Data and Network Security
IT-4541

By: Gert_DeLaet,_Gert_Schauwers

Prepared By:

Muhammad Shahid Azeem

Lecturer CS/IT
MCS, M.Phil (CS)

Dedication

All glories praises and gratitude to Almighty Allah Pak, who blessed us with a super, unequalled processor! Brain…

I dedicate all of my efforts to my students who gave me an urge and inspiration to work more.
I also dedicate this piece of effort to my family members who always support me to move on. Especially, my father (Ch. Ali Muhammad) who pushed me ever and makes me to stand at this position today.

Muhammad Shahid Azeem

Course Outline

1. Network Security Overview: Defining Trust, Weaknesses and Vulnerabilities, Responsibilities for Network Security, Security Objectives, the Need for Security, Risk and Vulnerability, TCP/IP Suite Weaknesses, Buffer Overflows, Spoofing Techniques, Social Engineering. [TB1: Ch. 1,2]

2. Understanding Defenses: Digital IDs, Intrusion Detection System, PC Card-Based Solutions, Physical Security, Encrypted Login, Firewalls, Reusable Passwords, Antivirus Software, Encrypted Files, Biometrics. [TB1: Ch. 3]

3. Cryptography: Introduction, Cryptography versus Cryptanalysis, Modern-Day Techniques. [TB1: Ch. 4]

4. Security Policies: Defining a Security Policy, Importance of a Security Policy, Development Process, Incident Handling Process, Security Wheel, Sample Security Policy. [TB1: Ch. 5]

5. Secure Design: Network Design-Principles, Network Design-Methodology, Return on Investment, Physical Security Issues, Switches and Hubs. [TB1: Ch. 6]

6. Web Security: Hardening, Case Study. [TB1: Ch. 7]

7. Router Security: Basic Router Security, Router Security to Protect the Network, CBAC, Case Study. [TB1: Ch. 8]

8. Firewalls: Firewall Basics, Different Types of Firewalls, Enhancements for Firewalls, Placing Filtering Routers and Firewalls. [TB1: Ch. 9]

9. Intrusion Detection System: Introduction to Intrusion Detection, Host-Based IDSs, Network-Based IDSs, IDS Management Communications-Monitoring the Network, Sensor Maintenance, Case Study: Deployment of IDS Sensors in the Organization and Their Typical Placement. [TB1: Ch. 10]

10. Remote Access: AAA Model, AAA Servers, Lock-and-Key Feature, Two-Factor Identification, Case Study: Configuring Secure Remote Access. [TB1: Ch. 11]

11. Virtual Private Networks: Generic Routing Encapsulation Tunnels, IP Security, VPNs with IPSec, Case Study: Remote Access VPN. [TB1: Ch. 12]

12. Public Key Infrastructure: Public Key Distribution, Trusted Third Party, PKI Topology, Enrollment Procedure, Revocation Procedure, Case Study: Creating Your Own CA. [TB1: Ch. 13]

13. Wireless Security: Different WLAN Configurations, What Is a WLAN? How Wireless Works, Risks of Open Wireless Ports, War-Driving and War-Chalking, SAFE WLAN Design Techniques and Considerations, Case Study: Adding Wireless Solutions to a Secure Network. [TB1: Ch. 14]

14. Logging and Auditing: Logging, SYSLOG, Simple Network Management Protocol, Remote Monitoring, Service Assurance Agent, Case Study. [TB1: Ch. 15]

Textbook(s):

Network Security Fundamentals by GertDeLaetand GertSchauwers, Cisco Press; 1st Edition (September 18, 2004). ISBN-10: 1587051672

This page is left blank intentionally

Chapter 01
NETWORK SECURITY OVERVIEW

1.1. Computer Security

The protection afforded to an automated information system in order to attain the applicable objectives of preserving the integrity, availability, and confidentiality of information system resources (includes hardware, software, firmware, information/ data, and telecommunications).

1.2. Security Objectives:

When performing security tasks, security professionals try to protect their environments as effectively as possible. These actions can also be described as protecting confidentiality, integrity, and availability (CIA), or maintaining CIA. CIA stands for

- **Confidentiality:** Ensure that no data is disclosed intentionally or unintentionally. It covers two related concepts:

  Data confidentiality: Assures that private or confidential information is not made available or disclosed to unauthorized individuals.

  Privacy: Assures that individuals control or influence what information related to them may be collected and stored and by whom and to whom that information may be disclosed.
- **Integrity:** Make sure that no data is modified by unauthorized personnel, that no unauthorized changes are made by authorized personnel, and that the data remains consistent, both internally and externally. This term covers two related concepts:

  Data integrity: Assures that information and programs are changed only in a specified and authorized manner.

  System integrity: Assures that a system performs its intended function in an unaffected manner, free from deliberate or inadvertent unauthorized manipulation of the system.
- **Availability:** Assures that systems work promptly and service is not denied to authorized users. Availability also ensures timely and reliable access to and use of information. A loss of availability is the disruption of access to or use of information or an information system.

Although the use of the CIA triad to define security objectives is well established, some in the security field feel that additional concepts are needed to present a complete picture. Two of the most commonly mentioned are

- **Authenticity**: The property of being genuine and being able to be verified and trusted; confidence in the validity of a transmission, a message, or message originator. This means verifying that users are who they say they are and that each input arriving at the system came from a trusted source.
- **Accountability**: The security goal that generates the requirement for actions of an entity to be traced uniquely to that entity. This supports nonrepudiation, deterrence, fault isolation, intrusion detection and prevention, and after-action recovery and legal action. Because truly secure systems are not yet an achievable goal, users must be able to trace a security breach to a responsible party. Systems must keep records of their activities to permit later forensic analysis to trace security breaches or to aid in transaction disputes.

The opposite of CIA is disclosure, alteration, and denial (DAD).A major security objective is measuring the costs and benefits of security. If user want to measure the cost of securing an entity, whether it is data on networks, data on computers, or other assets of an organization, he need to know something about risk assessment. Generally, the assets of an organization have multiple risks associated with them, such as:

- Equipment failure
- Theft
- Misuse
- Viruses
- Bugs

After the identification of assets at risk as well as the risks themselves, the probability of a risk occurring is determined. Although there are numerous threats that could affect an organization, not all of them are likely to occur in your environment. For example, an earthquake is highly possible if you live close to San Francisco but not if you live in New York City. For this reason, a realistic assessment of the risks must be performed. Research must be performed to determine the likelihood of risks occurring to certain resources at specific places. By determining the likelihood of a risk occurring within a year, you can determine what is known as the ***annualized rate of occurrence (ARO).***

Once the ARO is calculated for a risk, you can compare it to the economic loss associated with an asset. This is the value that represents how much money would be lost if the risk occurred. The ARO includes the price of the new equipment, the hourly wage of the person replacing the equipment, and the cost of employees unable to perform their work. This value, which provides the total cost of the risk, is the single loss expectancy (SLE).

