New Zealand Security of Information Technology Publication 203

THE SECURE USE OF PUBLIC DATA NETWORKS

CHAPTER 1

THE USE OF PUBLIC DATA NETWORKS

General

101. Public Data Networks (PDNs) have emerged as a cost effective communications medium for organisations with modest data transfer requirements, and provide a widely available means of electronic access to departmental information for businesses and the public. PDNs also offer access for departmental staff to a wide variety of information sources, and an increasing number of on-line, electronic commerce systems. One of the most popular PDNs is the Internet.

102. The Internet began life in 1969 as a research network developed by the US Defense Advanced Research Projects Agency known as ARPANET. Connectivity to this network has been continuing at an ever increasing pace, originally within the academic community, and by the early 1990s the Internet consisted of about 2.5 million host computers and 35 million users. At that time, Internet users still needed to use cryptic command-line access methods, but the emergence of the Windows-like MOSAIC interface, to become known as the World Wide Web (or just 'the Web') provided a useable interface for a much wider user community. US commercial and government interest in the Internet started to occur, and by the mid-199Os New Zealand government agencies had established a significant presence on the Internet.

Internet Services

103. The Internet offers a range of specific services:

World-wide e-mail: It is doubtful whether any other communications medium provides access to as many government-related people and organisations while also buffering messages so that participants need not be simultaneously accessible in order to communicate. Telephone connectivity exceeds Internet connectivity, but buffering is provided only by answering machines or voicemail systems that are less widely deployed and lack the capability to transmit the scope and variety of information that can be sent digitally.

Access to open information sources via WWW: Since the release of the NCSA Mosaic browser in 1993, use of the World Wide Web and its associated protocols and formats has mushroomed. A great deal of commercial and marketing information is available through the Internet Web system, and a wide variety of research publications, including documents not yet published in journals, can be found on the Web. All kinds of public domain software are

available from the wide range of commercial and academic servers on the Internet, and software patches for commercial programs are routinely distributed in this fashion. For many organisations, the Web is becoming a primary means of providing and retrieving public information.

Real-time communications: E-mail can occasionally provide close to real time communication if both parties are connected and active, but voice and video communications via the Internet are also becoming available, despite bandwidth limitations. There is a great deal of activity both in increasing the available bandwidth and in the development of efficient protocols to support video and voice transmission over the Internet. Use of the Internet for teleconferencing purposes is likely to grow substantially as these facilities become available because of the wide and relatively low cost access available to the Internet from many government facilities and the scarcity and high cost of competing teleconferencing equipment.

Electronic commerce: Following the lead of industry, and with the encouragement of the central IT Policy Unit, government departments will increasingly adopt electronic commerce as the preferred means to conduct business with industry suppliers. Business transactions conducted through electronic commerce are potentially more efficient and less error-prone than manual commerce systems.

Resilience in non-critical situations: While the Internet does not match the availability and reliability of the telephone and power grids in most parts of the world, it provides adequate reliability for most routine uses. Despite widely publicised forecasts of impending "Internet collapse" and substantial increases in its use, no major long term outages have been observed, and the outages that have occurred have generally been resolved fairly quickly. Increasing commercial use of the Internet is encouraging the development of more reliable communications services.

Easy accessibility and interoperability: Access to the Internet is readily available from departmental locations and mobile telephone connectivity. The Internet protocol suite has become a standard in the commercial marketplace, a goal that other programmes such as OSI have sought but not achieved. Commercial networking and operating systems use, or are being converted to use, the Internet protocols.

Low cost: Compared to alternative telecommunications facilities, Internet communications are cheap. Any assessment of the costs of alternative communications is likely to end heavily in favour of the Internet. For this reason, use of the Internet in place of dedicated leased lines is an attractive path for many departments.

Internet Security

104. The lack of security in the Internet is widely acknowledged. Traffic on the Internet is transmitted in the clear, unencrypted. It is transmitted in standard formats using public protocols. It passes through switches and

communication facilities that are publicly accessible and in many cases relatively uncontrolled. Under such conditions, almost any kind of attack is feasible. Eavesdropping unencrypted information on the Internet is trivial and, even if some level of encryption is applied, the source and destination of most traffic remains visible. Identifying the participants in an electronic dialogue is not difficult even if the contents of the dialogue are protected. On the other hand, it is relatively easy to generate packets with bogus source address information, permitting rogue users or systems to masquerade as legitimate users. Unprotected messages can be undetectably altered in transit, and denial of service attacks are simple to mount against specific portions of the network. For the foreseeable future, security on the Internet will be a user responsibility.

Global Information Infrastructure

105. The US Administration is leading the development of the next generation Internet which has been called the Global Information Infrastructure (GII) which is focused on the development of a higher performance and more reliable underlying network with access methods better suited to electronic commerce. This development is likely to have a substantial impact throughout the world. The GII is expected to supersede current private networking systems in the longer term, and departments will be under increasing pressure to exploit the GII for all their wide area communications requirements.

CHAPTER 2

COMPONENTS OF THE INTERNET

General

201. While there are substantial benefits to the use of the Internet, these are accompanied by a number of risks both in system connection and data transmission.

202. The Internet is a global, decentralised network of communication links, switching devices, and computers that communicates using the Internet protocol suite. The U.S. Federal Networking Council adopted the following definition of "Internet" on October 24,1995:

"Internet" refers to the global information system that (i) is logically linked together by a globally unique address space based on the Internet Protocol (IP) or its subsequent extensions/follow-ons; (ii) is able to support communications using the Transmission Control Protocol/Internet Protocol (TCP/IP) suite or its subsequent extensions/follow-ons, and/or other IP-compatible protocols; and (iii) provides, uses or makes accessible, either publicly or privately, high level services layered on the communications and related infrastructure described herein.

203. To understand where and how information flowing on the Internet can be disclosed, modified, or delayed, it is useful to understand something of its component parts: links, nodes, addressing and domain naming, and network and application protocols.

Links

204. The primary links of the Internet are provided by lines leased from common carriers in countries around the world They are part of the public switched telephone networks (PSTNs), although lines used for Internet traffic are in general distinct from those used for voice traffic - a line leased for Internet use will be dedicated to that purpose. The actual media that carry the traffic include terrestrial copper wire and optical fibre, radio and cellular phone links, and satellite links.

Nodes

205. The primary nodes of the Internet are called routers. These are computers programmed to accept and forward packets of data (which may represent real-time voice or video information as well as numbers or text) across the links. A router will be connected to two or more links. It examines each inbound packet for its destination address and, based on its routing tables, determines over which outbound link the packet should be forwarded.

206. Routing tables can be updated dynamically in order to permit the network to adapt both to outages and to new links. Although the original theory behind the Arpanet's switching scheme was to provide fully decentralised and dynamic routing, the tremendous size to which the Internet has grown has led, in practice, to somewhat more hierarchical and static routing regimes. Each packet in a series originating from a host attached to a local Internet Service Provider (ISP) and destined for a host attached to an ISP in another country is likely to traverse the same route, and that route is unlikely to include random ISPs in either country. Rather, the packets will be sent from a local ISP in the source country to major routers and circuits controlled by the large common carriers, across to the major carriers in the destination country, and then to the local ISP that connects the destination host. The routing on the links controlled by the major common carriers can vary, but it is relatively static and is unlikely (though not impossible) for it to cause packets to traverse arbitrary routers in out-of-the-way corners of the Internet. There may, of course, be extensive local area networks (LANs) in place at both the host and destination locations, and the packets may be broadcast widely within those areas depending on the configurations and protocols in use (see *NZSIT 202: LAN Security* for details of LAN operation and security).

207. Host computers are another class of Internet nodes, and represent the sources and destinations of the packets that are routed through the network. Additional types of network nodes include gateways and firewalls; both of

these can be considered types of routers. Gateways that connect networks using different protocols to the Internet will need to provide protocol translation services and addressing services. Firewalls can filter incoming and outgoing traffic, translate addresses and more.

Addressing

208. Internet Protocol Version 4 (IPv4) network layer addresses consist of 32 bits, usually represented as four decimal numbers from zero to 255, separated by dots: eg 132.250.80.57, and logically divided into a link number and a host number (originally, link number consisted of a net number and subnet number, but this distinction did not prove useful). A host may be connected to more than one link and may therefore be reached by more than one IP address. Packets contain both source and destination addresses, and software in the sending node is responsible for providing both of these. In addition to the source and destination, the IP header may contain a number of optional fields, including a method to specify a route through the network and to require the destination host to route packets back to the source over the same route (this is the loose source route option).

