guidance as to how it should be configured.

FSP - (The name does not stand for anything) is dubbed the Sneaky File Transfer Protocol. It uses UDP port 21 to implement a service similar to FTP. It is unofficial, and is mainly used by hackers to move their ill-gotten gains around [Ches94]. Discovery of FSP traffic will therefore probably be a cause for concern.

2.4. The "r" Commands

The "r" commands (rlogin, rsh and rexd) can be used for remote terminal access and remote execution of programs. These programs are used in a trusted environment to allow users remote access without the need to re-authenticate themselves [Chap95]. The host that the user is connecting to trusts the host they are connecting from to have correctly authenticated the user.

The trusted host model is inappropriate for use on untrusted networks. Nevertheless the "r" commands are still widely used and have been involved in some of the higher profile attacks of recent times [Shim95].

2.5. World Wide Web WWW

Most commentators agree that the World Wide Web is responsible for the explosive growth the Internet has seen in the last two years. To many organisations, the World Wide Web *is* the Internet. The commercial implications of the World Wide Web are staggering, and so, therefore, are the implications of inadequate security. Information based industries such as banking and insurance are likely to be transformed by Web developments. Sales, marketing and post-sales support in other industries will similarly be affected.

Developments are occurring more rapidly than the IETF committees(13) can process them. For example version 1.0 of the HTTP [Bern96] is the de-facto standard WWW protocol, however it has still to be sanctioned by the IETF, version 1.1 is now an Internet draft [Field96], and there are several HTTP enhancements and extensions at Internet draft stage [Host96, Luot96, Hopm96, Khar96, Mogu96, Kris96, Holt96]. Netscape are currently releasing a new version of their browser, with significant new functionality, every quarter. This rate of development is also proving too fast for users [Robe96] who are often still struggling to deploy the previous version.

Despite the fact that information sent to and from web browsers is currently visible to others [Reyn96], organisations are increasingly building links to corporate databases into their World Wide Web Pages [Book96].

The incidence of attacks on web sites is increasing rapidly [Mill96] and WWW security is a significant cause for concern [Delv94] [Bild96] as the HTML specification allows protocols other than HTTP to be used (e.g., FTP, TELNET, RLOGIN). HTML may therefore be used to bypass the filters normally applied to those protocols by a firewall. This can be rectified by using an HTTP proxy which filters the relevant protocols as required [Dalv94]. Other problems include :-

- unexpected input values can cause actions which were not intended by the author ;
- special characters may allow unauthorised access to the host;
- unexpectedly large input may cause a buffer overflow resulting in inappropriate actions;
- the potential for data driven attack especially for Trojan horses;
- Authentication/Confidentiality/Integrity Issues - Mutual authentication and protection from message stream modification (e.g., to support electronic commerce).

Potential Solutions -

· Type enforcement - In 1985, the US government published the Trusted Computer Security Evaluation Criteria, more commonly known as the Orange Book, which offers a range of ratings for secure systems. The ratings start with D, for systems with no security, and go to A1, which requires formal methods to verify security. One of the key elements of the Orange Book was mandatory access control, where all the resources in the computer (such as users, files, services, and programs) are labelled with a security level, or sensitivity. The label identifies the degree of sensitivity of each resource, such as Unclassified, Confidential, Secret, and Top Secret. Labels effectively assign the data on the system to separate classifications. Many initial Orange Book implementations were too restrictive. The DoD funded seven years of R&D to create a flexible implementation, resulting in the type-enforcement security model.

Type enforcement is a security mechanism based on least privilege that controls how users, programs and data interact. Type enforcement works by grouping all the processes of the system into classes based on least privilege. Each process group is called a domain. In a similar manner, the files on the system are grouped into classes called types. The Domain Definition Table describes each domain's access rights for each type. The table cannot be changed while the system is running. Type enforcement also creates what is termed an

"assured pipeline" to organise data flow between programs - to assure that information moves securely, type enforcement controls the data each program can read and write. Each program can only read from the stage in front of it, and write to the next stage, of the pipeline. No stage of the pipeline can be skipped [DOD85, Thom90, Thom91].

· Digest Authentication - Digest based authentication uses a challenge-response paradigm in which the server issues a unique challenge string to the client. The client concatenates his password to this string and computes a one way hash of the result and transmits the result of this back to the server. The server also concatenates the users password with the original message and generates a digest. If this digest matches the one returned by the user then the client has been authenticated without his password having been transmitted.

This scheme requires the server to know the users password, however, the scheme proposed by [Host96] has a nice refinement which means that the server only needs to know the digest of the users password.

- S-HTTP - The secure hypertext transfer protocol [Resc96a] is compatible with HTTP but incorporates security extensions that support sender authentication, message integrity and confidentiality, and non-repudiation(14) of origin.
- SSL - The Secure Sockets Layer developed at Netscape Communications Corporation takes a radically different approach to S-HTTP. Rather than enhancing World Wide Web security by extending the HTTP application protocol, SSL creates channel security between the application layer protocol, HTTP, and the transport layer protocol, TCP. SSL is backed by, amongst others, IBM, Microsoft and SpyGlasss, all of whom are incorporating SSL in client server applications [Pomp96, Lips96, Moel95b]. Netscape have submitted SSL 3.0 in Internet Draft form [Frei96] to the Internet Engineering Task Force (IETF).

SSL is far from flawless. Opponents have expressed concern about the weakness of the encryption borne out when an SSL 2.0 key was cracked in 1995 by two Berkeley graduate students. Netscape uses a 40 bit key in the international versions of its browser and servers, as required by the US State Department under the RC4 regulations(15) which limit the export of products with encryption keys longer than 40 bits. These regulations make Netscape's 128 bit encryption key illegal outside the U.S.

· PCT - Microsoft's Private Communications Technology Protocol [Simo96] is a superset of Netscape's SSL protocol and is intended to address the perceived shortcomings of SSL. PCT will spawn a second key specifically for authentication. This will not fall under the RC4 restrictions because the regulations deal only with bulk encryption. Microsoft also intends to develop a more robust random number generator, used to seed the encryption key , as this is also considered to be a weakness in SSL. PCT is intended to be backward compatible with SSL 2.0

2.6. The X Window System

The X Window System was developed as part of Project Athena(16). Version 11 of the X Window System, commonly referred to as X11, was released in September 1987. With Release 2 of X11 in March 1988, control of X passed from MIT to the X consortium, an association of computer manufacturers who support the X protocol.

The X Window System is a network oriented windowing system. It uses the network for communicating I/O between the windowing display software and applications. An application need not be running on the same system that opens the display. The program that controls each display is the X server and the applications are the X clients(17).

[Ches94] points out that applications that have connected to an X11 server can "do all sorts of things". For example the screen contents can be printed and key presses can be both detected and generated. This allows an attacker to read passwords as they are being typed. The protection mechanisms, for example the so called "magic cookie" system, built into X11 are generally considered to be inadequate [Ches94, Hugh95, CA-95:07a].