To plan for the probable risk, you need to budget for the possibility that the risk will happen. To do this, you need to use the ARO and the SLE to find the annual loss expectancy (ALE). To illustrate how this works, let's say that the probability of a web server failing is 30 percent. This would be the ARO of the risk. If the e-commerce site hosted on this server generates $10,000 an hour and the site is estimated to be down two hours while the system is repaired, the cost of this risk is $20,000. In addition to this cost, there would be the cost of replacing the server itself. If the server cost $6000, this would increase the cost to $26,000. This would be the SLE of the risk. By multiplying the ARO and the SLE, you find how much money needs to be budgeted to deal with this risk.

1.3. Defining Trust:

"*Trust is the likelihood that people will act the way as they are expected to act*". Trust is often based on past experiences. Trust can exist only between two individuals who know each other. A total stranger can't be trusted, but can be trusted over a certain period of time. An exception to this rule exists in the context of networking. A user might be willing to trust a stranger if he know that someone he trust trusts the stranger. This is, after all, the basis for Secure Sockets Layer (SSL) and certificate exchange.

Now that trust is defined, a list of resources can be developed that ranges from most trusted to least trusted, as shown in Figure 1.1.

1.3.1. Most Trusted

The most trusted network resources in an organization are internal servers, domain controllers, and storage devices attached to the network. Only a limited number of well-known people should have access to these devices.

1.3.2. Less Trusted

This category includes the internal users and the remote, authenticated users. On a certain level, an organization has to trust its users, internal or remote, because otherwise these users cannot perform their jobs. Despite the trust granted to them, some people in an organization use the passwords they have to do things they are not supposed to do. Although most employees can be trusted, it is because of the minority that abuses its privileges that this group is categorized as less trusted, not most trusted.

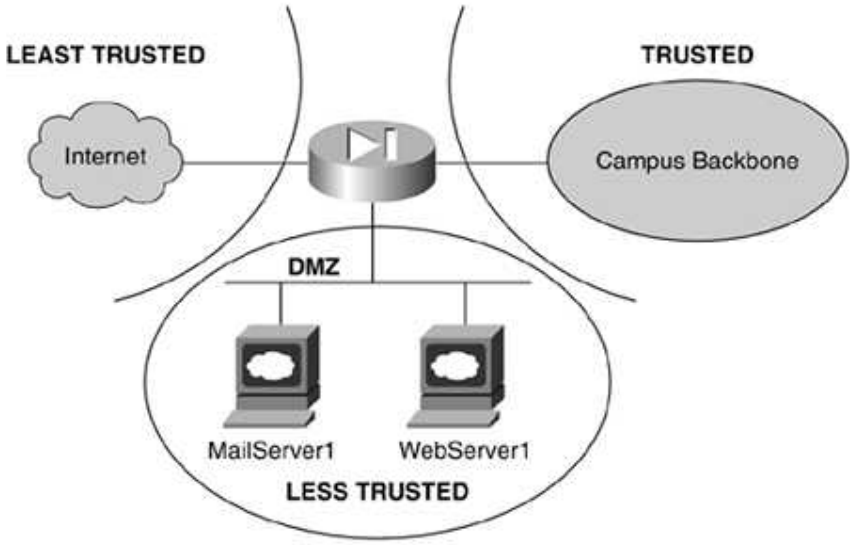


Figure 1.1: Security Zones

1.3.3. Least Trusted

The least trusted (sometimes referred to as untrusted) resources and users are Internet servers and remote, unauthenticated users. User can never trust an Internet server because he is not sure what is behind it. That is the reason for using digital certificates

1.4. Weaknesses and Vulnerabilities:

External and internal weaknesses and vulnerabilities must be considered. External weaknesses include malware, spyware, hackers, crackers, and script kiddies.

Malware is a group of destructive programs such as viruses or worms. The following list defines some types of malware:

- **Virus** A virus is a piece of code that is capable of attaching to programs, disks, or computer memory to propagate itself. Viruses also carry a payload with an action they must carry out. The action can be anything from displaying a message to erasing a computer hard disk.
- **Worm** Like viruses, worms replicate. They are capable of making copies of themselves, and they use e-mail and network facilities to spread to other resources.
- **Trojan horse** Trojan horses do not have the capability to replicate. By pretending to be a useful utility or a clever game, Trojan horses convince the user that they should be installed on a PC or on a server.
- **Spyware** This is software that gathers user information and sends it to a central site. The popular music-sharing program Kazaa came with spyware attached to the original program. It is even mentioned in the user license agreement, so that when users accept the agreement, they are giving permission to install the spyware and send personal user information to a central site.
- **Hoax** This is a special kind of malware. Hoaxes do not contain any code, instead relying on the gullibility of the users to spread. They often use emotional subjects such as a child's last wish. Any e-mail message that asks you to forward copies to everyone you know is almost certainly a hoax.

Often driven by a passion for computing, a hacker is a person who is proficient in using and creating computer software to gain illegal access to information. Hackers do no malicious damage whatsoever.

NOTE

The term hacker is used to describe an individual who attempts an unauthorized and malicious activity. The press and public have muddied the definitions so much that both now often mean people with malicious intent.

Crackers differ from hackers. A cracker uses various tools and techniques to gain illegal access to various computer platforms and networks with the intention of harming the system.

Script kiddies are a subclass of crackers. They use scripts made by others to exploit a security flaw in a certain system.

A common security mistake is to assume that attacks always come from outside your organization. Many companies build a massive wall around their buildings, but they leave all inside doors unlocked. The following list shows some of the potential threats from inside your organization:

- **Authenticated users:** These users already have access to the network. They are authenticated and authorized to use certain resources on the network. Often they use the access they have to get to confidential data such as payrolls or personnel records.
- **Unauthorized programs:** Users within your organization sometimes install additional programs and plug-ins that are not authorized by your organization. Often they open a hole to your network by doing this.
- **Unpatched software:** It is also very important to keep up with the latest updates or patches. Once a software bug or flaw is identified, vendors provide an update to their affected customers. It is good practice to check for updates and patches frequently, especially for your browser and operation system. If you are running a Microsoft operating system such as Windows 2000, you need to go to following URL:

1.5. Responsibilities for Network Security:

Many people are involved in the security process of an organization, ranging from senior management to the everyday user. Senior management enforces the security policy. Policies and rules that come from senior management that are based on the saying "Do as I say, not as I do" are usually ignored. If you want users to participate in maintaining security, they need to believe that you take it seriously. Users need to be aware of not only the existence of security, but also the consequences of not abiding by the rules. The best way to do this is by providing short security-training seminars in which people can ask questions and talk about issues. Another excellent security practice is to post articles describing security breaches in highly frequented areas (the coffee corner or the cafeteria).