209. The rapid growth of the Internet has led to a problem in allocation of IP addresses, and a larger address space is required. Internet Protocol Version 6 (IPv6, also known as the Internet Protocol next generation or IPng) expands the IP address space substantially, allowing 128 bits for source and destination addresses. IPv6 addresses are assigned to interfaces, not nodes; any of the unicast addresses assigned to any of the node's interfaces can be used to identity that node. LPv6 will eventually supersede the full IPv4 addressing scheme.

210. Hosts or networks on the Internet have fixed IP addresses. However, workstations accessing the Internet through a dial-up Serial Link Interconnection Protocol (SLIP) or Point to Point Protocol (PPP) are more usually allocated a dynamic IP address by the Internet Service Provider OSP). This address is unlikely to be the same from one session to the next.

Domain Name System

211. Packets traverse the Internet using numerical addresses, but users and programs usually deal with more mnemonic addresses in the form of Domain Names. The Domain Name System (DNS) provides the infrastructure for translating domain names into IP addresses and is fully described in the Internet standards RFC-1034, RFC-1035, and RFC-2065.

212. A domain name is a sequence of character strings separated by periods, eg dse.defence.govt.nz. Although software that processes domain names preserves the case of the character strings, the name is case insensitive so DSE.Defence.govt.NZ has the same meaning as the previous example. The domain name space is a tree structure. The resource records stored for a

domain name that corresponds to a host can carry additional information such as the operating system and version number associated with the host. Other resource records can designate a host that processes incoming mail for the specified domain, identify a name server for a domain, or map an alias to the real domain name for a host. DNS can also provide an inverse mapping from IP address to host name; however, there is no enforced relationship between the inverse mapping and the forward mapping. Many host names can map to a single IP address, but given an IP address, DNS returns a single corresponding domain name.

213. The size of the Internet dictates that the DNS database is distributed among many servers (called name servers), none of which has a complete copy. A name server knows the parts of the domain tree for which it has complete information; it is said to be an *Authority* for those parts of the space. Authoritative information is organised into units called *Zones,* and these zones can be automatically distributed to the name servers that provide redundant service for the data in a zone. A name server must periodically refresh its zones from master copies in local files or foreign name servers. A name server may also cache information about other parts of the domain tree for which it is not an authority.

214. To use DNS, a program typically invokes a local procedure called a *Resolver.* The resolver contacts a name server (usually, but not necessarily, on a different machine) to satisfy the request. The name server either provides the information requested by the resolver or a pointer to another name server that the resolver can try.

215. The original DNS database was designed to support queries to a statically configured database. Changes to the database were expected to be infrequent and were to be made as external edits to a zone's Master File. However, mechanisms have been designed for permitting dynamic updates to DNS records.

216. Although DNS is a critical operational part of the Internet infrastructure, it has no strong security mechanisms to assure data integrity or authentication. However, extensions to provide these services to security aware resolvers or applications through the use of cryptographic digital signatures have been proposed in RFC 2065 and are under review and are expected to include the proposed dynamic update mechanisms.

Protocols: IP, ICMP, UDP, and TCP

217. The Internet Protocol (IP) provides an unreliable, connectionless, best-effort delivery service that routes datagrams (packets) towards a specified IP address. IP also includes a protocol for reporting errors, the Internet Control Message Protocol (ICMP); routers use ICMP messages to report delivery failures, misroutings, congestion, and related problems to each other. Both Transmission Control Protocol (TCP) and User Datagram Protocol (UDP) are layered on top of IP. Application programs can use TCP or UDP to send messages to applications running on remote hosts.

218. TCP uses IP services to implement a reliable, connection-oriented transport service. TCP incorporates features to ensure that the messages it receives for transmission are delivered to the correct address uncorrupted, without duplication, and in order. User Datagram Protocol (UDP) uses IP services to provide connectionless datagram service although, due to its connectionless nature, UDP packets may be lost, duplicated, or delivered out of order.

219. Both UDP and TCP incorporate the notion of logical ports to distinguish traffic sent to the same IP address but for different recipients. A port number in both protocols is a 16-bit integer. An application on one machine can send UDP datagrams to different processes *(listening to* different UDP ports) on a remote machine by addressing the datagrams to different ports. UDP queues traffic for different ports independently. For TCP, the connection is the fundamental abstraction, and a connection is specified by its two endpoints. Each endpoint is a pair of integers (host, port), where *host* is the host's IP address and *port* is a TCP port number on that host. This arrangement permits, for example, a program that accepts incoming mail to use only one local TCP port even though it may be communicating over many connections concurrently. TCP and UDP port numbers are independent, since each message specifies its protocol as well as its destination IP address and port number, and in both protocols certain protocol numbers are used, by convention, as the addresses for particular services. Such a reserved, and advertised, port number is called a *well-known port.* The well known port for e-mail delivery, for example, is TCP port 25; UDP port 53 provides access to the Domain Name Service (as does TCP port 53).

220. A TCP connection is opened with a three-way sequence:

a. the initiating host sends a "SYN" message with its IP address and an arbitrary number sequence number N,

b. the destination host replies by sending an acknowledgment and another arbitrary sequence number M, and

c. the initiator completes the protocol by acknowledging the second message.

221. Closing connections is slightly more complicated version of connection establishment, as there are various contingencies to consider, such as messages en route at the time that one of the participants requests that the connection be closed.

Application Layer Protocols: Telnet, FTP, SMTP, HTTP, SNMP

222. The protocols that most Internet users see most directly are those at the application layer. These protocols support the transport of e-mail (SMTP) and files (FTP), the initiation of terminal sessions on remote hosts (Telnet), and the operations of World Wide Web Browsers (HTTP) and many other functions. These protocols may use TCP, UDP, or both to accomplish their

functions. Typically, a host that supports a particular service, such as e-mail delivery, will have a process or processes that listen for requests phrased in the appropriate protocol (such as SMTP) over a particular port, as noted above.

223. The Simple Network Management Protocol (SNMP) provides for the management of multivendor networks and became an Internet Recommended Standard in 1989. It operates over many different transport mechanisms and is not tied to any specific machine architecture. In addition, the simplicity of the protocol eases implementation and reduces vendor development costs. SNMP allows the network administrator to address queries and commands to network nodes and devices, and provides the functions of network monitoring, control, reporting, analysis, and fault-isolation.SNMP operates on three basic concepts: manager, agent, and the Management Information Base (MIB):

- A manager is a software program housed within a Network Management Station. The manager has the ability to query agents, receive agent responses, and set specific variables using various SNMP commands.
- An agent is a software program housed within a managed network device (such as a host, gateway, terminal server, etc.). An agent stores management data and responds to the manager's data requests.
- The Management Information Base (MIB) is a database of managed objects, accessible to agents and manipulated via SNMP to provide network management information.

CHAPTER 3

INTERNET VULNERABILITIES

General

301. Any system that is connected to the Internet and uses the Internet to communicate with other systems inherits many well-known vulnerabilities. Some of these arise in the Public Switched Telephone Networks (PSTNs) that carry Internet traffic among routers and hosts, while others come from the routers and hosts themselves. Many have to do with the ways in which people use the Internet and come to rely on it as a means to access both information and processing resources.

302. New vulnerabilities related to the Internet and Internet-connected hosts are being reported on a daily basis. This chapter provides some information on the major classes of vulnerabilities. The GCSB maintains a database of known vulnerabilities and departments should contact the GCSB for up-to date, detailed information related to specific system configurations.

Network Vulnerabilities: Sniffing

303. The major benefit of the Internet is its low cost which results from the use of shared communication channels. Consequently, computers on a shared

channel can receive information that is intended for other machines. To capture another user or system's information being transmitted across the shared channel is called *sniffing*.

304. The most popular way of connecting computers is through Ethernet LANs. The Ethernet protocol works by sending packet information to all the hosts on the same circuit. The packet header contains the proper address of the destination machine and only the machine with the matching address is supposed to accept the packet. A machine that is accepting all packets, regardless of the packet header, is said to be in promiscuous mode.

305. In a normal networking environment, account and password information is passed along Ethernet in clear-text and it is not hard for an intruder or a local user on the same network to put their workstation or host into promiscuous mode and, by sniffing, compromise all user accounts on the network.