Some research has suggested that it is feasible to allow X11 securely over the Internet using an application gateway [Winf93], however most of the literature recommends extreme caution when considering this [Hugh95, Ches94, Siya95].

(1) The TCP/IP 5 layer model does not have the Presentation or Session layers found in the ISO OSI 7 layer model. If session or presentation services are needed by the application layer they must be provided by the application itself.

(2) The JANET security unit recently found that a sniffer placed on the central Ethernet network at Cambridge University had been in place for four weeks before being detected

[Farr96]. Whilst the attacker did not delete or alter any files, he is believed to have gained access to more than 10 000 files including sensitive drug and genetic research.

(3) A NFS file server exports its file system to each of the clients listed in its configuration database. Ensuring that the client has a known IP address is the only client authentication performed by the server before allowing the client to mount the server's file system.

(4) Robert Tappan Morris would later release a program that exploited several weaknesses in the TCP/IP protocols to propagate itself around the Internet. This became known as the Internet Worm, and the widespread confusion it caused led to the foundation of CERT. Morris' father was a computer security expert working for the NSA [Mark88]. An detailed account of these events can be found in "Cyberpunks" [Hafn91], and [Eich89] presents an excellent analysis of the virus itself.

(5) The reports that this advisory refers to came initially from Tsutomu Shimomura, a security expert in the San Diego Supercomputer Centre, whose computers were attacked by Kevin Mitnick on Christmas eve of 1994. The fascinating story of how Shimomura tracked, identified and apprehended Mitnick is told in Shimomura's book "Takedown" [Shim95].

(6) Some of the more frequently used well known ports are listed in appendix IV.

(7) As DNS is intended to provide information it breaks the principle of security through obscurity. (This is discussed in Chapter 3)

(8) SMTP can be used for a denial of service attack by overloading a system with messages - as has happened to whitehouse.gov in the past.

(9) Each RPC based service is identified by a four byte "RPC Service Number". This caters for 4 294 967 296 different services.

(10) Many result directly from the fact that it is difficult to protect them with packet filtering techniques as they do not use fixed TCP or UDP port numbers.

(11) NIS was originally known as the Yellow Pages but this infringed British Telecom's trademark. The Network Information Service is still abbreviated in some texts to NIS/YP.

(12) NFS was designed for use in local area networks, the Andrew File System (AFS) was designed for use across wide area networks and better tolerates poor performance and lower degrees of trust [Chap95]

(13) For more information about how Internet standards are processed see [Hove96] which describes the organisations involved in the IETF. It includes descriptions of the IESG and Working Groups and their relationship with the Internet Society.

(14) This is a means of proving the origin of a message, i.e. preventing the sender from repudiating its origin.

(15) Named after the RC4 algorithm on which SSL is based.

(16) Project Athena was a collaborative research project between the Massachusetts Institute of Technology and IBM and DEC that began in 1983. Two of the most notable outcomes of this project are the X Window System and the Kerberos Authentication System.

(17) This may be counter-intuitive. Ordinarily the application acts as the server and the workstation as the client. In this case the workstation is the server because the workstation is accessible to other systems across the network., and acts as a display server

CHAPTER 3

Firewalls and Internet security - Computer Security Risks and Attacks

3.1. Introduction

[Chap95] identifies three generic risks when connecting a trusted network with one that cannot be trusted :-

- Intrusion
- Denial of Service
- Information theft

In a presentation titled "A Taxonomy of Internet Attacks - What you can expect to see" [Ranu95c] Marcus Ranum described eight types of attack from the Internet :-

- Social Engineering
- Impersonation
- Exploits
- Transitive Trust
- Data Driven
- Infrastructure
- Denial Of Service
- Magic

3.2. Generic Risks

A discussion of the generic threats identified by Chapman and of the types of attack identified by Ranum is given below.

3.2.1. Intrusion

Intrusion occurs when an attacker gains access to the system and is able to use it and modify it in the same way as a legitimate user. In some cases rigorous password protection can protect against this type of attack, with accounts locking after three failed access attempts etc. However policies need to be geared against social engineering attacks as well, where an attacker uses ploys such as posing as a senior manager and demanding an immediate password change to allow very important and urgent work to continue. Some attacks in this category will exploit weaknesses in operating system security and will not require the attacker to knock at the door, the door opens itself for them.

3.2.2. Industrial Espionage and Information theft

Industrial espionage is on the rise, [NCSA96] reports that there are currently 122 countries actively engaged in industrial and economic espionage to the benefit of their

respective states. A study in 1992 sponsored by the American Society for Industrial Security (ASIS) found that proprietary business information theft had increased 260 percent since 1985 [NIST95]. The data indicated 30 percent of the reported losses in 1991 and 1992 had foreign involvement. The study also found that 58 percent of thefts were perpetrated by current or former employees. The three most damaging types of stolen information were pricing information, manufacturing process information, and product development and specification information. Other types of information stolen included customer lists, basic research, sales data, personnel data, compensation data, cost data, proposals, and strategic plans [NIST95].

Most experts are pessimistic about the extent and the scale of the problem, for example the following extract is from a recent book by an acknowledged expert on Internet security [Chap95].

"....Espionage is much more difficult to detect than run-of-the-mill break-ins, however. Information theft need not leave any traces at all, and even intrusions are relatively rarely detected immediately. Somebody who breaks in, copies data, and leaves without disturbing anything is quite likely to get away with it at most sites.

In practical terms most organisations can't prevent spies from succeeding. The precautions that governments take to protect sensitive information on computers are complex, expensive and cumbersome; therefore are used on only the most critical resources. These precautions include electromagnetic shielding, careful access controls, and absolutely no connections to unsecured networks."

Traditional warfare may even be giving way to "Information Warfare". The implications that the failure of the communications infrastructure would have for technology dependent Western society have prompted both Britain and the USA to develop formal Information Warfare Policies. Information warfare represents a global challenge that faces all late-industrial and information age nation states. It also represents the cheapest way for less developed nation states and religious or political movements to anonymously and grievously attack major nations and industrial corporations [NCSA96]. During a World-wide Threat Assessment briefing to the US Senate Select Committee on Intelligence, John Deutch, Director of the US Central Intelligence Agency said :-

"While intelligence sources have only identified a handful of countries that have instituted formal information warfare programs, I am concerned that the threat to our information systems will grow in coming years as the enabling technologies to attack these systems proliferate and more countries and groups develop new strategies that incorporate such attacks."

3.2.3. Denial of Service

A denial of service attack seeks to deny use of resources to legitimate users. This type of attack can be achieved in a multitude of ways, for example by corrupting routing tables etc. causing messages to be re-routed, by overloading resources with junk messages, by damaging stored data, by locking user accounts, and so on.

Example 1 - The attacker ICMP bombs router off the network.

Example 2 - The attacker floods network link with garbage packets.

Example 3 - The attacker floods mail hub with junk mail (or many users send many messages to one address.)(1)

There is little that a network administrator can do to prevent denial of service attacks as an attacker can always attack upstream of the point of connection to the Internet and disrupt service. This is one of the reasons that people are wary of using the Internet for mission critical or time critical connectivity.