In addition, governments are now playing a significant role in security by enacting laws to create a legal structure to surround emerging technologies such as wireless and voice communication over IP. In this way, governments have created legal requirements that need to be taken into account when making security decisions. The following list describes some of these legal requirements:

- **HIPAA:** The Health Insurance Portability and Accountability Act restricts disclosure of health-related data along with personally identifying information.
- **GLB:** The Gramm-Leach-Bliley Act affects U.S. financial institutions and requires disclosure of privacy policies to customers.
- **ECPA:** The Electronic Communications Privacy Act specifies who can read whose e-mails and under what conditions.

1.6. Risk and Vulnerability:

Attackers choose their targets based on vulnerabilities they have observed. Individuals and organizations often try to shield themselves from one instance or form of an attack, but they must keep in mind that the attacker can easily shift focus to newly exposed vulnerabilities. Even if you experience some success in tackling several attacks, risks always remain, and the need to confront threats is going to exist for the foreseeable future.

`

Attackers continue to benefit from certain tactical advantages. Time, location, place, and method of attack are just some of the parameters the aggressor can choose to act unpredictably and unexpectedly. After reducing vulnerability in one area, you can expect attackers to alter their plans by pursuing other exposed and unprotected targets. Most of the time, the attacker has no time pressure at all and can carefully and patiently plan an attack weeks, months, or even years in advance. As a security administrator, you can be assured that new plans are underway that have not yet been considered by your organization.

With the increasing popularity of the Internet, terrorist groups might seek to cause damage by means of a cyber-attack. They can exploit the Internet to collect information and to recruit, command, and control their accomplices. Terrorists can even raise funds for their activities through the Internet. Terrorist groups can also use the Internet to expand their technical capabilities to further explore cyber-attacks. They can develop their skill sets with the intention of targeting commercial and governmental computer-driven applications in order to disturb financial networks such as stock market exchanges and international banking. Other targets that are increasingly threatened include energy delivery, aviation, and security networks.

Adequate security protection against cyber-attacks is an ongoing process. It is implemented through new technologies, system redesign, and adaptation of existing procedures.

The enterprise or organization should always be conscious of designing systems and procedures that eliminate vulnerabilities and reduce risks. If an identified vulnerability cannot be eliminated immediately, reduction of the associated risk to an acceptable level should be the primary goal. When risks cannot be reduced to a level that is acceptable through network design, security equipment, alerts, and alarms, the alternative of personnel awareness through training and procedures should be utilized.

1.7. TCP/IP Suite Weaknesses:

Communication on the Internet is based on the Transmission Control Protocol/Internet Protocol (TCP/IP) protocol suite. The TCP/IP protocol suite was developed in the mid-1970s as part of research by the Defense Advanced Research Projects Agency (DARPA).

With the introduction of personal computers as standalone devices, the strategic importance of interconnected networks was quickly realized. The strategic importance of networks was first realized in the development of local-area networks (LANs) that shared printers and hard drives. The importance of networks increased in a second phase with the development of worldwide applications such as e-mail and file transfers. The globalization of business caused web applications to be developed to support customers and clients all over the world with a focus on increasing efficiency and productivity for organizations. Now TCP/IP is seen as the de jure standard for Internet communication, enabling millions of users to communicate globally. Computer systems in general communicate with each other by sending streams of data (bytes), as displayed in Figure 1.2.

This section presents a brief overview of the IP protocol and TCP protocol characteristics and then examines some of the TCP/IP weaknesses. Readers should not expect a full description of the TCP/IP protocol suite, but rather information relevant to a discussion of the weaknesses. Figure 1.3 maps the TCP/IP protocol stack to the OSI model and serves as a framework for the discussion.

1.7.1. IP

The IP layer of the TCP/IP stack corresponds to the OSI network layer. IP is a connectionless protocol providing routing of datagrams in a best-effort manner. The following sections present topics that will help you to further understand the design weaknesses of the protocol.

The IP datagram is a combination of a number of bytes (IP header) that prefixes the data received from the transport (and higher) layer. Figure 1.4 shows the complete IP header format, but only the relevant fields are discussed.

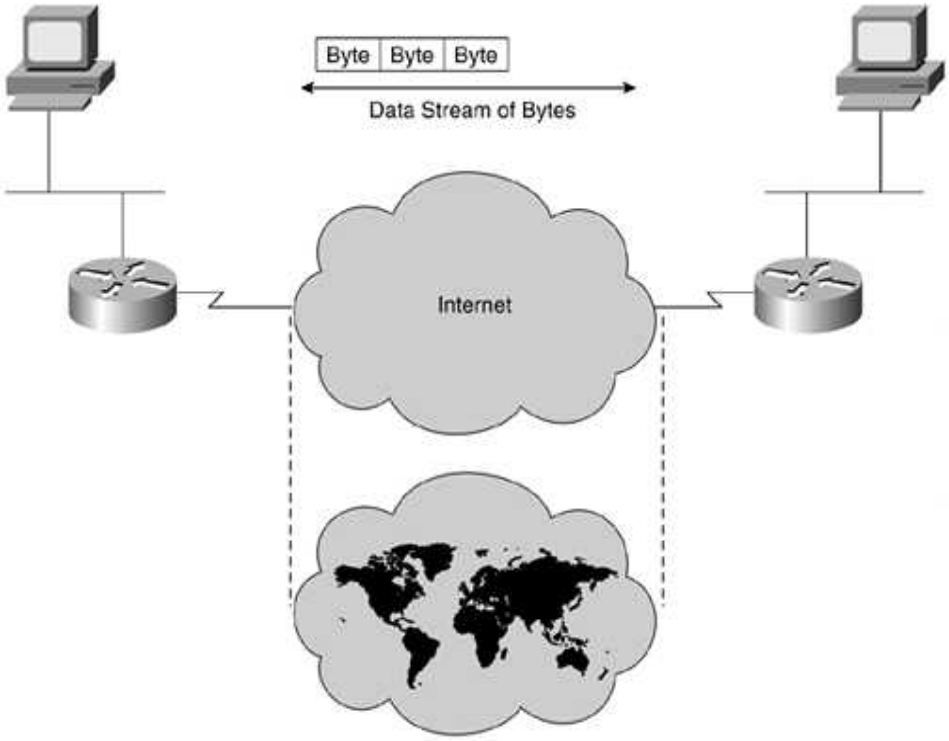


Figure 1.2: Internet Communication

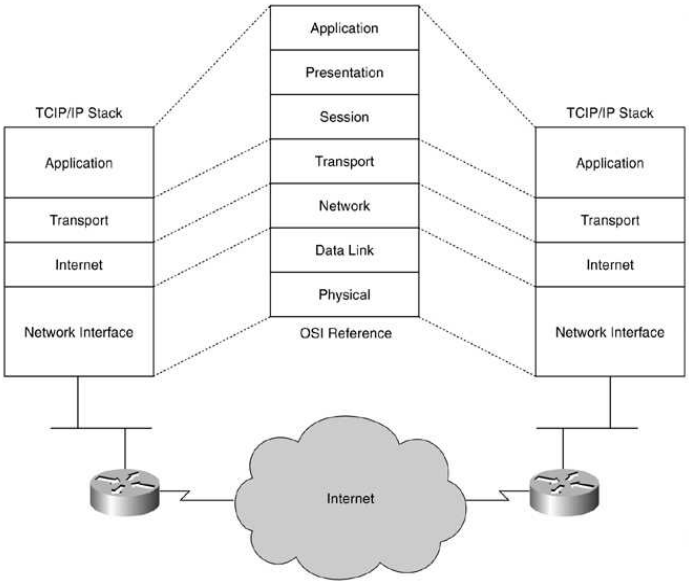


Figure 1.3: TCP/IP Protocol Mapped to the OSI Model. Four layers of the TCP/IP protocol stack map to seven layers of the OSI model.