Network Vulnerabilities: Wardialing

306. Although the system security policies of most protected networks will limit uncontrolled connections to the Internet, there are often unauthorised modem connections to the telephone system. These connections can be exploited by an attacker who can identify such a connection.

307. A common method of identifying such connections is wardialing, in which an attacker programs a computer to connect to a series of telephone numbers. Any answers which indicate the number is either a modem or a fax machine are logged for further investigation.

Protocol Vulnerabilities: Data Link Layer Security

308. Address Resolution Protocol (ARP), which is used to translate Ethernet addresses on a LAN to IP addresses, is open to manipulation. For instance, Unix System V does not check whether received ARP packets are associated with an outstanding request. This could allow malicious responses to ARP requests and unsolicited updates to ARP tables, resulting in denial of service or man-in-the-middle attacks in which one address masquerades for another.

Protocol Vulnerabilities: Network Layer Security.

309. There are a range of vulnerabilities at the network layer. The Simple Network Management Protocol (SNMP) has poor authentication, and unless the routers are correctly configured, they are vulnerable to malicious reconfiguration. IP allows direct network injection of IP packets. ICMP has no authentication, and can be subverted to send unsolicited address mask reply packets.

310. Network security can be further compromised by the protocols used to manage the network routers.

Protocol Vulnerabilities: Transport Layer Security

311. TCP provides no enforcement of port numbering, and ports may receive messages in any protocol. TCP checksumming of IP packets is not strong, leading to a potential for forgery, injection of data and tailgating of packets. The quality of random numbers for TCP initial sequence numbers varies across UNIX systems, leading to a potential to inject packets into a connection between two users.

Protocol Vulnerabilities: IP Origin Forgery

312. The origin of an IP message can be forged relatively easily. While this is not on its own a serious vulnerability, many higher level protocols unwisely use the IP origin as a form of identification. For instance, the r-commands, which allow managers of one UNIX system to control another, use IP source as a primary authentication method.

Application Vulnerabilities

313. Many of the higher level protocols can be exploited to attack systems connected to the Internet. Many vulnerabilities are well known, such as those in versions of the Sendmail system that allow an attacker to gain root privileges quickly. At this point, the attacker can stop audit of his or her actions, delete any previous audits, install Trojan horse software, read, modify or delete user applications or data, and then use the current system as a platform for launching an attack on further systems.

314. Many other protocols and software components contain similar bugs and vulnerabilities. Many of these bugs originate from simple errors in the program code, such as a failure to check array bounds. Indeed, with the applications originating in university rather than commercial environments, and being written primarily with functionality rather than robustness in mind, it is not surprising that so many protocols and servers are open to attack.

Application Vulnerabilities: World Wide Web.

315. The risks associated with World Wide Web (WWW) security vulnerabilities affect confidentiality, integrity and availability of intormation. A WWW server provides access to local information that the whole Internet can potentially access. If a Web server can be subverted then any or all the following may occur:

a. confidential documents held on the server may be accessed by unauthorised users;

b. public documents, e.g. Web pages, may be changed thus destroying the integrity of the information. This has been graphically illustrated by recent attacks on Web servers belonging to the US Central Intelligence Agency, US Department of Justice, and the British Labour Party, and

c. remote users may execute commands on the Web server's host machine providing access to operating system level functionality.

316. Recently identified Web server vulnerabilities include a bug which allows remote users to download and read the contents of executable scripts thus potentially accessing sensitive information such as database passwords or network configurations.

317. Many Web servers provide additional functionality through server side processing. This includes the use of Common Gateway Interface (CGI) scripts normally written in a scripting language such as Perl. CGI scripts may unintentionally leak information about a system or may be tricked into executing commands. An example is a widely published CGI script which inadvertently allows remote users to write files into directories to which they don't have access.

318. Client side network browsers may also suffer from security vulnerabilities. Most browsers provide functionality to allow additional applications to be specified when accessing a particular document type. It would seem natural to specify a word processing application to view documents. This, however, potentially leaves the application vulnerable to macro viruses.

Application Vulnerabilities: Java and ActiveX

319. Increasingly, there has been a move towards providing rich networked Web functionality via technologies such as Java and ActiveX. Java programs are precompiled and stored on a Web server. These mini-applications are known as Java applets and are downloaded to a network browser and executed locally. ActiveX is a technology for distributing software over the Internet. The ActiveX analog of a Java applet is called a *control.* An ActiveX control may be embedded in a Web page where it may be accessed via a network browser. Both these systems introduce vulnerabilities:

a. Java has been explicitly designed to address security issues through various mechanisms which essentially restrict the behaviour of applets. However, a number of implementation problems have been identified allowing Java applets to execute arbitrary machine instructions, interfere with other applets and bypass the Java Security Manager.

b. ActiveX controls are not restricted in any way and rely on digital certification. If a browser accesses an ActiveX control which is not signed or

the certificate is not recognised then a dialog box is used to warn the user. This behaviour is controlled by properties held in the browser. ActiveX controls have been published on the Internet which close down machines, format hard disks and install viruses. Relying on users to determine which ActiveX controls are safe to use is a dangerous strategy. History teaches that users quickly become bored with "warning" dialog boxes that appear very frequently. Users either stop paying attention to them or find a way to prevent their appearance.

Trojan Horses

320. A Trojan horse is a malicious, security-breaking program that is disguised as something benign, such as a directory lister, archiver, game, or, in one notorious 1990 case on the Macintosh, a program to find and destroy viruses. More recent Trojan horses have used the macro facilities within MS Word, allowing them to be constructed using an intuitive programming language.

Viruses

321. A virus is a program that searches out other programs and infects them by embedding a copy of itself in them. When these programs are executed, the embedded virus is executed too, thus propagating the infection. This normally happens invisibly to the user. A virus cannot infect other computers without assistance. More recent viruses have been based on document macro capabilities. The virus may do nothing but propagate itself and then allow the program to run normally. Usually, however, after propagating silently for a while, it may write messages on the terminal or play tricks with the display (some viruses include sophisticated display hacks). The damage a virus does, once it gains control of a machine, is limited only by the good will and competence of its author. Nothing prevents a virus from causing irreversible damage, such as destroying all of a user's files.

322. Viruses are a serious problem, especially among PC and Macintosh users (the lack of security on these machines enables viruses to spread easily, even infecting the operating system). The production of special anti-virus software has become an industry, and a number of exaggerated media reports have caused outbreaks of near hysteria among users; some users blame everything that doesn't work as expected on virus attacks. In some cases, the fear of a virus infection can waste as many resources as an actual occurrence, since users may be driven to adopt extreme precautions and broadcast unnecessary warnings.

Documents

323. A final vulnerability exists when exporting complex documents. It is often impossible to fully review the contents of a computer file, so that

classified text can escape without proper sanitisation. As an example, most users of MS Word use the *fastsave* option. When text is deleted from a document and the document is then saved again, Word does not delete the text: it merely inserts a note not to display the text. The text can be recovered using another program which cannot recognise Word's instruction to itself.

CHAPTER 4

INTERNET SECURITY MEASURES

Introduction

401. A number of off-the-shelf technology products are available to mitigate the vulnerabilities outlined in the preceding chapter. These countermeasures can be grouped under the following broad headings: integrity and authentication, data confidentiality, intrusion detection, security management, and firewalls and guards. Departments can use judicious combinations of these countermeasures to create corporate "Intranets" that are connected to the Internet while being reasonably protected from its threats.

402. There are a large number of current and emerging Internet security solutions and their quality ranges widely. Some solutions are written by university students and amateurs, while others are developed by professional programmers in commercial organisations. In neither case should any degree of qua1ity be assumed. A poor quality security countermeasure may be worse than no security at all, as it will give users a false sense of security. While Internet products will not, in general, be of sufficient value to warrant formal evaluation, the GCSB can provide informal assessments of Internet products on request. Departments should in all cases contact the GCSB for advice on Internet security products.

Network Authentication

403. Kerberos is a publicly available security architecture that can provide reliable authentication and data integrity over open networks such as the Internet. Kerberos is a secret key authentication system that involves a central database keeping copies of the secret keys of all users. It typically uses DES for encryption and authentication, and allows entities to communicate over networks and to prove their identity to each other using the concept of time-limited *tickets,* containing relevant security information, which are issued from the central management point.