3.2.4. Malicious Code

Malicious code can be thought of as an indirect denial of service attack. Most users are now familiar with the threat posed by viruses, worms, Trojan horses and genetic algorithms. However new forms of malicious code are appearing all the time. A new type of virus attacks documents rather than programs using the advanced features in desktop productivity tools such as word processors.

Currently the two high risk areas for infection with malicious code are when downloading files or in binary attachments to mail messages. However new technologies are being developed that extend World Wide Web viewers by downloading and executing

software on the client rather than the server. Such programs are known as applets and greatly increase the risk of infection from malicious code.

3.3. Types of Attack

3.3.1. Social Engineering

Example 1 Inexperienced user is tricked into changing password

Example 2 Attacker masquerades as administrator and asks for password for some reason or gives user new password and tells them to change it.

The infamous computer criminal Kevin Mitnick, subject of the book "Cyberpunk" [Hafn91], used social engineering techniques extensively. He was, for example, able to obtain a Pacific Bell internal memorandum by posing as a Pacific Bell executive and asking the author's secretary to fax a copy to him. Mitnick had attacked the telephone company's computerised exchange and was able to divert the call to a friend's fax machine. The friend's fax machine had been reprogrammed to indicate the message had reached its correct destination. The details of the memo appeared on the front page of the New York Times shortly after in July 1988 in an article by John Markoff, the same journalist that would later that year break the story of Robert Tappan Morris and the Internet "Worm" Virus [Mark88] see also [Eich89].

Mitnick was also able to convince Neill Clift, a British computer researcher and VMS security expert that he was an employee of Digital Equipment Corporation. At Clift's request Mitnick supplied technical manuals that Clift believed could only have come from Digital and released detailed information about security weaknesses.

People generally like being helpful and co-operative and attackers exploit this ruthlessly. Social Engineering is very hard to protect against as it is essentially hitting a "soft" target and requires "soft" means of addressing it such as staff education, clear policies and mechanisms for reporting problems.

3.3.2. Impersonation

Any attack where the attacker captures valid user-id and password and reuses them to gain access to system.

Example 1 A user uses Telnet program to connect to system from remote site and an attacker with network sniffer such as tcpdump or nitsniff etc. captures the login session. The attacker is later able to login to system with captured user-id and password.

Example 2 The attacker writes a shell script to present a false login session to the user. The user enters his correct user-id and password which the script records before initiating a real login session to allow the user to login. The user thinks he has entered his password incorrectly and is none the wiser.

Impersonation attacks are primarily sniffer and spoofing attacks [Tard95], with attackers seeking to capture passwords. It is a mistake to dismiss attacks on passwords as being of little danger. The miscreant who attacked Eindhoven University of Technology in 1990 causing Wietse Venema to develop a tool called TCP-Wrappers used password guessing as his primary means of gaining accounts [Vene92]. This individual (Venema referred to him as "his pet") frequently deleted all files on target systems. Venema's "pet" has earned himself an interesting footnote in computer security history. He was known as "Berferd(2)" to Cheswick and Bellovin who described his activities in "An evening with Berferd" [Ches92] and in their book subtitled "Repelling the Wily Hacker" [Ches94]. As if this level of interest wasn't enough, Tsutomu Shimomura knew him as "Adrian" . Shimomura tracked him as he attacked and damaged computers all over the Internet and wrote about the experience in "Takedown" [Shim95]. Much of the early work on firewalls refers extensively to this one hacker, who escaped arrest or punishment as he conducted his attacks from Holland which had no laws to prevent him.(3) There are several excellent accounts of how system administrators have pursued hackers and of the tools they developed in doing so [Stoll89, Bell92, Bell94, Hafn91, Shim96].

3.3.3. Exploits

These are attacks that seek to exploit a hole in a piece of software. Most of CERT's advisories fall into this category. For example the UNIX sendmail program runs with system

privileges. Sending a message with the "To" and "From" fields completed as shown has given root access to the sender :-

To : mrinvisible@nonexistnat.com

From "| /bin/sed '1,//d' | sh"

Exploits succeed because badly written software is the norm, security is generally added as afterthought, too many programs run with excessive privilege violating the least privilege principle, and few programs use the operating systems underlying security features [Ranu96c].

3.3.4. Transitive Trust

Transitive trust attacks take advantage of the trust models used by remote services (such as the "r" commands discussed in chapter 2).

Example 1 Many networks use ".rhost" files so that users can log in from "trusted" hosts without giving a password. An attacker who gains access to a host and scans for exported file systems using a remote procedure call is able to build a trust model of the network. The attacker then compromises a user account on one of the remote computers to gain a foothold on an entirely new network. This is one of the attack strategies that the 1988 Internet "Worm" Virus used to propagate itself [Eich89].

3.3.5. Data Driven

Data driven attacks take the form of Viruses and Trojan Horses. For example an attacker can email the victim a postscript file with hidden file operations in it. If the victim displays the file on his workstation with a postscript interpreter (such as Ghostscript), the postscript interpreter will execute the file operations. These may perform actions such as adding the attacker's host name to the victim's ".rhosts" file allowing the attacker to gain access to the victim's computer.

The World Wide Web is currently particularly vulnerable to data driven attacks. The emergence of languages such as Java that will run code on the client computer present attackers with significant new potential for this type of attack.

A firewall can help to screen out some data driven attacks. Some firewalls vendors are incorporating anti-virus software into their products, and some are able to control executable files. However firewalls in general provide little protection from data driven attacks.

3.3.6. Infrastructure

Infrastructure attacks include DNS Spoofing, ICMP Bombing and Source Routing.

Example 1 ICMP Bombing. ICMP (Internet Control Message Protocol) is used to re-route traffic on the fly and by routers to notify a host when a destination system or network is unreachable. An attacker can use widely available tools such as "icmpbomb" or "nuke" to send ICMP "host unreachable" packets to a target system effectively knocking the network off the Internet.

Most firewalls and routers can screen ICMP traffic. However ICMP is used for legitimate purposes such as Ping and screening ICMP messages in routers can cause network problems. Firewalls that are a single point of connectivity correctly interpret ICMP without letting it through.

Firewalls can block and log all source routed packets and tools like TCP wrappers can detect source routed packets and trigger alarms. Many routers can block source routed packets.

3.3.7. Magic

These are attacks that nobody has thought of yet. Such attacks if and when discovered will be full of surprises. An illustrative (and possible) example is Racing Authentication, where an attacker is able to sniff packets as a legitimate user logs in with SecurID or other similar authentication token. The attacker mirrors the user's keystrokes and takes a guess at last digit of SecurID code, thereby winning the "race" with the user to login. If the attack is successful (an average of 1 in 10 should be) then the attacker is granted access, and the user probably just thinks they have made a typing error.

3.3.8. Combination attacks

Attackers are likely to use a combination of the above methods when seeking to gain unauthorised access or to deny service etc.