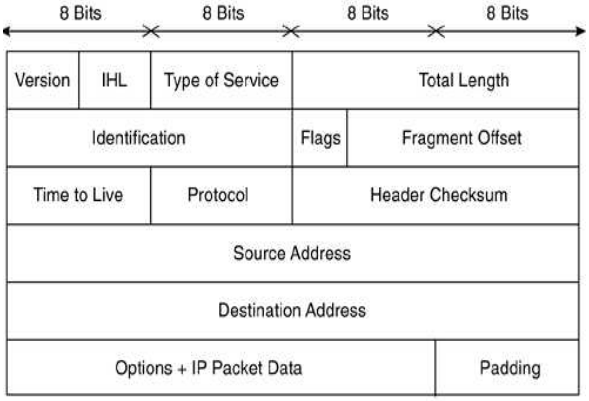


Figure 1.4: IP Datagram Format

IP addressing (both the source IP address and the destination IP address) is used to identify the end stations involved in the transport of datagrams for communication.

End stations with source IP addresses and destination IP addresses on the same segment have direct delivery of packets. When source and destination end stations are not on the same network, there can be multiple paths. Path selection and decision is made by specialized computer systems whose primary function is routing network traffic. These systems are referred to as routers for the remainder of this book.

IP fragmentation offset is used to keep track of the different parts of a datagram. Splitting larger datagrams may be necessary as they travel from one router to the next router in a small packet network, for example, because of interface hardware limitations. The information or content in the offset field is used at the destination to reassemble the datagrams. All such fragments have the same Identification field value, and the fragmentation offset indicates the position of the current fragment in the context of the original packet. Also important to keep in mind is the existence of the IP Options field. This makes the IP header variable in length. Table 1.2 illustrates all the fields of the IP header.

Header Field	Description
Version	Indicates the format of the Internet header (4 bits)
Internet Header Length (IHL)	Specifies the length of the Internet header in 32-bit words (4 bits)
Type of Service	Provides an indication of the abstract parameters of the quality of service desired (8 bits)
Total Length	Specifies the length of the datagram, measured in octets (16 bits
Identification	Value assigned by the sender to aid in assembling the fragments (16 bits)
Flags	Various control flags (3 bits)
Fragment Offset	Indicates where in the datagram this fragment belongs (13 bits)
Time to Live	Indicates the maximum time the datagram is allowed to remain in the Internet system (8 bits)
Protocol	Indicates the next level protocol used (8 bits)
Header Checksum	A checksum on the header (16 bits)
Source Address	The source IP address (32 bits)
Destination Address	The destination IP address (32 bits)
Options	The Options field is variable in length
Padding	Internet header padding used to ensure that the Internet header ends on a 32-bit boundary

Table 1.1: IP Header Fields

1.7.2. TCP

The TCP or transport layer of the TCP/IP stack corresponds to the OSI transport layer. TCP is a connection-oriented protocol providing delivery of segments in a reliable manner. Some TCP characteristics are highlighted in the next section because they might be used to exploit some vulnerability in the TCP/IP protocol suite.

The TCP segment is a combination of a number of bytes (TCP header) that prefixes the data received from the upper layers. Figure 1.5 shows the complete TCP header format, but as with the discussion of the IP header, only the relevant fields are covered in this chapter.

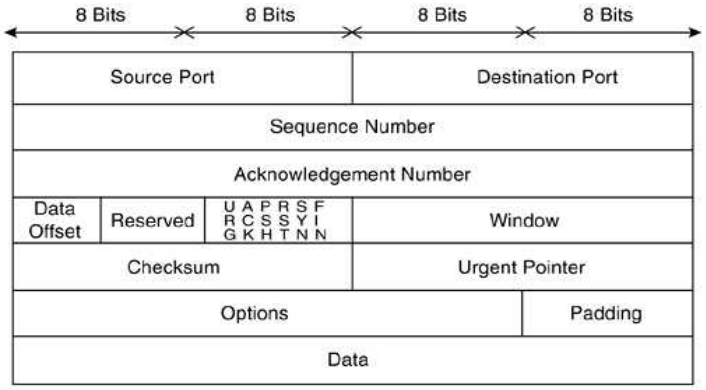


Figure 1.5: TCP Segment Format

TCP uses port or socket numbers to pass information to the upper layers. This mechanism enables the protocol to multiplex communication between different processes in the end stations. In other words, the port numbers keep track of the different conversations crossing the network at the same time. Port numbers assigned by the operating system are also called sockets.

NOTE: The port numbers are divided into three ranges: the Well-Known Ports, the Registered Ports, and the Private Ports.

Application Layer	Port Number
FTP	21
Telnet	23
SMTP	25
HTTP	80
HTTPS	443

Table 1.2: IP Header Fields

An established connection between two end stations can be uniquely identified by four parameters: source and destination IP addresses and source and destination port numbers. It is important to understand the underlying mechanism in order to configure extended access lists on routers to implement pass/block filtering decisions based on these numbers. Firewalls can also be configured to filter based on TCP ports.

Data exchange using TCP does not happen until a three-way handshake has been successfully completed. The connection needs to be initialized or established first on sequence numbers. These numbers are used in multiple packet transmissions for reordering and to ensure that no packets are missing. The Acknowledgment number defines the next expected TCP octet and is used for reliability of the transmission. The sequence number in combination with the Acknowledgment number serves as a ruler for the sliding window mechanism. This sliding window mechanism uses the window field to define the size of the receiving buffers. In other words, the window field is used to define the number of octets that the sender is willing to accept.

1.7.3. TCP/IP Security Issues

Most of TCP/IP weaknesses are likely because the development of the protocol dates from the mid-1970s. Vendors of network equipment and operating systems have made code improvements over time to disable many of the attacks that are described in the following sections.