Network Encryption

404. Session encryption is used to provide security services between the network components participating in the session, eg a browser and a web

server. When employing session layer encryption, all data transmitted between the agents during a session will be encrypted but will be in the clear when presented to the applications. Session layer authentication provides authentication of the application programs used at either end of the session. The Secure Sockets Layer (SSL) module, developed by Netscape and widely deployed in Web browsers and servers, is an example of a mechanism providing session layer encryption services. Departments planning to deploy SSL should contact the GCSB to obtain advice on approved SSL products.

405. IP layer encryption is used by some commercially available encryptors although these tend to be relatively expensive. A pair of these encryptors, each acting as a gateway between a private network and the Internet, will route all traffic to each other via Internet routers, which will see cleartext IP headers but encrypted payloads. The address of the actual destination system can be hidden in the encrypted payload, so that the Internet routers only see the addresses of the encrypting gateways. Firewalls that provide Virtual Private Networks (VPNs) use this approach. The Internet Protocol of TCP/IP has a security extension (IPSEC) that permits flexible use of encryption at the IP layer.

Protection of Passwords

406. One of the significant vulnerabilities for computers connected to the Internet is the system password file. If a system password is compromised, then the computer may be wide open to a variety of attacks. This can be countered by using unguessable and non-dictionary passwords combined with mechanisms that ensure passwords are never passed over the Internet in the clear. Alternatively, authentication mechanisms such as the RACAL WatchWord and SecureId permit the use of one-time passwords over the network and thus the disclosure of the password does not compromise the system. The S/KEY standard (RFC 1760) defines a software based one-time password system for use across the Internet. Departments should contact the GCSB for advice on specific S/KEY implementations.

Protection of Workstations

407. There are a number of ways in which a user connected to the Internet by dial-up or dedicated access can be attacked. The two major problems departmental staff will strike are likely to be the use of *Magic Cookies* by Internet Web Servers, and attacks emanating from active Web components such as Java or ActiveX.

408. Magic cookies are files created by Web servers on the hard disks of users who connect to them. These files are used to store information relating to use of that Web Server, in order to provide the Web server with customer-specific information at subsequent accesses. However, the fact that the Web server can write to a client workstation opens a ready channel for a variety of

attacks. Departments should ensure all staff terminals with Web access to the Internet have their browsers configured to reject cookies.

409. The ability of active components to operate on client workstations is a significant risk for departments. The ability of applets to be carried on the HTTP stream means that traditional firewall controls are of little use in protecting the firewalled systems, and additional applet blocking strategies need to be implemented. A basic blocking strategy is to check for the CA-FE-BA-BE hexadecimal string (the applet tag) in the downloaded stream. However, the best solution is for departments' staff to ensure that their terminals are configured to reject Java applets. If departments wish to field active component Webs, then the applets should be digitally signed and verified in user browsers.

Protection of Electronic Mail

410. The primary method of protection for electronic mail is the use of public key encryption. Internet mail security solutions typically provide message confidentiality through symmetric encryption, key exchange through public key encryption, and digital signatures for authentication, integrity, and non-repudiation also through public key encryption. The use of public key encryption requires infrastructure support to ensure the trustworthiness of the cryptographic processes, and this is described in detail in Chapter 5.

Protection of Web Servers

411. Web servers are vulnerable to remote attack typically through unauthorised access to the host computer in order to maliciously modify the Web script pages. The first line of defence is to ensure the server operating system's security controls are properly configured and that all known operating system problems have been addressed. Following this, a second line of defence is to provide integrity verification on the HyperText Markup Language (HTML) files which make up the Web The GCSB can provide detailed advice on Web server integrity checking utilities on request.

Interconnection

412. A firewall is a system (one or more pieces of hardware and software) that acts as a barrier between two network segments. Typically, one segment is a LAN and the other segment the Internet. A firewall enforces the LAN security policy against access from the Internet. Firewalls come in three types: packet filtering, which is usually implemented with a screening router, circuit gateways, and application gateways, which are usually implemented with a dual-homed host and proxy servers. In the dual-homed implementation, the firewall uses two separate network connections and does not allow data to pass directly between the two. Guards provide a similar controlled connection

between networks of differing sensitivity levels, but with less application functionality.

413. When operating as an application gateway, the firewall will examine specific application protocols to decide whether connections are permissible. The range of supported application protocols varies but most firewalls examine such popular ones as Telnet, HTTP or FTP. Configured as a proxy server, the firewall interacts over the Internet on behalf of internal users, making it appear that all outgoing traffic emanates from the firewall and shielding internal node addresses from the Internet.

414. A full description of firewall technology is provided in Chapter 6.

Intrusion Detection

415. It is prudent to assume that attacks will occur and some will succeed in penetrating the connection security mechanisms. It is important, therefore, that a means to detect and respond to these attacks is installed to protect critical information services. Commercial and government intrusion detection products exist that detect and provide alerts to known attack methods.

416. Intrusion detection systems can be categorised into two classes: those that are network based; and those that are host based.

a. Network based intrusion detection systems examine every packet as it enters the network for strings matching known attack methodologies. When a possible attack is discovered, the system administrator will be alerted and must take action to prevent further intrusions from occurring. These tools are similar to virus checkers, in that, as new attacks are discovered the tool must be modified to allow the discovery of these potential new attacks. There are a range of commercial intrusion detection systems, including Net Ranger, NetProwler, Network Flight Recorder, and RealSecure. Tripwire is a freely available and widely used utility that can alert a system administrator to changes in file systems that may signal an intrusion. The GCSB can provide further details of current network intrusion detection systems on request.

b. Host based intrusion detection systems are required to be run on each individual host within a network. These systems run as background tasks in host computers and detect probes of host ports, password guessing and other known attack methods, which are captured as part of host auditing features.

417. There is some overlap between host based and network based string matching capabilities and running the two in concert will provide, in some cases, the same detection alert. However, intrusion detection tools do not detect 100% of potential attacks and will often have a significant level of false alarms. Departments should, as a minimum countermeasure, install a network based intrusion detection system behind the Internet connection firewall.

Security Management Tools

418. A number of tools exist which can be used by both attackers and system managers to test the security of a system. Some of the more widely available ones are listed below:

419. **SATAN.** The Security Analysis Tool for Auditing Networks (SATAN) is a network vulnerability toolkit, using a web front-end. In its simplest (and default) mode, it gathers as much information about remote hosts and networks as possible by examining such network services as finger, NFS, NIS, ftp and tftp, rexd, and other services. The information gathered includes the presence of various network information services as well as potential security flaws - usually in the form of incorrectly configured network services, well-known bugs in system or network utilities, or inadequate security policies. It can then either report on this data or use a simple rule-based system to investigate any potential security problems. SATAN output can be examined using a web browser. While the program is primarily geared towards analysing the security implications of the results, a great deal of general network information can be gained when using the tool - network topology, network services running, and the types of hardware and software being used on the network

420. **ISS**. Internet Security Scanner (ISS) can scan a host or a network to test for a common set of security flaws and errors in configuration. It is designed to carry out a simple vulnerability test on a network of computers in order to highlight systems that are vulnerable and may be used to gain access to more secure and important systems.

421. **tiger.** The tiger software suite is produced by the Texas A & M University. It is a collection of Bourne shell scripts, C programs and data files which are used to perform a security audit of UNIX systems. It has pre-defined configuration databases for AIX 3.x as well as many other systems. tiger has one primary goal: to report ways in which root can be compromised. The primary assumption made is that any UID other than 0 can be obtained and that any GID can be obtained by unauthorised persons. tiger also checks other means of gaining root access, e.g. cron, inetd and setuid executables and whether any user, other than root, can alter any of the configuration files associated with these utilities. Specific checks are performed to see if anything in the root executable path can be modified by a normal user.

422. **crack.** crack is a dictionary based password guessing tool. UNIX passwords are encrypted using the DES algorithm and so, in themselves, are secure. However, when compared with a set of dictionary words which have been similarly encrypted, it is simple to identify those passwords which appear in the dictionary.

423. **snoop.** snoop is a promiscuous-mode IP packet sniffer. The Sun operating system Solaris provides to administrators a command *snoop* for the capture and inspection of ethernet packets. By default it uses both the network interface and the streams buffer modules to capture packets, then displays a single line summary of each packet seen. It can also provide a far more detailed packet description, including information from the ethernet layer

upwards to the top layer. This utility can also be used to monitor certain protocols and to sniff for authentication data, for example the rexec service.