Exampe 1 The attacker tells a new user who is using IRC (Internet Relay Chat) to obtain a utility program that will help them to use system better. This phase of the attack can be categorised as Social Engineering. The user downloads the program and runs it causing all his messages to be deleted, and exposing the password file to the attacker. This phase of the attack can be categorised as Data-Driven.

3.3.9. Security Analysis Tools

There are several tools that will probe a computer to test for known vulnerabilities [Farm93, Farm94, Drew95, Tabi96]. Some of these tools are public domain, for example Farmer and Venema's SATAN tool. These tools can be used by system administrators to perform security audits, however they can also be used by attackers to probe for weaknesses.

(1)As happened to the law company that sent unsolicited mail advertising their services

(2) Berferd was the account name he captured at Bell Labs. The account name itself was based on an episode of the Dick Van Dyke show when Dick Van Dyke's brother called him "Berferd" - because he looked like a "Berferd".

(3) This changed in 1992. The first Dutch hackers to be arrested at the end of February 1992 were much less harmful though [Vene92].

CHAPTER 4

Firewalls and Internet security - Network Security Policy

Much of the literature on firewalls concentrates on diagramming the numerous possible configurations of routers, host systems, interfaces, and sub-nets. It is imperative, however, not to lose sight of the broad definition of a firewall as a part of security policy [Wack95].

The role of a security policy is to ensure that each of the four fundamental components that make up computer security, Authentication, Access Control, Integrity and Confidentiality are adequately addressed. Typical questions that need to be answered when developing a network security policy are :-

- What resources are we trying to protect ?
- Which people do we need to protect the resources from ?
- How likely are the threats ?
- How important is the resource ?
- What measures can be implemented to protect the resource ?
- How cost effectively and in what time frame can these be implemented ?
- Who authorises users ?

These questions should be revisited periodically, as network security is very dynamic.

The security policy identifies the threats that need to be protected against and defines the level of protection required. The security policy will itself contain several different policies, for example a Network Service Access Policy and System Specific Policies and will be based on a security strategy.

4.1. Security Strategies

Several generic strategies are documented in the literature.

4.2. Least Privilege

The principle of least privilege is to grant only those privileges that are required.

Systems that allow permission to be granted or revoked by operation providing fine-grain control are well suited to this. There is generally an overhead in terms of increased system maintenance.

Adopting a least privilege strategy limits exposure to attacks and importantly limits the damage that can be done when an attack is successful.

Many of the common security problems on the Internet can be viewed as failures to follow the principle of least privilege.

4.3. Defence In Depth

The defence in depth strategy is summed up by the term "Belt and Braces", i.e. use as many security mechanisms as you can and arrange them so that they back each other up.

One of the problems with firewall systems is that they provide an all or nothing solution to security. If the firewall is breached (and this has happened [CSI95a, CSI96, Shim95]) the internal network is a soft target. This was noted by Bellovin who coined the term "Hard on the outside, soft and chewy on the inside" [Bell92] to describe it. Some firewalls however do implement the principle of defence in depth using techniques such as Type Enforcement.

An important aspect of defence in depth that is often overlooked is the need to avoid common mode failures. For example if an attacker can exploit a security weakness in brand X's router then there is little point in having two of them. However brand Y's router may not have the same weakness and therefore the principle of defence in depth is met. This principle is important in the context of firewalls as many of the products commercially available are variations of Trusted Information Systems Gauntlet or their tool kit.

4.4. Choke Point

A choke point is a single point through which all incoming and outgoing network traffic is funnelled. As all traffic passes through a choke point it is the natural place to focus monitoring and control efforts such as Internet firewalls. It is also the natural place at which to break the connection with the external network if necessary.

Choke points are often criticised as an all-eggs-in-one-basket solution. This concern can be addressed by building some redundancy into the choke point. The key point is that the choke point provides control.

The largest threat to a choke point strategy is if an attacker is able to bypass the choke point. As Firewalls generally act as choke points this is a significant issue, especially given the ease with which SLIP(1) or PPP(2) connections to Internet Service providers can be established.

As choke points can experience high levels of network traffic it is important to ensure that there is sufficient bandwidth available at the choke point to prevent a network traffic bottleneck. Any monitoring and logging software should also be able to cope with the level of network traffic.

4.5. Fail Safe Stance

If a system is going to fail, it should be designed to fail into a safe state(3). This principle is particularly important in the design of Internet firewalls. Packet filters and application level gateways, both of which are discussed in the next chapter, should fail in such a way that traffic to and from the Internet is stopped.

4.6. Security Through Obscurity

This strategy is based on the hope that if you keep a low profile, would be attackers won't find you, and if they do, they will pass you by. Many companies do not publish the telephone numbers of their dial-in modems, only divulging the numbers on a need to know basis. Whilst this is a sensible precaution it is a poor basis for long term security. Information

tends to leak out, and attackers are often skilled at eliciting information from staff using social engineering techniques.

Many organisations assume that an attacker won't be interested in them, and that they are therefore unlikely to be the target of an attack. The rationale behind this stance assumes that a site is targeted because the attacker is interested in the information stored on it.

Such assumptions are, at best, naive. Several studies into the factors that motivate the perpetrators of computer crime [Stol89, Hafn91, Bell92, NIST95, Ranu95c, Shim95] have found that there is no single factor. Some attackers regard themselves as electronic freedom fighters battling against the commercialisation of the Internet, others wish to sell the information they glean, others are motivated by the power and control they wield over the lives of the system administrators they affect, and many attacks are motivated by revenge.

Attacks generally involve several computers and a multitude of accounts. An attacker may capture accounts and gain unauthorised access to several systems before reaching his real target. A site can be compromised for no other reason than to provide a staging post for attacks on other sites, and to the attacker, it means little more than another IP address.

4.7. Simplicity

Software is complex. As the size of a piece of software grows it becomes increasingly difficult to test all eventualities. Complex code will probably have unknown loopholes that an attacker can exploit. These loopholes may be convoluted but that will not prevent an attacker from trying to exploit them, some of the exploit attacks against sendmail have been extremely intricate.

Simplicity is an important factor in sound network defences. Application level gateway network Security systems should have all extraneous functionality removed and should be kept as small and simple as possible.

4.7.1. Host Based Security

Host based security is probably the most common computer security model in current use. The major problem with the host based security model is that it does not scale well. The major impediment to effective host security in modern computing environments is the complexity and diversity of those environments [Chap95]. Even if all hosts are identical, the sheer number of them at some sites makes securing each of them difficult. Effectively implementing and maintaining host security takes a significant amount of time and effort, and is a complex task.

Whilst the host security model might be appropriate for small sites, and whilst all sites should implement some level of host security, it is not cost effective for larger sites, requiring too many restrictions, and too many people [Chap95].

4.7.2. Network Based Security

Network security is designed to address the problems identified with host security. The network security model concentrates on controlling network access to hosts and services rather than on securing the hosts themselves.

Network security approaches include building firewalls to protect trusted networks from untrusted networks, utilising strong authentication techniques, and using encryption to protect the confidentiality and integrity of data as it passed across the network.