1.7.3.1.IP Address Spoofing

In this type of attack, the attacker replaces the IP address of the sender or in some rare cases the destination, with a different address. IP spoofing is normally used to exploit a target host. In other cases, it is used to start a denial-of-service (DoS) attack. In a DoS attack, an attacker modifies

the IP packet to mislead the target host into accepting the original packet as a packet sourced at a trusted host. The attacker must know the IP address of the trusted host to modify the packet headers (source IP address) so that it appears that the packets are coming from that host.

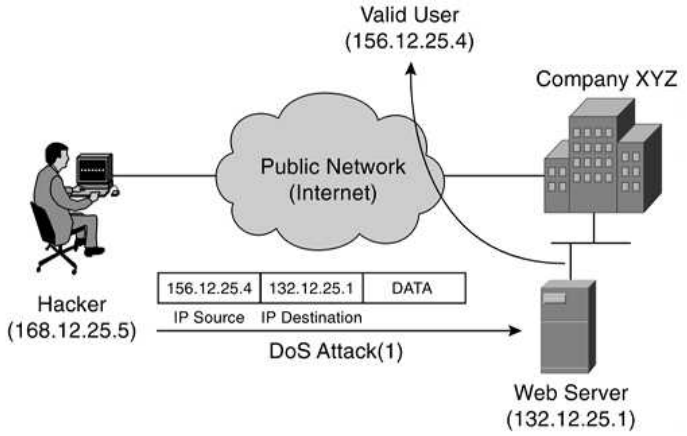


Figure 1.6: DoS Attack Using IP Spoofing

For all DoS attacks launched against a host the attacker is not interested in retrieving effective data or information from the intended victim. The attacker has only one goal: to deny the use of service that the web server provides to valid users without being revealed. Therefore, the return address or source IP address can be spoofed.

In Figure 1.6, the attacker has the IP address 168.12.25.5 and is connected to the Internet. For normal traffic interaction between a workstation with a valid source IP address (168.12.25.5) and the web server (132.12.25.1), the packet is constructed with a source IP address of 168.12.25.5 and a destination IP address of 132.12.25.1. The web server returns the web page using the source IP address specified in the request as the destination IP address, 168.12.25.5, and its own IP address as the source IP address, 132.12.25.1.

Let's now assume that a DoS attack is launched from the attacker's workstation on Company XYZ's web server using IP spoofing. Imagine that a spoofed IP address of 156.12.25.4 is used by the workstation, which is a valid host. Company XYZ's web server executes the web page request by sending the information or data to the IP address of what it believes to be the originating end station (156.12.25.4). This workstation receives the unwanted connection attempts from the web server, but it simply discards the received data. It's becoming clear that multiple simultaneous attacks of this sort deny the use of service that the web server provides to valid users. Locating the origin of the attacker launching the DoS attack is very complex when IP address spoofing is used.

1.7.3.2.Covert Channels

A covert or clandestine channel can be described as a pipe or communication channel between two entities that can be exploited by a process or application transferring information in a manner that violates the system's security specifications.

More specifically for TCP/IP, in some instances, covert channels are established, and data can be secretly passed between two end systems. Let's take Internet Control Message Protocol (ICMP) as an example. In the following types of circumstances, ICMP messages are sent to provide error and control mechanisms:

- Testing connectivity/reachability using datagrams echo and Echo-Reply messages
- Reporting unreachable destinations for datagrams Destination Unreachable message
- Reporting buffer capacity problems for forwarding datagrams Source Quench message
- Reporting route changes in the path for datagrams Redirect messages

ICMP resides at the Internet layer of the TCP/IP protocol suite and is implemented in all TCP/IP hosts. Based on the specifications of the ICMP Protocol, an ICMP Echo Request message

should have an 8-byte header and a 56-byte payload. The ICMP Echo Request packet should not carry any data in the payload. However, these packets are often used to carry secret information. The ICMP packets are altered slightly to carry secret data in the payload. This makes the size of the packet larger, but no control exists in the protocol stack to defeat this behavior. The alteration of ICMP packets gives intruders the opportunity to program specialized client-server pairs. These small pieces of code export confidential information without alerting the network administrator. Blocking ICMP packets that exceed a certain limit size is the only solution to protect against this vulnerability.

An example of a tool that uses this covert channel technique is Loki. The concept of the Loki tool is simple: It is a client-server application that tunnels arbitrary information in the data portion of ICMP_ECHO and ICMP_ECHO REPLY packets. Loki exploits the covert channel that exists inside of ICMP_ECHO traffic. Figure 2-6 illustrates this tool.

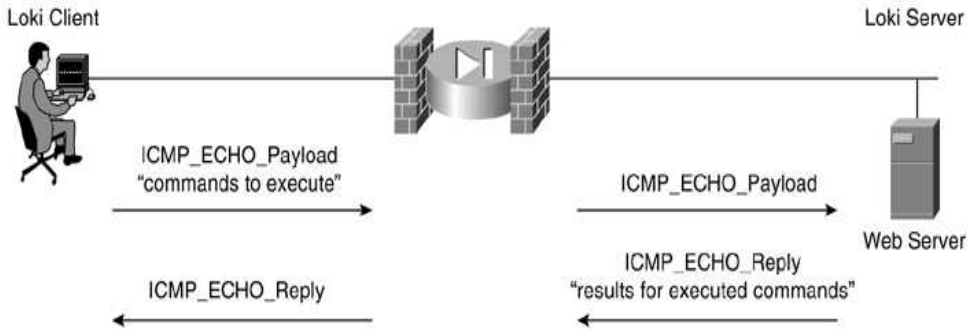


Figure 1.7: Loki tool

In general, covert channels are prevalent in nearly all the underlying protocols of the TCP/IP protocol suite.

1.7.3.3.IP Fragment Attacks

The TCP/IP protocol suite, or more specifically IP, allows the fragmentation of packets. IP fragmentation offset is used to keep track of the different parts of a datagram. The information or content in this field is used at the destination to reassemble the datagrams. All such fragments have the same Identification field value, and the fragmentation offset indicates the position of the current fragment in the context of the original packet.

Many access routers and firewalls do not perform packet reassembly. In normal operation, IP fragments do not overlap, but attackers can create artificially fragmented packets to mislead the routers or firewalls. Usually, these packets are small and almost impractical for end systems because of data and computational overhead.

Let's go into a little more detail. The ingeniously constructed second fragment of a packet can have an offset value that is less than the length of the data in the first fragment. Upon packet reassembly at the end station, the second fragment overrides several bytes of the first fragment. These malformed IP packets cause the operating system at the end station to function improperly or even to crash.

1.7.3.4.TCP Flags

Data exchange using TCP happen after a three-way handshake has been successfully completed. This handshake uses different flags to influence the way TCP segments are processed. There are 6 bits in the TCP header that are often called flags. In Figure 1.4, six different flags are part of the TCP header: Urgent pointer field (URG), Acknowledgment field (ACK), Push function

(PSH), Reset the connection (RST), Synchronize sequence numbers (SYN), and sender is finished with this connection (FIN).