CHAPTER 5

PUBLIC KEY INFRASTRUCTURE

Introduction

501. Encryption in public data networks is typically implemented using asymmetric, or public key, cryptography. In this form of cryptography, a pair of associated cryptographic keys are allocated to a user. One of the cryptographic keys is made public (the public key), and the other is known only to the owner (the private key). By making the public keys widely accessible in the form of security certificates, many of the problems of key management are avoided, in particular the need to preposition cryptographic keys before a secure dialogue can take place. However, there is still a need to issue keys and the tokens that hold them, and provide a method of promulgating advice regarding compromised keys.

502. A Public Key Infrastructure (PKI) is a supporting public key management system for the deployment of public key encryption technology throughout an organisation or community of interest. Such an infrastructure will typically comprise a key generation system, a Root Certification Authority (Root CA) which keys for subordinate authorities to operate, a number of subordinate Certification Authorities (CAs) which provide keys to users, electronic directories to hold security certificates, and Local Registration Authorities (LRAs) to provide access to the certification services on the CAs. The functionality of a certification authority is detailed in the ISO 11770-3 standard, and also in the draft Internet PKI standard (PKIX).

503. Contemporary PKIs require that users be allocated two key pairs, one for applying confidentiality in key exchanges and the other for the generation of digital signatures. The public keys are held in accessible directories, and are protected from unauthorised modification by encapsulation in the form of X.509 digital certificates. Private keys are typically held in some form of security token such as a PCMCIA card or diskette.

504. The public keys produced for use in exchanging encryption keys or verifying digital signatures are typically encapsulated into X.509 certificates. These certificates are then stored on the departmental directory server and, to the extent both possible and appropriate, replicated onto other directories throughout Government. In addition, each user will maintain a local disk cache of commonly used certificates to avoid unnecessary network overheads.

CHAPTER 6

FIREWALLS AND GUARDS

Introduction

601. The term 'firewall' is a generic term covering a range of different types of devices providing a secure connection between an internal LAN and an external network such as the Internet. Firewalls in their various incarnations are an important security technology for the Internet. This chapter describes the basic firewall architectures.

Screening Router

602. A screening router is a basic component of most firewall architectures, and usually consists of a commercial router. In some cases routing can be host-based, particularly on hosts using the Unix operating system. Screening routers filter the packets passing between the network connections in accordance with a previously defined routing table. Filtering is usually done on IP packets based on some, or all of the following fields:

- source IP address,
- destination IP address,
- TCP/UDP source port, and
- TCP/UDP destination port.

603. Some routers are able to distinguish the network interface on which a packet arrives, and use this information to decide how the packet should be filtered This is particularly useful when traffic needs to be segmented from specific networks, and in eliminating IP source address spoofing. Packets which arrive at the external interface are known as *inbound packets,* while packets arriving at the internal interface are known as *outbound packets.*

604. Due to the inexpensive nature of screening routers, they have been used in many networks as the sole component of the firewall architecture. Usually there are direct communication paths between multiple hosts on the internal LAN and the Internet. In normal operation, the risks to the internal network are proportional to the number of servers it hosts and the number of connections to the external network. As the number of hosts and connections grows it becomes impossible to identify all possible threats should the router be compromised.

605. Routers are generally used to block connections from or to specific hosts or networks, and to block connections to specific ports. The ability to filter on TCP/UDP ports adds considerable flexibility in defining security policies, as it allows the router to control which TCP services can be accessed e.g. telnet, ftp, smtp, finger etc. Filter rules are defined using a table of conditions and actions which are applied to each packet until a decision to route or drop is reached. If a packet meets all of the conditions specified in the row of the table, the action specified in that row is carried out. Some systems apply the rules in a systematic manner from first to last, while others enforce an order based on the criteria in the rules, such as source and destination address.

Dual Homed Gateway

606. The dual-homed gateway is a common and easily implemented application level firewall. A dual homed gateway is a host machine which has two network connection ports; one connected to the external network and one connected to the internal network. With IP forwarding disabled a complete block of traffic between the two networks is ensured A dual-homed gateway may also perform the same packet level functions as the screening router.

607. There are two ways a user can access the external network via the internal network. The first is by direct logon to the dual-homed gateway. This is not advisable as it makes the dual-homed gateway directly vulnerable to password cracking, and provides access directly to the firewall through software vulnerabilities, such as bugs or compilers being present on the host. The second way for a connection to be made is through the application layer using a proxy server. A proxy server is an application which routes IP traffic from one port to another. Such an application can provide user authentication, auditing, and logging facilities. These features are a great improvement over screening routers which generally provide no more than rudimentary facilities.

608. The problem with using proxy servers is that they usually have to be written for each service that is offered However, basic proxy servers for standard TCP/IP services, such as Telnet, FTP, WWW, etc., are generally available for the UNIX environment.

Screened Host Gateway

609. The screened host gateway is implemented using a screening router and a bastion host. It is one of the most popular firewall architectures. The bastion host is usually placed on the internal network, with the screening router configured such that the bastion host is the only machine reachable from the Internet. To restrict Internet access further the screening router is generally configured to block all traffic not destined to specifically authorised ports on the bastion host. This has the effect of controlling the number of available services.

610. There are some major benefits in the use of screened host gateways. These include reduction of router programming complexity, and improved connectivity for local users. As all traffic is passed through one single point i.e. the bastion host, the rules for configuring the router table need only consider the bastion host's IP address. All other packets arriving at the inbound or outbound ports of the screening router can be discarded, which greatly simplifies the required packet filter rules.

611. If the internal network is a virtual local area network configuration with no sub-nets or additional routing, then the screened host gateway can be implemented without changes to the original LAN. Users then have the ability

to connect directly through the bastion host to the external network without excessive routing overhead.

612. The zone of risk in a screened host gateway incorporates only the screening router and bastion host. The security of this firewall architecture is determined by the accuracy of the packet filter rules in relation to the security policy, and the level of assurance regarding the software running on the bastion host. If an attacker gains entry to the bastion host then the threats to the internal network are similar to those of the dual homed gateway.

613. There is a potential problem with the architecture of the screened host gateway due to the reliance placed upon the screening router to control the traffic flow to and from the bastion host. If the screening router is compromised, either through misconfiguration of the packet filtering rules or through an attacker gaining access to the screening router via a proprietary maintenance account, then the entire internal network is at risk. In effect, compromising the screening router effectively subverts the bastion host. Once an attacker has control of the screening router, any packets can be sent to hosts on the internal network.

614. A more secure implementation is to use a screening router connected to a dual homed gateway. This architecture ensures that the bastion host is not circumvented if the screening router is compromised. The attacker has to overcome the dual homed gateway before the internal network is at risk. This architecture, however, offers no improvement if an attacker is able to enter through the screening router and compromise the bastion host directly.

Screened Sub-nets

615. A screened sub-net firewall architecture consists of an isolated network known as the *exterior network,* positioned between the external and internal networks. This configuration allows non-critical hosts, such as WWW servers and anonymous FTP sites, to be placed on the exterior network. The advantage of removing these servers from the internal network is realised when one is compromised. As they have no connection with the internal network a compromise does not affect the safety of the internal network. The servers also benefit from the protection afforded by the external screening router. Bastion hosts are placed on the exterior network to provide interactive terminal sessions, or application level firewalls. The screened sub-net is generally considered to be the most secure firewall architecture.

616. The bastion host provides the sole point of access to machines on the internal network, and forces all services through the firewall to be provided by application gateways. Protecting the bastion host are two screening routers, one between the external network and sub-net (known as the external router), the other between the sub-net and internal network (known as the internal router). Therefore the zone of risk for this configuration consists of only the two routers, and the bastion host, as well as any other hosts placed on the sub-net.

617. The strength of this firewall architecture comes from the fact that an attacker must subvert the external router, followed by the bastion host, and finally the internal router. If the screening routers are configured so they cannot be managed remotely from the network, then subverting this firewall architecture without setting off alarms and appearing in audit logs would be very difficult.

618. As with the screened host gateway, if the screening routers can be directly compromised by logging into them and reconfiguring their routing tables then the bastion host can be circumvented and the internal network put at risk.

619. A drawback with this configuration is the extra level of complexity added to the definition of packet filter rules, especially if there are hosts on the sub-net other than the bastion host. The overall threats to this firewall architecture are the same as those described for the dual homed and screened host gateways.