4.7.3. Security Policy

RFC1244 - The Site Security Handbook [Holb91] presents a useful guide to developing a site security policy. It is currently being revised and is due to be re-issued

shortly. The guide lists and discusses issues and factors that a site must consider when setting their own policies and makes some recommendations.

Useful guidance on some of the higher level requirements necessary for network security policy to be effective can be found in [NIST95] :-

To be effective, policy requires visibility. Visibility aids implementation of policy by helping to ensure policy is fully communicated throughout the organisation. Management presentations, videos, panel discussions, guest speakers, question/answer forums, and newsletters increase visibility. The organisation's computer security training and awareness program can effectively notify users of new policies. It also can be used to familiarise new employees with the organisation's policies.

Computer security policies should be introduced in a manner that ensures that management's unqualified support is clear, especially in environments where employees feel inundated with policies, directives, guidelines, and procedures. The organisation's policy is the vehicle for emphasising management's commitment to computer security and making clear their expectations for employee performance, behaviour, and accountability.

To be effective, policy should be consistent with other existing directives, laws, organisational culture, guidelines, procedures, and the organisation's overall mission. It should also be integrated into and consistent with other organisational policies (e.g., personnel policies). One way to help ensure this is to co-ordinate policies during development with other organisational offices.

4.7.4. Policy Site Security

The Site Security Policy is an overall policy regarding the protection of the organisation's information resources. This includes everything from document shredders to virus scanners, and remote access to floppy disk tracking. At the highest level, the overall organisational policy might state:

- Information is vital to the economic well-being of the organisation.
- Every cost-effective effort will be made to ensure the confidentiality, integrity, authenticity, availability and utility of information.
- Protecting the confidentiality, integrity, and availability of information resources is a priority for all employees at all levels of the company.

Below this come site-specific policies covering physical access to the property, general access to information systems, and specific access to services on those systems. The firewall's network service-access policy is formulated at this level.

4.7.5. Network Service Access Policy

The Network Service Access Policy is a higher-level, issue-specific policy which defines those services that will be allowed or explicitly denied from the restricted network, plus the way in which these services will be used, and the conditions for exceptions to this policy.

While focusing on the restriction and use of internetwork services, the network service access policy should also include all other outside network access such as dial-in and SLIP/PPP connections. This is important because restrictions on one network service access can lead users to try others. For example, if restricting access to the Internet via a gateway prevents Web browsing, users are likely to create dial-up PPP connections in order to obtain this service. Since these are non-sanctioned, ad hoc connections, they are likely to be improperly secured while at the same time opening the network to attack.

For a firewall to be successful, the network service access policy should be drafted before the firewall is implemented. The policy must be realistic and sound. A realistic policy is one that provides a balance between protecting the network from known risks while still providing users reasonable access to network resources. If a firewall system denies or restricts services, it usually requires the strength of the network service access policy to prevent the firewall's access controls from being modified or circumvented on an ad hoc basis. Only a

sound, management-backed policy can provide this defence against internal resistance. Here are the typical network service access policies that a firewall implements:

- Allow no access to a site from the Internet, but allow access from the site to the Internet; or, in contrast,
- Allow some access from the Internet, but only to selected systems such as information servers and e-mail servers.

Firewalls often implement network service-access policies that allow some users access from the Internet to selected internal hosts, but this access would be granted only if necessary and only if it could be combined with advanced authentication.

4.7.6. Firewall Design Policy

The Firewall Design Policy is a lower-level policy which describes how the firewall will actually go about restricting the access and filtering the services as defined in the network service access policy.

The firewall design policy is specific to the firewall. It defines the rules used to implement the network service access policy. This policy must be designed in relation to, and with full awareness of, issues such as firewall capabilities and limitations, and the threats and vulnerabilities associated with TCP/IP. Firewalls generally implement one of two basic design policies:

- Permit any service unless it is expressly denied; or
- Deny any service unless it is expressly permitted.

A firewall that implements the first policy allows all services to pass into the site by default, with the exception of those services that the network service access policy has identified as disallowed. A firewall that implements the second policy denies all services by default, but then passes those services that have been identified as allowed. This second policy follows the classic access model used in all areas of information security.

The first policy is less desirable, since it offers more avenues for getting around the firewall. For example, users could access new services currently not denied by the policy (or even addressed by the policy). For example, they could run denied services at non-standard TCP/UDP ports that are not specifically denied by the policy. Certain services, such as X Windows, FTP, Archie, and RPC are difficult to filter [Chap92], [Ches94]. For this reason, they may be better accommodated by a firewall that implements the first policy. Also, while the second policy is stronger and safer, it is more restrictive for users; services such as those just mentioned may have to be blocked or heavily curtailed.

Certain firewalls can implement either design policy but one particular design, the dual-homed gateway, is inherently a "deny all" firewall.

Systems which require services which should not be passed through the firewall could be located on screened subnets separate from other site systems.

In other words, depending on security and flexibility requirements, certain types of firewalls are more appropriate than others, making it extremely important that policy is considered before implementing a firewall. Failure to do so could result in the firewall failing to meet expectations.

4.7.7. System Specific Policies

System-specific policy is often implemented through the use of access controls. For example, it may be a policy decision that only two individuals in an organisation are authorised to run a particular program. Access controls are used by the system to implement (or enforce) this policy.

4.8. Incident Handling

When a site that is not protected comes under sustained attack one of two things can happen. The site can rapidly develop a policy and defences or it can withdraw from the Internet. Internet security incidents, such as break-ins and service disruptions, have caused significant harm to several organisations' computing capabilities. Many organisations have an ad hoc response when initially confronted with an attack which can exacerbate the damage caused by the attack. For this reason it is often cost-effective to develop an in-house capability for the quick discovery of, and controlled response to, network security incidents.

The primary benefits of an incident handling capability are the ability to contain and repair damage resulting from network attacks. An incident handling capability also assists an organisation to prevent, or at least to minimise, damage from future incidents. Incidents can be studied internally to gain a better understanding of the organisation's vulnerabilities so that more effective safeguards can be implemented [NIST95].

4.9. Disaster recovery (4)

It is prudent to assume that an attack may fundamentally compromise an organisation, for example deleting large amounts of data. It is for such eventualities that organisations develop disaster recovery plans. The basic steps in establishing a disaster recovery plan are :-

1. Identify the mission or business critical functions.
2. Identify the resources that support the critical functions.
3. Anticipate potential contingencies or disasters.
4. Select contingency planning strategies.
5. Implement the contingency strategies.
6. Test and revise the strategy.

(1) Serial Line Internet Protocol - A means of using IP over serial (telephone) lines.

(2) Point to Point Protocol - A replacement for SLIP

(3) For example an electrical switch should always fail to the open (i.e. off) position and break any circuit.

(4) A detailed discussion of the development and role of a disaster recovery plan is beyond the scope of this report. Interested readers should consult [NIST95].

CHAPTER 5

Firewalls and Internet security - Firewall Theory & Architectures.