Figure 1.8 illustrates this three-way handshake in a little more detail, elaborating on some of the flags used.

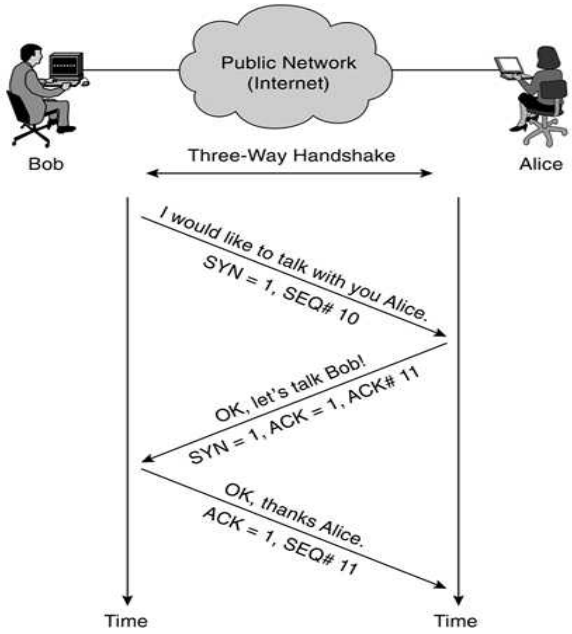


Figure 1.8: Three-Way Handshake Using TCP Flags

Bob wants to start talking with Alice, so he initiates the TCP session with the SYN bit (flag) set in the first TCP segment. If Alice is happy to talk to Bob, she responds with the SYN flag and ACK flag set to 1. If she is unwilling to talk to Bob, she responds with an RST (reset) flag set to 1.

Abuse of the normal operation or settings of these flags can be used by attackers to launch DoS attacks. This causes network servers or web servers to crash or hang. Table 1.3 illustrates some invalid combinations of these parameters.

SYN	FIN	PSH	RST	Validity
1	1	0	0	Illegal combinations
1	1	1	0	Illegal combinations
1	1	0	1	Illegal combinations
1	1	1	1	Illegal combinations

Table 1.3: invalid Combination of three way handshake parameters

The attacker's ultimate goal is to write special programs or pieces of code that are able to construct these illegal combinations resulting in an efficient DoS attack.

1.7.3.5.SYN Flood

The TCP/IP protocol suite relies on the use of multiple timers during the lifetime of a session. These timers include the Connection Establishment timer, the FIN_WAIT timer, and the KEEP_ALIVE timer. The following list elaborates on the three-way handshake mechanism presented in Figure 1.8:

- Connection Establishment timer Starts after SYN is sent during the initial connection setup (step 1 of the three-way handshake).
- FIN_WAIT timer Starts after FIN is sent and the originator is waiting for an acknowledgement to terminate the session.
- KEEP_ALIVE timer Counter restarts after every segment of data is transmitted. This timer is used to periodically probe the remote end.

All these timers are critical for proper and accurate data transmission using TCP/IP. These timers (or lack of certain timers) are often used and exploited by attackers to disable services or even to enter systems. For instance, after step 2 of the three-way handshake, no limit is set on the time to wait after receiving a SYN. The attacker initiates many connection requests to the web server of Company XYZ (almost certainly with a spoofed IP address). The SYN+ACK packets (Step 2) sent by the web server back to the originating source IP address are not replied to. This leaves a TCP session half-open on the web server. Multiple packets cause multiple TCP sessions to stay open.

Based on the hardware limitations of the server, a limited number of TCP sessions can stay open, and as a result, the web server refuses further connection establishments attempts from any host as soon as a certain limit is reached. These half-open connections need to be completed or timed out before new connections can be established.

This vulnerability can be exploited by the attacker to actually remove a host from the network for several seconds. In the meantime, this temporarily disabled platform can be used to deposit another exploit or to install a backdoor.

1.7.3.6.Closing a Connection by FIN

These types of attacks also known as connection-killing attacks. In normal operation, the sender sets the TCP FIN flag indicating that no more data will be transmitted and the connection can be closed down. This is a four-way handshake mechanism, with both sender and receiver expected to send an acknowledgement on a received FIN packet. During an attack that is trying to kill connections, a spoofed FIN packet is constructed. This packet also has the correct sequence number, so the packets are seen as valid by the targeted host. These sequence numbers are easy to predict. This process is referred to as TCP sequence number prediction, whereby the attacker either sniffs the current Sequence and Acknowledgment (SEQ/ACK) numbers of the connection or can algorithmically predict these numbers.

Once the packet is constructed and sent, the receiving host believes the spoofed sender has no more data to be transmitted. Any other packets received are ignored as false and dropped. The remaining packets for completing the four-way handshake are provided by the spoofed sender. Similar connection-killing attacks are launched using the RST flag.

1.7.3.7.Connection Hijacking

TCP connections can be hijacked by unauthorized users without much difficulty. In Figure 1.9, an authorized user (Employee X) sends HTTP requests over a TCP session with the web server.

The web server accepts the packets from Employee X only when the packet has the correct SEQ/ACK numbers. As seen previously, these numbers are important for the web server to distinguish between different sessions and to make sure it is still talking to Employee X. Imagine that the cracker starts sending packets to the web server spoofing the IP address of Employee X, using the correct SEQ/ACK combination. The web server accepts the packet and increments the ACK number.

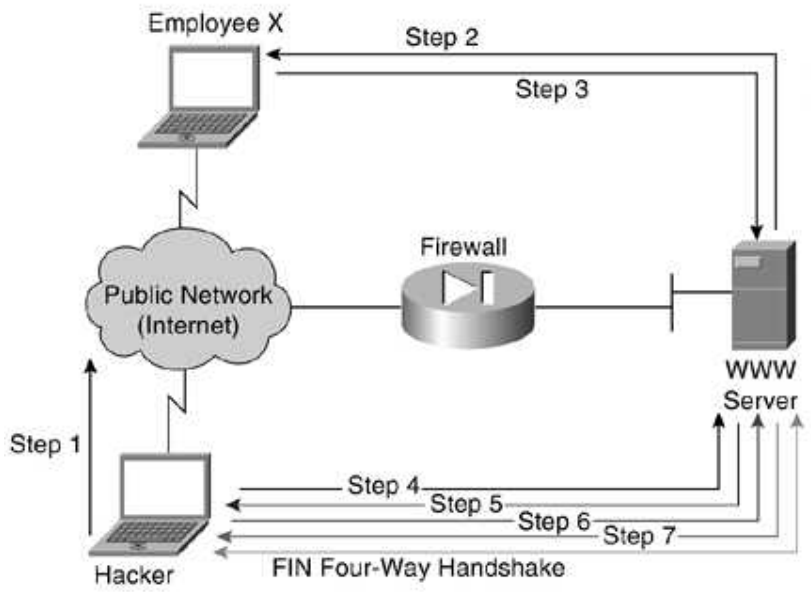


Figure 1.9: Connection Hijacking

In the meantime, Employee X continues to send packets but with incorrect SEQ/ACK numbers. As a result of sending unsynchronized packets, all data from Employee X is discarded when received by the web server. The attacker pretends to be Employee X using the correct numbers. This finally results in the cracker hijacking the connection, whereby Employee X is completely confused and the web server replies assuming the cracker is sending correct synchronized data.