Guards

620. Guards provide a controlled connection for specific protocol messages between networks of differing sensitivity levels. For example, the US Standard Mail Guard (SMG) enables users on SECRET networks to send and receive unclassif~ed e-mail without attachments. It is installed between a high-level network and a low-level network, such as the Internet, and ensures that no message traffic from the high-level network that is labelled SECRET passes through to the low-level network unencrypted. It also ensures that no attachments, which may contain Trojan Horses, can pass from the low-level network to the high-level network. Other forms of guards can automatically downgrade highly formatted data.

Gateway Accreditation Guidelines

621. Firewalls and guards provide adequate security only when correctly configured, and an accreditation regime should be adopted to ensure the initial and ongoing assurance of the system. This section details the specific areas of concern that must be addressed when accrediting a gateway between unclassified but sensitive local systems and public data networks such as the Internet. Accreditation is a formal process which verifies that a security solution has been adequately implemented.

622. Gateway accreditation focuses on four distinct areas:

- a review of the gateway access policy;
- a review of the gateway design;
- carrying out a vulnerability test of the gateway installation and configuration; and
- a review of the gateway management procedures.

623. **Access Policy.** A realistic access policy is one which balances the known risks to a local network against the need for reasonable user access to external resources. The policy should be based on the principle of denying all services except for those that have been specifically permitted. This policy should be fully supported by the departmental executive and management staff. The accreditation review of the access policy will examine the following items:

a. which IP packets should be filtered, and the criteria for filtering;

b. which services are permitted;

c. the conditions under which remote network connections will be permitted;

d. the functionality and assurance requirements for authenticating remote users;

e. the policy on collection, distribution, and transmission of e-mail;

f. the conditions under which files can be transferred in and out of the local network;

g. the conditions on provision of restricted and publicly accessible Web servers;

h. the departmental controls on any third party involvement in gateway operations;

i. the policy on intrusion detection and incident response, and

j. configuration management policy.

624. **Design.** The gateway design will define the rules used to implement the network services access policy. This policy must be designed around the known capabilities and limitations of the specific gateway product or products, and the known vulnerabilities of TCP/IP. Of particular importance to the design review is the architectural location of components of the gateway. In particular, publicly accessible hosts should not be on the internal, sensitive network but should be on a separate screened sub-domain. This will ensure that remote users do not penetrate the bastion host firewall. The gateway accreditation will review the design of:

a. facilities to detect and warn of potential intrusion;

b. facilities to detect and warn of unauthorised system and configuration changes;

c. facilities to detect malicious content (viruses, trojan horses) in downloaded files and mail messages;

d. filtering and routing rules;

e. use of proxy services;

f. dial-in access capabilities;

g. authentication mechanisms;

h. remote user warning and liability disclaimer notices;

i. usage monitoring mechanisms; and

j. encryption and key management services.

625. **Installation and Configuration.** The physical location of gateway components will be reviewed against the potential for damage, theft, destruction, or malicious modification. A vulnerability test should be carried out against the following requirements:

a. perimeter elements (windows, walls, doors, floors, and ceilings) should be of such a design to protect against removal or displacement which might allow unauthorised entry without obvious damage;

b. doors should be solid core and fitted with an approved lock and with key or card entry access controls during working hours;

c. windows should be fixed or fitted with an approved lock for use when the area is unoccupied;

d. the gateway configuration will be tested with scanning tools to ensure that it does not have any known vulnerabilities and that the gateway installation conforms to the gateway access policy; and

e. the internal network will be reviewed to ensure that there are no back-door access methods such as wide area network connections or modems fitted to workstations on the internal network.

626. **Management Procedures.** The establishment of gateway policies and the installation of the system provides a foundation for secure interconnectivity, but this needs to be supported by an ongoing security management regime which will ensure the policy and design principles are maintained for the life of the system. The management processes that will be reviewed as part of gateway accreditation are ownership and accountability, configuration management, activity logging and monitoring, and incident response. The review will in particular look at the operation of:

a. the skill levels of gateway administrators;

b. configuration control board processes;

c. access control processes;

d. intrusion detection processes;

e. security review and audit processes;

f. change control processes; and

g. logging systems and audit reduction tools.

Firewall Limitations

627. A firewall architecture is a powerful tool for network security, but there are things it cannot do. Packet level firewalls can restrict the TCP and IP layers, while the application level firewalls can restrict access to the services used. Firewall architectures rely on the correct implementation and operation of these components. Further, a firewall can only defend against known security threats, and will always be vulnerable to new ones.

628. Firewalls cannot protect against installation errors or software bugs, nor can they protect against malicious programs imported through ftp (although some firewall products are beginning to offer additional protection against viruses, e.g., Borderware has an add on unit which uses McAfee's virus scanner to scan all files which are imported through the firewall).

CHAPTER 7

RECOMMENDED ARCHITECTURES FOR INTERNET USE

General

701. There are two types of connection to the Internet: a dedicated connection from a router on a local area network (or, less commonly, directly from a host computer); and a dial up connection from a computer modem. In the latter case, the computer may be a stand alone system dedicated to Internet use, or it may be a workstation on a network The recommended architectures detailed in this chapter deal with connectivity of systems handling only unclassified information. Details of connectivity for systems handling classified information are provided in Chapter 8.

Stand Alone Dial Up Access

702. The safest architecture for Internet connectivity consists of a stand alone computer and modem, used only for accessing the Internet, typically operating as a 'public kiosk' for general use.

703. The normal management procedures of system integrity checking (such as anti-virus software) and data backups provide the controls necessary to recover the system in the event of catastrophic attack.

704. Application encryption services using a public-key based product will be required for the protection of private or sensitive information being transmitted across the Internet. Departments should use public key systems which are evaluated and comply with the SSC PKI standards.

705. Network encryption services using products such as the Secure Sockets Layer (SSL) may be required if any form of sensitive information, such as credit card information, is being exchanged between the local browser and an Internet Web server. Departments should use full 128-bit SSL encryption rather than the weaker 40-bit encryption often provided as default with browsers.

706. Java/ActiveX features should be disabled to avoid malicious attack by hostile applets if such functionality is not critical to the use of the system.

707. All software imported through downloading from the Internet should be carefully scanned for virus infection. Word documents should also be scanned for document-based virus attack.

Workstation Dial Up Access

708. The security requirements of a LAN workstation which is used for dial up connectivity (Figure 7.1) through an Internet Service Provider are the same as for stand-alone dial up access, but with the more rigorous requirement that Java/ActiveX controls must be turned off.

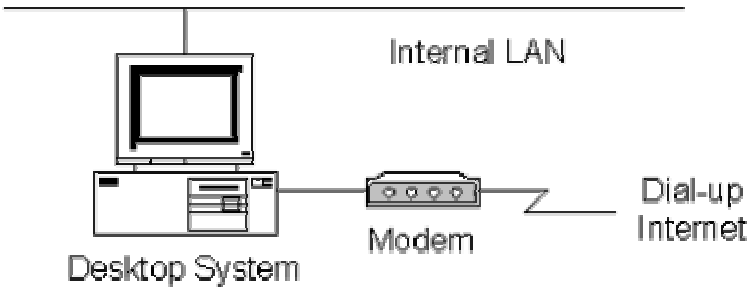

Figure 7.1: Workstation Access to the Internet

Dedicated Line Access

709. Where a department wishes to offer Internet access to a departmenta1 Web server, a dedicated link between the departmental LAN and an existing Internet Service Provider will be required to providc a permanent connection (Figure 7.2). Such connections must be dynamically monitored in some way. While routers are commonly used, on their own they provide inadequate

protection and fully configured firewalls are necessary to provide adequate protection.

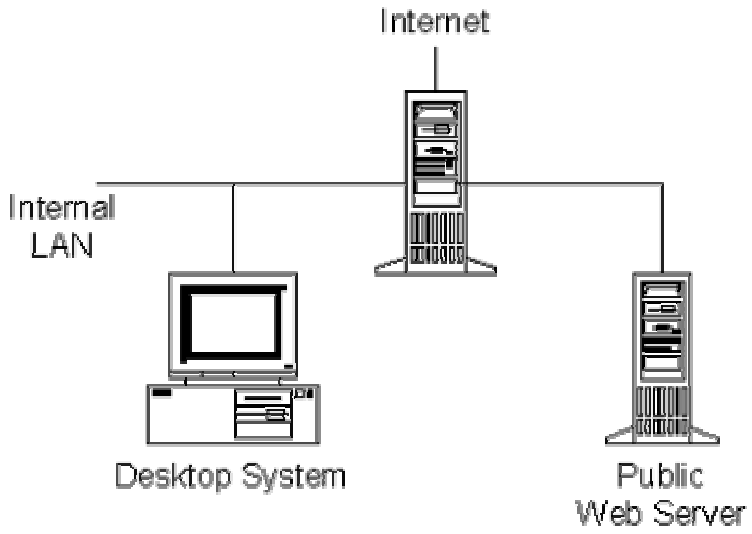

Figure 7.2: Dedicated Connection to the Internet

710. An approved firewall should be used, and the installation should be accredited by departmental security authorities to ensure that the Internet gateway conforms to the established policy for access to departmental resources.