5.1. What is an Internet Firewall?

Internet firewalls are a means of protecting networks by implementing access control to and from the Internet. In practice this is achieved by controlling the means of communication between the two networks, the TCP/IP suite of protocols. [Wack95] describes a Firewall as an approach to security. He uses the term Firewall to mean the strategies and policies and the term Firewall System to refer to the hardware and software elements that implement the policy.

[Chap95] notes that in practice an Internet firewall is more like a moat around a castle than a firewall in a modern building.(1)

A Firewall System is a collection of components that is placed between two networks and possesses the following properties :

- All traffic from inside to outside, and vice-versa, must pass through it [Chap95].
- Only authorised traffic, as defined by the security policy, is allowed to pass through it.
- The system itself is immune to penetration [Ches94]

In other words a Firewall System is a mechanism used to protect a trusted network whilst it is connected to an untrusted network.

Typically, the two networks in question are an organisation's internal network (trusted) and the Internet (untrusted). But there is nothing in the definition of a firewall that ties the concept to the Internet. Although the majority of firewalls are currently deployed between the Internet and internal networks, there are good reasons for using firewalls when connecting any trusted network with a less trusted network, be it internal or external.

5.1.1. What can a firewall do ?

A firewall can enforce security policy. The firewall is the means by which the network access security policy is implemented. Internet services considered to be insecure can be restricted and access to or from certain hosts can be restricted.

A firewall can log activity effectively.

A firewall can limit your exposure to the untrusted network by controlling/restricting access to/from it to the level defined in the security policy. This includes controlling what users use the Internet for.

A firewall can be a focus for security decisions - a choke point. All traffic to or from the Internet must pass through it. By focusing defences on this point they can reduce internal system security overhead since they allow an organisation to concentrate security efforts on a limited number of machines.

What can't a firewall do ?

Whilst firewalls provide good protection at the lower levels of the TCP/IP model, they provide almost no protection against higher level protocols [Ches94].

[Ches94] notes that any data that is passed by the firewall still has the potential to cause problems which, were these to be exploited deliberately would be labelled as denial of service or data driven attacks. For example a firewall offers no protection against viruses contained in files transferred via ftp or as MIME attachment to an e-mail message(2).

A firewall can't protect against malicious insiders. A firewall cannot differentiate between hosts on the same side of a network therefore any Internet Host can spoof any other Internet Host and any internal host can spoof any other internal host.

A firewall can't protect against connections that don't go through it (i.e. backdoors). Firewalls can restrict the access to certain facilities and users will sometimes bypass the firewall to gain access to those facilities. A good example would be a firewall that didn't allow access to the World Wide Web. Users on that network may establish point to point connections with an Internet service provider over a normal telephone line and introduce

Internet connectivity behind the firewall. This type of threat can only be addressed by management procedures which are embodied in the organisations security policies.

A firewall can't protect against completely new threats if the security strategy is different from "deny everything unless specifically permitted." Again this is dealt with within the security policy by basing it on just such a strategy.

5.2. Firewall Components

5.2.1. Packet Filters

A packet filtering system selectively routes packets between internal and external hosts according to rules that reflect the organisation's network security policy. Packet filtering may occur in a router, in a bridge, or on an individual host and operates at the network layer.

5.2.2. Screening Router

The type of router used in a packet filtering firewall is called a screening router(3) [Chap95]. It is configured with rules to block or filter protocols and addresses and is installed at the external network gateway. Internal users usually have direct access to the Internet while all or most access to site systems from the Internet is blocked. However, the router could allow selective access to systems and services, depending on the policy. Inherently dangerous services such as NIS, NFS, and X Windows are usually blocked [Chap95].

The screening router passes or rejects an IP packet based on information contained on the packet's header. The main information used is :-

IP Source and Destination Address - By filtering packets on the IP source and destination address the screening router is able to effectively block access to or from any site or host that is not trusted.

TCP or UDP source and destination port - The screening router makes use of the TCP "well known ports" to permit, deny, or re-route access to particular Internet services. For example many firewalls block all inward traffic except for email by rejecting all externally sourced

packets bound for any port other than port 25, the Simple Mail Transfer Protocol port. The screening router can also route all World Wide Web traffic (port 80) to a particular host.

A screening router is also able to base routing decisions on information not found in the packet header, for example the source and destination interfaces.

A packet filtering router can implement either of the design policies discussed earlier(4), however it suffers a number of disadvantages:-

- there is little or no logging capability. It is often therefore difficult for an administrator to determine whether the router has been compromised or is under attack;
- packet filtering rules are esoteric and difficult to test thoroughly, which may leave a site open to untested vulnerabilities;
- if complex filtering rules are required, the filtering rules may become unmanageable, and
- each host directly accessible from the Internet will require its own copy of advanced authentication measures.

These disadvantages become magnified as the security needs of a protected site become more complex and stringent. Screening routers alone are therefore considered to be inadequate for effective security(5) [Ches92, Ches92, Ranu93, Ches94, Chap95] and several firewall architectures, such as the screened host and screened subnet, have evolved to overcome these limitations. These provide additional security in packet filtering firewall implementations by utilising additional routers, hosts and perimeter networks. Before examining these architectures it will be useful to examine application level gateways, the means by which control of network traffic is extended from the network and transport layers to the application layer.

5.2.3. Application Level Gateways

Application level gateways are specialised application or server programs that run on a firewall host. These programs provide a safety barrier between the internal user and the Internet. Instead of connecting to the Internet directly with, say, a World Wide Web browser, the internal user connects to the application level gateway instead. The application level

gateway then establishes the connection with the required world wide web server on the Internet and acts as a go-between for the session.

Application gateways operate at the application layer and can therefore provide access controls at the application protocol level(6) and can handle store and forward as well as interactive traffic [Siya95] [Chap95].

The main disadvantage of application level gateways is that they require special purpose code to provide each service that is relayed. However, this means that they therefore implement a policy of "deny everything unless explicitly permitted" by default, which is often advantageous from a security perspective.

Application level proxies understand the application protocol and are therefore able to control the session based on the operations being requested. For example an application level proxy is able to block FTP PUT commands whilst permitting FTP GET commands.

The custom application acts as a "proxy" between the client and the server(7). Because all data between the client and the server is routed through the application proxy it is able to both control the session and provide detailed logging. This ability to log and control all incoming and outgoing traffic is one of the main advantages of application level gateway.

5.2.4. Circuit Level Gateway

Another type of application level gateway is called the circuit level gateway [Ches94]. Circuit-level proxies do not interpret the application protocols but they authenticate the user before establishing the circuits. They relay packets between the two communicating end-points but are not able to do any additional processing or filtering based on the protocol.

The advantage of circuit level gateways is that they provide services for a wide range of different protocols however they require special client software that has had system calls replaced with secure equivalents from a library such as Socks [Kobl92]. This re-introduces the problem that host based security does not scale well. As the size of the network increases the task of managing secure clients becomes increasingly time consuming and prone to error.

In general application level proxies use modified procedures and circuit level gateways use modified clients(8) [Chap95].