The following steps outline the different phases of a connection-hijacking attack, as shown in Figure 1.9:

Step 1. The attacker examines the traffic flows with a network monitor and notices traffic from Employee X to a web server.

Step 2. The web server returns or echoes data back to the origination station (Employee X).

Step 3. Employee X acknowledges the packet.

Step 4. The cracker launches a spoofed packet to the server.

Step 5. The web server responds to the cracker. The cracker starts verifying SEQ/ACK numbers to double-check success. At this time, the cracker takes over the session from Employee X, which results in a session hanging for Employee X.

Step 6. The cracker can start sending traffic to the web server.

Step 7. The web server returns the requested data to confirm delivery with the correct ACK number.

Step 8. The cracker can continue to send data (keeping track of the correct SEQ/ACK numbers) until eventually setting the FIN flag to terminate the session.

Sniffing Internet traffic is not necessarily easily accomplished. Most hijacking attacks require access to the local wire or the broadcast domain. An excellent tool to monitor the local wire is Ethereal.

These connection-hijacking attacks often occur unnoticed. The Employee X session hangs, but most Internet users reconnect the session and observe this incident as a network problem.

Luckily, it is true that not all session hangs are caused by connection-hijacking attacks but involve different causes.

1.7.4. Countermeasures

As a network administrator, it is important to understand the vulnerabilities that exist in network in order to implement effective countermeasures. TCP/IP vulnerabilities are nothing new, but the number of TCP/IP attacks is increasing considerably with the growth of the Internet. Subsequent chapters in this book refer to these TCP/IP vulnerability issues, and more prevention and protection methods are discussed.

1.8. Buffer Overflows

A buffer is a temporary data storage area used to store program code and data. When a program or process tries to store more data in a buffer than its capacity, a buffer overflow occurs.

Buffers are temporary storage locations in memory that are able to store a fixed amount of data in bytes. When more data is retrieved than can be stored in a buffer location, the additional information must go into an adjacent buffer, resulting in overwriting the valid data held in them.

Buffer overflows are nowadays very common security vulnerabilities. Buffer overflows are especially useful for crackers trying to infiltrate remote networks, where anonymous users try to gain access or control of a host. These types of attacks represent one of the most serious security threats on the Internet, making up the majority of all security attacks because the vulnerabilities are common and easy to exploit. The attacker has the ability to inject and execute the code on a remote system, gaining full or privileged access.

1.8.1. Buffer Overflow Mechanisms

Buffer overflow vulnerabilities exist in different types. But the overall goal for all buffer overflow attacks is to take over the control of a privileged program and, if possible, the host. The attacker has two tasks to achieve this goal. First, the dirty code needs to be available in the program's code address space. Second, the privileged program should jump to that particular part of the code, which ensures that the proper parameters are loaded into memory.

The program code, or shell code, is the software that provides the interface between the human operator and the operating system of a computer. In other words, it is the command interpreter that provides a user interface to the kernel.

The first task can be achieved in two ways: by injecting the code in the right address space or by using the existing code and modifying certain parameters slightly. The second task is a little more complex because the program's control flow needs to be modified to make the program jump to the dirty code.

1.8.2. Buffer Overflow Protection

Several approaches can be used to defend hosts from buffer overflow vulnerabilities and attacks. The most important approach is to have a concerted focus on writing correct code.

Software development teams need to understand how to write secure applications. Tools and techniques have been developed to help programmers write pieces of code that are immune to buffer overflow attacks.

A second method is to make the data buffers (memory locations) address space of the program code non-executable. This type of address space makes it impossible to execute code, which might be infiltrated in the program's buffers during an attack. As previously discussed, trying to inject the code into the program's space is just one element of the buffer overflow attack. Another essential part is taking over the flow control of the program under attack. This threat can be eliminated by implementing array-bound control or array-bound checks during debugging phases of the program development. The implementation of these checks ensures that buffers stay in the correct predefined range and also verifies that buffers cannot be overflowed at all.

1.8.3. Countermeasures

`

This chapter has touched so far only on buffer overflow vulnerabilities, attacks, and some defenses. Understanding these buffer overflow mechanisms is important because they form a major part of all existing remote penetration issues in today's internetworking infrastructure. Subsequent chapters in this book refer to these remote penetration vulnerability issues and discuss more prevention and protection methods (access filters, intrusion detection systems, and auditing tools).

1.9. Spoofing Techniques:

Spoofing methods are used by crackers to compromise computer systems. Many people mistakenly think that spoofing is an actual attack. In reality, spoofing is just one step in a process whereby an attacker tries to exploit the relationship between two hosts. Two spoofing techniques are discussed with some guidelines on spoofing prevention.

1.9.1. Address Resolution Protocol Spoofing

The Address Resolution Protocol (ARP) provides a mechanism to resolve, or map, a known IP address to a MAC sub-layer address. In Figure 1.10, two hosts are attempting to start a conversation across a multi-access medium such as Ethernet.

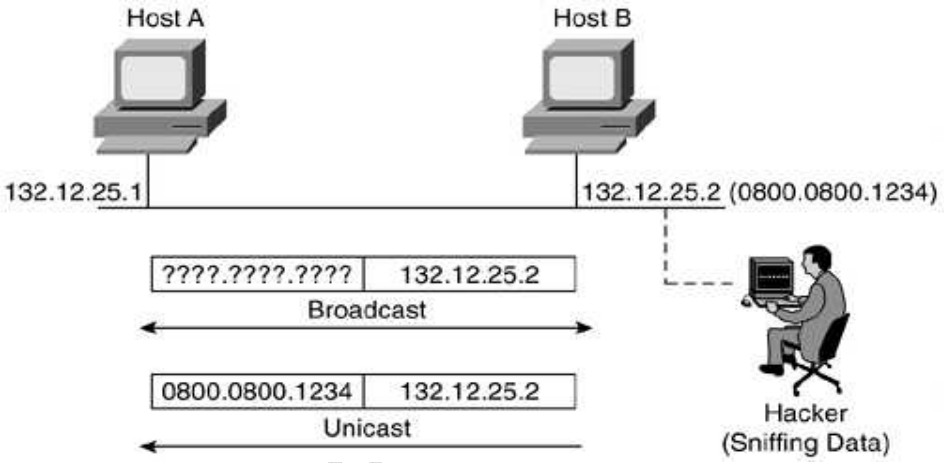


Figure 1.10: Connection Hijacking

Host A wants to initiate the conversation with Host B but requires both the IP address and the MAC address. During the conversation setup, Host A is aware only of Hosts B's IP address, 132.12.25.2. To determine a destination MAC address for a datagram, the ARP cache table locally in Host A is checked first. If the MAC address is not in the table, Host A sends an ARP request, which is a broadcast on the wire looking for a destination station Host B with IP address 132.12.25.2. Every host on the network receives this broadcast. Host B hears the message, finds out the message is destined for it, and replies with an ARP reply containing its MAC address and IP address.