711. Departmental hosts which are made available for restricted access through the Internet will, in addition to the above countermeasures, require some form of user authentication. This can be done in a number of ways:

a. password schemes which are simple to implement but have been shown to be easily defeated;

b. more advanced one-time authentication systems such as the RACAL Watchword or SecureId challenge-response schemes, which not only provide much more effective authentication but also counter the problem of network password sniffing; and

c. cryptographic authentication using digital signature verification techniques to authenticate users.

CHAPTER 8

CLASSIFIED INFORMATION AND PUBLIC DATA NETWORKS

General

801. Systems and LANs handling classified information are not to be connected to public data (unclassified) networks, and particularly not high threat networks such as the Internet. Current interconnectivity and filtering technologies, such as routers, mailguards, and firewalls, provide insufficient assurance for the avoidance of accidental spillage of classified information and its protection against attacks from the unclassified domain. Air gap procedures must be used when transferring information from an external network into the classifed LAN, and rigorous manual review procedures are to be used to protect against inadvertent spillage of classified information when transferring information through an air gap out of a classified LAN.

802. Not withstanding the above, there are some high grade "data diodes" which provide sufficient assurance to allow connectivity, using specifcally defined protocols, between classified and unclassified networks Certain high grade encryption devices may also be uscd to provide protection of classified information that may then be transmitted across public data networks Further information on these devices can be obtained on specific request from the GCSB.

ANNEX A

OVERVIEW OF TCP/IP

A.1. The Transmission Control Protocol/Internet Protocol (TCP/IP) is the collection of communications protocols which are used on the Internet. This set of protocols is the most widely accepted form of networking between computers.

A.2. Communication between heterogeneous computer systems is a complex and diverse problem, which cannot be accomplished by a single all-encompassing protocol. TCP/IP splits the task of communicating into four layers: application, transport, network, and link. Each layer addresses a different functional requirement:

a. **Link Layer** - contains the protocols which provide access to a communication network. This usually consists of the host operating system's device driver, and the network interface card. The link layer handles all aspects of physical interfacing with the communications medium. The function of these protocols is to route data between hosts attached to the same network. Other services may also be provided, such as flow and error control.

b. **Network Layer** - provides the functionality which allows data to traverse multiple networks between hosts. Therefore, the network layer provides the inter-network routing function. At this layer there are three different protocols; Internet Protocol (IP), Internet Control Message Protocol (ICMP), and the Internet Group Message Protocol (IGMP). The protocols at this layer are usually implemented in hosts and network devices, such as routers, which route data between different networks.

c. **Transport Layer** - provides functionality which enables data to be delivered between two processes on different host computers. At this layer there are two different protocols which provide this functionality; Transmission Control Protocol (TCP) and User Datagram Protocol (UDP).

d. **Application Layer** - handles the details of the applications.

Internet Protocol

A.3. The IP protocol is an unreliable, connectionless datagram service. There are no guarantees that a given IP datagram wil1 reach its destination. In fact, there is no guarantee that the datagram received is the same as was sent. The official IP specification can be found in RFC 791.

A.4. An IP datagram can be a maximum of 65535 bytes long, which is limited by the two-byte packet length field in the IP header. In practice, few datagrams of this size are sent, as most link layer protocols support physical frame lengths of a few thousand bytes. If an IP packet has to be broken down because it is too large for the underlying link layer this is known as *fragmentation,* the reconstruction of fragments at the destination is known as *re-assembly.* The largest amount of encapsulated data a network interface can transmit is called the maximum transmission unit (MTU). For example, Ethernet supports an MTU of 1500 bytes.

A.5. For packets with a destination on the same physical network as the original sender, the MTU is known by the sending host because it is a parameter that is a part of every network interface specification. A transport layer protocol can then use the MTU to limit the size of the message it passes to IP, therefore an IP packet will never require fragmentation. However, when a transport layer protocol builds a message destined for a host on a different network it has no way of knowing the route, nor the MTU of each physical network the packet will traverse to its destination. In this case a default MTU of 576 bytes is used which supports a 512 byte message, a 20 byte TCP header, and a 20 byte IP header. Most link layer protocols support an MTU of at least 576 bytes.

A.6. The IP datagram contains the following fields:

a. **4-bit Version.** The version number of IP; currently version 4 (IPv4) is deployed.

b. **4-bit Header Length.** The length of the IP header in 32-bit words. The header is always padded out to a multiple of 32-bit words.

c. **8-bit Type Of Service (TOS).** The type of service or priority for this packet. Type of service processing is not frequently used, therefore the default value of 0 is generally used.

d. **16-bit Total Length.** The length of the IP packet (including the header) in bytes.

e. **16-bit Identification, 3-bit Flags, and 13-bit Fragment Offset.** This field is used for fragmentation and re-assembly control.

f. **8-bit Time To Live (TTL).** TTL is the maximum time in seconds, time-to-live, that the packet may exist. This field is decremented by at least 1 each time the IP header is processed by a router or host. Unless the packet is queued in a buffer for a long period of time, this field actually indicates the maximum number of intermediate routers a packet may cross before it is dropped. When this field reaches 0, it must be dropped unconditionally by the IP. This feature prevents a packet from looping around the network forever because of a routing error.

g. **8-bit Protocol** This field indicates the type of protocol message encapsulated within the IP header. For example, the protocol field value is 6 for TCP and 17 for UDP.

h. **16-bit Header Checksum.** This checksum covers only the IP header. The checksum is constructed by taking the 16-bit 1's complement of all the 16-bit words in the header. This field allows the header to be checked for errors which may have occurred in transmission.

i. **32-bit Source IP Address.** The IP address of the interface from which the packet originated.

j. **32-bit Destination IP Address.** The IP address of the packet's final network interface destination. As each IP packet contains its source and destination address it can be routed independently to its destination.

k. **Options.** This field can contain various IP options, although most IP packets do not. Options include the following:

source routing enables an IP packet's route to be specifically controlled. Is used in source routing attacks.

route recording records the route the packet takes in the options field.

time-stamping adds a time-stamp by each intermediate router.

security can contain seldom used security options.

l. **Padding.** This pads the IP header to an even 4-byte boundary. This is occasionally needed because not all IP options are even multiples of 32 bits.

A.7. An IP packet can travel through many routers or hosts before it reaches its destination. On receipt of an IP packet, the router looks at its IP destination address and compares this with its routing table; returning a result which decides which port the IP datagram will be sent out on. Routing tables are constantly updated to reflect the status of the various interconnected networks. It is not uncommon for IP datagrams which are part of the same conversation to take different paths before arriving at their destination. It is the job of higher layers of the protocol stack e.g. TCP, to reassemble and resequence the IP datagrams.

A.8. The use of dynamic paths between source and destination points, and the ease at which they can be manipulated, means that any plain text sent across the Internet is, in essence, available for anyone to see.

IP Addresses

A.9. IP addresses are 32 bits long, and divided into two parts; the network and the host address. The boundary is dependant on the first one to four high-order bits, and indicates which network addressing scheme is being used.

Network Class	High-order Bits	Network	Host	Number of Addresses
A	0	7	24	16,777,214
B	10	14	16	65,534
C	110	21	8	254
D	1110	Multicast group		268,435,456
E	1111	(Experimental use)		

Figure A.1: IP Address Formats

A.10. The host part of the IP address is usually broken into a sub-net and host address. Sub-nets are used to route IP datagrams within an organisational network domain. It is up to the organisation to determine the number of bits used for the sub-net. For example it is common to divide a Class B network into 254 sub-networks.

A.11. IP addresses are not usually used in their numeric formats, instead they are translated into a more human readable domain name, such as *gcsb.govt.nz,* instead of <129.15.4.1>. This is done by the Domain Name System which, in essence, is a distributed database.