5.3. Firewall Architectures

The packet filtering technologies that are used in screening routers provide an efficient and general way to control network traffic. They have the advantage that no changes are required to host or client applications because they operate at the transport and network layers. Application level gateways extend control of network traffic to the application layer, and have the advantage that because they can understand the application protocol they can implement a finer degree of control and provide detailed logs.

Firewalls bring these components together to provide extremely effective network based security control. To illustrate this, several "standard" Internet firewall architectures or configurations are presented.

5.3.1. Dual-Homed Host

The simplest firewall architecture utilises a dual homed host. A dual-homed host is a computer that has separate network connections to two networks, as illustrated in figure 3. Such a host could act as a router between the two networks, however, this routing function is disabled when dual-homed hosts are used in firewall architectures.

Because the routing function is disabled the host isolates the two networks from each other whilst retaining the ability to see traffic on both networks. Systems inside the internal network can communicate with the dual homed host via one network interface, and systems on the Internet via the other, however these systems cannot communicate with each other directly.

In a dual homed host architecture the dual homed host itself is critical to the network's security. Such hosts are often referred to as Bastion Hosts in the firewall literature [Ches94, Wack95, Stall95, Siya95].

A dual homed host can only provide services by proxying them(9). Where proxies are not available a screened host or screened subnet architecture provide extra options for providing new and/or untrusted services.

5.3.2. Screened Host

In this architecture, illustrated in figure 4, the primary security is provided by packet filtering and a bastion host sits on the internal network providing the required application.

The screening router's packet filtering rules are configured such that the bastion host is the only host accessible from the Internet. Connections to the Internet may be routed through an application proxy on the bastion host, or in some cases, allowed directly through the screening router, depending on the network security policy.

[Stall95] and [Siya95] argue that the screened host architecture adds an additional layer of security to the dual homed host architecture, as an attacker has to first bypass the screening router, and then the bastion host. [Wack95] on the other hand argues that the screened host architecture, whilst more flexible than the dual homed host architecture, is less secure because the screening router is allowed to pass certain "trusted" services around the bastion host. Noting that the screened host architecture may, at first sight, appear to be less secure than the dual homed host architecture, [Chap95] states that this is misleading as the dual homed host itself may fail in some unexpected way, and that the two are therefore as secure as each other in practice.

However [Wack95]'s argument still holds - the additional flexibility afforded in the screened host architecture is provided at some cost to security.

5.3.3. Screened Subnet

With both the dual homed host and screened host architectures, the trusted network is vulnerable if the bastion host is compromised. The impact of the bastion host being compromised can be reduced by isolating it on a perimeter network(10). The simplest way to provide a perimeter network is to add an additional screening router to the screened host architecture. This architecture, illustrated in figure 5, is called the screened subnet architecture. The bastion host is then located on the perimeter network between the two screening routers.

An attacker that successfully compromises the bastion host now will only be able to access the perimeter net. The trusted network is still protected by the internal screening router. Whilst the attacker will be able to use packet sniffer software on the perimeter

network, he will not be able to collect passwords for, or to examine sensitive files on, the trusted network unless these are passed via the DMZ, which is itself a security weakness.

CHAPTER 6

Firewalls and Internet security-Variations of these architectures

The main components of firewalls have been presented, and the main firewall architectures have been examined. There are many variations of these architectures, for example providing internal and external demilitarised zones. These are discussed at some length in the literature. [Chap95] and [Siya95] both provide considerably more detail about additional architectures, and about configuring and implementing those discussed here.

(1) The Firewall literature is full of references to dragons, castles and other medieval lore. Two protagonists are largely responsible for this :

Steven M. Bellovin, a computer security researcher with AT&T titled his 1992 paper "There be dragons" [Bell932] and used several dialogues from J.R.R. Tolkien's Lord of the Rings and The Hobbit about the existence of, and dangers posed by, live dragons.

Marcus J Ranum who developed the first commercial application gateway or proxy server popularised the term Bastion Host referring to the extra defences (Bastions) that medieval castles had to protect them from intruders [Chap95].

Alluding to medieval castles and battles with dragons has had an interesting effect on the press, who are able to present stories of computer networks being attacked and damaged as a latter day struggle between romanticised forces of Good and Evil.

(2) In the case of viruses contained in e-mail one solution is to check all incoming mail with anti-virus software. Such software can generally now examine compressed or encoded files and will quarantine any messages with viruses or that it is not able to confirm does not contain a virus. The broader point still remains however, that if an application is poorly implemented the firewall will offer no protection from the consequences [Ches94].

(3) An ordinary router uses each packet's IP destination address to determine whether or not it *can* route the packet. A screening router looks more closely at a packet to determine whether or not is *should* route the packet [Chap95].

(4)If the router does not filter on source port or filter on inbound as well as outbound packets, it may become difficult to implement a " deny everything unless specifically permitted" policy [Chap95].

(5) [Moli95] presents an architecture for advanced packet filtering and argues screening routers alone can provide effective security.

(6) For application level gateways to be effective some form of IP level segregation between the Internet and the trusted network, such as a screening router or dual-homed host that doesn't route packets, must have been implemented.

(7) Application level gateways are often called "proxy servers".

(8) This is because the user connects to the host that is providing the application proxy rather than to the real host on the Internet. The real host's address must somehow be provided to the application proxy so that it can establish the other half of the connection. In the case of an application level gateway the real address can be included in the application protocol because the application level gateway understands the application protocol. However circuit level gateways do not understand the application protocol and the real address must be provided by some other means, hence the use of modified client software that provides the destination address to the proxy application.

(9) Strictly speaking this is not true, as services could be provided if users are allowed to log in to the dual-homed host itself. However this is a significant security risk [Ches94, Chap95, Siya95, Hugh95] and most users find it too inconvenient [Chap95].

(10) The perimeter network is often referred to as the demilitarised zone (DMZ).

CHAPTER 7

Firewalls and Internet security - Future Developments

There appear to be two areas that will significantly affect the future role and development of Internet Firewalls :-

1. The IP Security (IPSec) standards (RFC1825-1829).
2. The level of fragmentation in the firewall market.

7.1. The IP Security (IPSec) standards

In the middle ages tunnels provided a safe means of communication for besieged castles. Encryption tunnels that lead from a firewall to a firewall are analogous to this and are provided by the next generation of Internet Protocols (IPng).

The popularity of Internet firewalls to provide access control and protocol filtering services between protected sites and the Internet is due to the lack of robust security mechanisms in the TCP/IP protocol suite. The lack of robust authentication, integrity and confidentiality facilities necessitate a firewall in order for useful services such as NFS and the "r" services to be used safely.

However a firewall is restrictive. It may be desirable to use certain vulnerable services such as X or NFS, between remote sites, but a firewall will normally block such services. To be effective all or most traffic must pass through the firewall which can lead to bandwidth bottlenecks and performance problems depending on the load and type of traffic. If the firewall does prove to be a bottleneck internal users who do not wish to use the firewall will often also be affected.

To address these and other issues associated with IPV4, the next generation of Internet Protocols (IPng) and IPv6 (specifically the next version of IP) incorporate optional security headers. Because the security headers provide the basis for robust authentication, integrity,

and confidentiality, services deemed insecure with IPv4 could be quite secure with IPv6 (provided the security headers option is used !).

As a result the threat posed to a system using IPv6 should be significantly less than that posed to an IPv4 based system depending on the extent to which the security headers are used.

IPv6's security headers correct some problems that current firewall technology cannot correct, such as session stealing, in which an attacker can take over an established connection such as with TELNET. A practical method for defeating session stealing in IPv6 is continuous reauthentication, in which each packet would be authenticated to ensure it has originated from the legitimate user.

IPv6 security headers and related items have been defined, but how the headers will be used in conjunction with security gateways and other systems is still open to debate and experimentation. IPv6 security services could be used directly between hosts with no security gateway intervention, which would indicate that the security gateway may become involved only in those security functions that IPv6 does not handle, e.g., robust user authentication in TELNET.

Discussing firewalls, RFC1825 reads:-

"Firewalls are not uncommon in the current Internet [Ches94]. While many dislike their presence because they restrict connectivity, they are unlikely to disappear in the near future. Both of these IP mechanisms(1) can be used to increase the security provided by firewalls.

Firewalls used with IP often need to be able to parse the headers and options to determine the transport protocol (e.g., UDP or TCP) in use and the port number for that protocol. Firewalls can be used with the Authentication Header regardless of whether that firewall is party to the appropriate Security Association, but a firewall that is not party to the applicable Security Association will not normally be able to decrypt an encrypted upper-layer protocol to view the protocol or port number needed to perform per-packet filtering OR to verify that the data (e.g., source, destination, transport protocol, port number) being used for access control decisions is correct and authentic. Hence, authentication might be performed not only within an organisation or campus but also end to end with remote systems across the

Internet. This use of the Authentication Header with IP provides much more assurance that the data being used for access control decisions is authentic.

Organisations with two or more sites that are interconnected using commercial IP service might wish to use a selectively encrypting firewall. If an encrypting firewall were placed between each site of a company and the commercial IP service provider, the firewall could provide an encrypted IP tunnel among all the company's sites. It could also encrypt traffic between the company and its suppliers, customers, and other affiliates. Traffic with the Network Information Centre, with public Internet archives, or some other organisations might not be encrypted because of the unavailability of a standard key management protocol or as a deliberate choice to facilitate better communications, improved network performance, and increased connectivity. Such a practice could easily protect the company's sensitive traffic from eavesdropping and modification.

Some organisations (e.g., governments) might wish to use a fully encrypting firewall to provide a protected virtual network over commercial IP service. The difference between that and a bulk IP encryption device is that a fully encrypting firewall would provide filtering of the decrypted traffic as well as providing encryption of IP packets."

Some firewall developers are deploying components of this technology already. The Secure Wide Area Network (S/WAN) Initiative was announced earlier this year and aims to provide encrypted IPv4 tunnels between different vendors firewalls. The S/WAN Initiative is intended to promote multi-vendor virtual private networks among firewall and TCP/IP vendors. The initiative will make recommendations and additions to the underlying IPSec standard to achieve this goal. Four key management protocols are included in the S/WAN initiative, SKIP [Aziz95], Photuris [Karn96] and Oakley [Orma96]. S/WAN is promoted by a group comprising RSA Data Security (who own the RSA encryption algorithm) and eighteen partners including IBM and Checkpoint software [Andr96b]. These vendors are already testing vendor to vendor interoperability.

7.2. The level of fragmentation in the firewall market

As was shown in the case studies organisations have difficulty differentiating firewall products. To be able to understand the differentiating factors requires significant technical knowledge as most are based on the same technology. Vendors increase the level of

confusion by focusing on small technical details, and presenting them as key differentiating features.

The firewall market is characterised by a mixture of large mainstream computer suppliers such as IBM Corp., SUN Microsystems Inc., Digital Equipment Corp. and relatively small start-ups. The firewall market has seen phenomenal growth with researchers estimating that it will grow 70% from 1.1 Billion USD in 1995 to 16.2 Billion USD in 2000 [CSI95b]. Commercial systems have 84% market share, however twelve different products each had 1% share in 1995 and the largest single share belonged to non-commercial systems (16%). The market is still very fragmented and analysts expect significant change including a large reduction in the number of suppliers as the market develops [CSI95b]. Some analysts have suggested that large suppliers will dominate the market and that smaller companies will sell their technology and leave the market [McGa95]. However there are indications that the familiar computer industry model of David beating Goliath may apply again. One company in particular is attracting the sort of attention Netscape received in its early days. Founded in 1993 V-ONE (Virtual Open Network Environment) is being heralded by security experts as "a shining example of Internet innovation - perhaps as much as a year ahead of the rest of the pack" [Netw96]. V-ONE announced on April 3, 1996 that they had been selected by the National Security Agency (NSA) to provide firewall protection to the United States Federal Government. They have developed software the company calls Security Middleware [Ranu96] that allows businesses to use smart cards to send and receive secure transactions on the Internet. Marcus Ranum, Chief Scientist at V-ONE and developer of Trusted Information System's Gauntlet, the first commercial firewall, claims this is pushing firewalls to the next level [Info96]. Security Middleware is a layer of security existing between an application and its remote user. It differs from browser based security methods in that it provides strong authentication, a stronger method of encryption, and fine grain access control. V-ONE's Smartwall product combines a dual homed application level gateway with strong token based authentication and encryption, and is seen by many as the future direction for firewalls [Basc96, Moel96, Info96, Mere96, Elec96, Wing95, Rodr95].

(1) This refers to the two IP security mechanisms provided by IPSec, IP Authentication Header (AH) and IP Encapsulating Security Payload (ESP).

The IP Authentication Header is designed to provide integrity and authentication without confidentiality to IP datagrams. The lack of confidentiality ensures that implementations of

the Authentication Header will be widely available on the Internet, even in locations where the export, import, or use of encryption to provide confidentiality is regulated (RFC1827).

The IP Encapsulating Security Payload (ESP) is designed to provide integrity, authentication, and confidentiality to IP datagrams (RFC1826)

CONCLUSION

This report, pointed out clearly that our dependent life on computers are in constant need of security due to various facts that many want to "dive into" it and obtain our personal information. The reasons are different. Some may just want to prove themselves, some may want to sell the information and some may want to use it their advantage.

Whatever the reasons are, we need to feel secure and firewalls are built for this purposeonly. This report outlined how and why our information is in danger, how we can protect it and an in -depth explanation of how firewalls work.

To conclude, firewalls and the current network systems around the world are getting more and more developed and future developments will prove better as people gain experience. In the near future we may not have firewalls instead and use Artificial Intelligence methods in order not to project our data but instead to enable only the authorized ones to access it.

REFERENCE

[1] http://www.matlus.com

[2] http://www.symantec.com

[3] http://www.webopedia.com