A.12. There is a significant amount of address space wastage caused by sub-net partitioning in IPv4. For example, if an organisation has a single class C network address, they have a possible 255 host addresses. However they may only have 50 hosts on their network, thus 205 host addresses have been wasted If it were not for sub-nets, the current 32-bit IP addressing scheme could accommodate a possible 232 host addresses. This problem has been addressed in the IPv6 standard, by using 64-bit IP addresses.

Address Resolution Protocol (ARP)

A.13. In most cases IP datagrams are sent over data links such as Ethernet or Token Ring. However, these devices have their own addressing schemes, e.g. 48 bits in the case of Ethernet. When Ethernet frames are sent between

hosts on a LAN, it is the 48 bit Ethernet address that determines which interface will receive it. The Address Resolution Protocol (ARP) is used to provide dynamic mapping between the 32 bit IP address and the 48 bit Ethernet address. This mapping usually requires a table lookup.

A.14. The ARP cache, maintained by each host, is an important aspect of efficient operation of the ARP. The cache contains an up-to-date table of mappings from IP to Ethernet addresses. Usually the expiration time for mappings in the cache is twenty minutes. If an ARP mapping in the cache has expired, or the required one is not found, then ARP sends out a broadcast packet containing the desired IP address. The destination host replies with a packet containing the IP and Ethernet address pair. This is placed in the ARP cache, reducing the amount of ARP traffic that would otherwise be required.

A.15. ARP is not a secure protocol. If an untrusted host has access to the LAN, it can broadcast phoney ARP messages redirecting all traffic to itself, in order to impersonate another machine or modify its data streams.

Transmission Control Protocol

A.16. The Transmission Control Protocol (TCP) is a connection oriented protocol, which provides reliable virtual circuits. TCP is designed to work in a very general environment of interconnected networks. The TCP interfaces on one side to user or application processes and on the other side to a lower level protocol such as Internet Protocol.

A.17. TCP, unlike IP, is a reliable service. Lost or damaged packets are retransmitted, and received packets are reassembled to match the original transmission sequence. Each packet contains a sequence number which maintains the original transmission order. All packets except the first TCP packet, which is used to initiate the session, contain an acknowledgment number corresponding to the sequence number of the last successfully received packet.

A.18. TCP supports the multiplexing of multiple circuits over a single channel. Every TCP packet includes the originating host address (orig.host) and port number (orig.port), as well as the destination host address (dest.host) and port number (dest.port). This vector (or 4-tuple), <orig.host, orig.port, dest.host, dest.port>, uniquely identifies the circuit being used for the communication.

A.I9. Communication over the Internet is usually in line with the Client/Server model. Servers generally listen to ports numbered below 1024, which are known as "well known ports" (see Appendix A). These ports offer standard TCP/IP services such as Telnet, FTP, SMTP, etc. A server continually listens to its associated port waiting for a Client process to initiate a connection. Port numbers for Client processes are generally allocated "high", i.e. above 1023. However, it is unwise to trust services with regard to port number alone, as the allocation of port numbers is a convention, not an enforced scheme.

A.20. When two processes wish to communicate, their TCP's must first establish a connection. Since connections must be established between potentially unreliable hosts over a potentially unreliable communication system, a three-way handshake mechanism with clock-based (or random) sequence numbers is used to prevent packets that get delayed in the network from being delivered later and being misinterpreted as part of the existing conversation.

A.21. However, this three-way handshake presents a vulnerability if an attacker can guess the target's initial sequence number. The problem lies with the predictability of the initial sequence number which allows an attacker to fool the target into thinking that it is communicating with a trusted host. To achieve this, the attacker must have attained a valid login initially. Trusted hosts are authenticated using their source IP address - this applies to Berkeley Unix remote commands such as, *rlogin* and *rsh.* Therefore, if an attacker knows the trusted host's IP address, can correctly guess the initial sequence number, and initiates supporting "r" commands, he or she could execute malicious commands on the target. This is known as a *sequence number attack.*

top

User Datagram Protocol

A.22. The User Datagram Protocol (UDP) provides a datagram mode of packet-switched computer communication and assumes that IP is used as the underlying protocol. UDP provides a procedure for application programs to send messages to other programs with a minimum of protocol overhead The protocol is connectionless, so delivery and duplication protection is not guaranteed. It is well suited for transaction based processes, such as Sun's Remote Procedure Calls.

A.23. UDP tends to behave badly when used to transmit streams of data. As it lacks flow control UDP messages can swamp hosts and routers, causing extensive packet loss. Therefore it can be used for *denial of service* attacks. It is far easier to spoof UDP packets than TCP, as there are no handshaking protocols and packets are unique - there is no notion of sequence number. It is not recommended that the source address be used for authentication. The specification for UDP indicates that it is up to the user interface to specify port numbers and host addresses.

Internet Control Message Protocol

A.24. The Internet Control Message Protocol (ICMP) is an integral part of the IP suite. It is used to communicate error and informational messages to the IP layer or higher layers, such as TCP or UDP. ICMP can be used to inform hosts of a better route to a destination, report problems on a route, or terminate

connections due to network problems. ICMP Messages received on a host are specific to a particular connection, or in relation to a packet sent from the host.

A.25. Some ICMP messages make use of the *code* field, which provides more specific information on the connection's status. The ICMP protocol also forms the basis of a useful program called *Ping,* named after the operation of locating objects by sonar. Ping issues an ICMP echo request message to a host, and expects an ICMP echo reply in return. If it does not receive a reply then this indicates that there could be a problem with the destination host or the intervening network. However, many firewalls are configured to block ICMP messages, therefore Pings to these hosts may not receive replies.

A.26. A popular abuse of ICMP is generation of *denial of service* attacks. It is possible to use several types of ICMP messages, such as *Destination Unreachable* and *Time to live Exceeded, to* reset existing connections. Some older implementations of ICMP do not limit their action to a specific connection, but will tear down all connections between the host and gateway on receipt of these messages.

top

Domain Name Service

A.27. The Domain Name System (DNS) is a distributed database used by TCP/IP applications to map between hostnames and IP addresses. It is also used to determine the destination of mail within an organisation. All hosts connected to the Internet have unique IP addresses, which are used to communicate between one another. However, IP addresses are not easily remembered by humans, so the DNS provides a way of associating an ASCII based identifier to an IP address and mapping between them. No single site on the Internet contains all mapping information. Instead every Internet site, e.g. University, company, etc., maintains its own database and runs a server program that other systems (i.e. clients) across the Internet can query. DNS provides the protocol which enables clients and servers to communicate with each other.

A.28. Applications access the DNS through *resolvers* which on Unix hosts are typically the *gethostbyname()* and *gethostbyaddr()* library functions. The first resolver, gethostbyname, takes a host name and returns its IP address. The IP address returned by querying a DNS is not related to the choice of name for a host. The second resolver, gethostbyaddr, takes an IP address and returns the corresponding host name. The resolver may have to contact more than one DNS to complete the mapping. Normally a resolver will generate a UDP based query to the DNS, which replies with the correct mapping, or returns information on an alternative DNS which can be queried further. TCP can also be used for queries, however this is normally reserved for *zone transfers.* A zone transfer allows backup servers to obtain a full copy of their portion of the DNS name space. This can also be used by attackers to obtain a list of potential targets.

A.29. The Internet, seen through DNS, resembles a hierarchical tree. At the top are the root name servers, which contain information about the contents of the top level domains (TLD), i.e., .net, .com, .edu, .org, .int, .gov, .mil, and the two-letter country codes from ISO-3 166 (.nz, .us, .uk .fi, jp, etc.). Under the top-level domains are the second level domains, such as nist.gov. Further levels can be defined, but are the responsibility of the second level DNSs maintained by the organisation, in this case the nist. Each name reflects its position within the DNS tree, for example, www.nist.gov represents a host computer (www) in the nist domain which is inside the gov top level domain. The domain name for a node is constructed by starting at that node and working up the tree to the root, separating each label with a period. A query about www.nist.gov sent to a primary domain server will be answered (as long as the primary domain server knows about the domain nist.gov) with a pointer (IP address) to nist.gov's name server which holds the information about the host www.nist.gov. The nist.gov DNS will then return the IP address for the host www.nist.gov. At this point a direct connection can be made to www.nist.gov from the querying host.