There is no real authentication; the verification between two hosts is based only on the hardware address, which is a weak part of the ARP process. Using ARP spoofing, the cracker can exploit this hardware address authentication mechanism by spoofing the hardware address of Host B. Basically, the attacker can convince any host or network device on the local network that the cracker's workstation is the host to be trusted. This is a common method used in a switched environment.

1.9.2. Domain Name Service Spoofing

Domain Name Service (DNS) is used for network clients who need an IP address of a remote system based on their names. The host sends a request to a DNS server including the remote system's name, and the DNS server responds with the corresponding IP address. DNS spoofing is the method whereby the hacker convinces the target machine that the system it wants to connect to is the machine of the cracker. The cracker modifies some records so that name entries of hosts correspond

to the attacker's IP address. There have been instances in which the complete DNS server was compromised by an attack.

1.9.3. Countermeasures

ARP spoofing can be prevented with the implementation of static ARP tables in all the hosts and routers of your network. Alternatively, you can implement an ARP server that responds to ARP requests on behalf of the target host. To counter DNS spoofing, the reverse lookup detects these attacks. The reverse lookup is a mechanism to verify the IP address against a name. The IP address and name files are usually kept on different servers to make compromise much more difficult. This chapter has touched so far on only two spoofing and antispoofing examples, but more prevention and protection methods (access filters, intrusion detection systems, and auditing tools) are discussed in the next chapters.

1.10. Social Engineering

In the world of information technology, social engineering exists in different forms but can be best described as the "***practice of tricking people into revealing passwords***." As a security administrator, it is your duty to be familiar with this threat and to educate your network users because social engineering can impact everyone in the organization.

1.10.1. Techniques

A number of techniques can be used in a social engineering attack. Three classic social engineering tricks are reverse social engineering, e-mails and phone calls, and authority abuse. This section outlines some of the most frequently used techniques.

During a reverse social engineering attack, the user is persuaded to ask the attacker for help. For instance, after gaining simple access to the user's system, the attacker breaks an application in the workstation, resulting in the user requiring and asking for help. The attacker then modifies the error messages to contain the attacker's contact information. The user contacts the attacker asking for assistance. This gives the attacker an easy way to obtain the required information.

Sending e-mails or phone calls is a much more direct approach, but it is less likely to be successful. An attacker calls a target individual asking the target to provide a username and password for completing a task quickly. This is by far the easiest type of social engineering attack to launch, but many individuals today are careful enough not to provide that information.

Here is a sample scenario. By pretending to be part of the technical support organization or just an important user, the attacker can pressure the target. For example, an attacker posing as a senior manager or system administrator could request usernames and passwords from subordinates to meet important deadlines or to resolve a problem quickly.

An alternative form of social engineering is as simple as guessing someone's password. Children's names, birthdays, and phone numbers are likely candidates to be guessed as passwords.

1.10.2. Countermeasures

As with all security threats, ways can be found to reduce the success of a social engineering attack. However, for social engineering attacks, the human factor can be easily influenced by an external event. A solid security policy defines expectations for users as well as for support personnel. In conclusion, your role as a security administrator requires you to understand the implications of social engineering threats and how these threats can be manifested. Only through such understanding can you take appropriate actions and ensure that protection of the organization is guaranteed on an ongoing basis.

1.11. Important Questions:

1. Which resources in a network are considered the most trusted?

2. List five types of malware.
3. What is a hoax?
4. What is the difference between a hacker and a cracker?
5. Attacks often come from inside your organization. List three potential threats from inside an organization.
6. Who is involved in the security process of an organization?
7. Name two legal requirements made by government agencies.
8. What is CIA?
9. What is SLE?
10. What is ALE?
11. What is IP fragmentation offset used for?
12. Name the method attackers use to replace the IP address of the sender or, in some rare cases, the destination address with a different IP address.
13. What is a covert TCP/IP channel?
14. The Ping of Death attack is a good example of what type of attack?
15. What happens during a buffer overflow?
16. List the two tasks the attacker must perform during a buffer overflow attack.
17. List two spoofing attacks.
18. During an ARP spoofing attack, does the attacker exploit the hardware address or the IP address of a host?
19. List two anti-spoofing measures for an ARP spoofing attack.
20. There are a number of techniques that can be used in a social engineering attack. List three techniques.

Chapter 02
UNDERSTANDING DEFENSES

2.1. Introduction:

Immense numbers of tools, techniques, systems, services, and processes are available to protect your data in today's challenging network environment. This chapter presents an overview of the techniques used to counter the network weaknesses. Because this chapter is an overview, many of the techniques are described at a basic level.

The chapter begins with a detailed explanation of digital IDs and how digital IDs can protect a network. Intrusion protection and intrusion prevention techniques are covered briefly in this chapter; will be discussed later in detail. This chapter describes how PC card based solutions counter network weaknesses. It also covers different encryption techniques that can be used to protect the network environment. The chapter continues with a discussion of how the physical security of a site can be achieved using access control and biometric techniques. Discussions of antivirus software and the basic functionality of firewalls conclude the chapter.

2.2. Digital IDs

A digital identity, or digital ID, is a means of proving user's identity or that a user has been granted permission to access information on network devices or services. The system or method behind digital IDs is similar to non-electronic means of identification. For instance, entering a private dancing club requires an ID check of a membership card to validate your claim to have the right to enter the venue. Using a photo ID on the card prevents others from abusing the card and impersonating valid members of the club.

Digital IDs are often required for electronic bank transactions, secure e-mail transmissions, and online shopping.

A digital ID is a means of proving that user has been granted permission to access information on network devices or services. To better understand the concept, let's examine the process of online shopping for a book from Cisco Press. Before the customer can trust the vendor, Cisco Press in this case, some sort of authentication needs to occur. The authentication occurs during the establishment of a connection. When the customer places an order, the customer's workstation web browser requests the certificate of the server. The certificate provides a form of authentication for the identity of the web server and also can serve as a way to guarantee that valid content is provided on the server.

The certificates combine the digital IDs and a set of keys to encrypt and validate the connection. These certificates are issued by a "***certification authority (CA)***" and are signed with the CA's private key. A CA is an organization that is trusted by both parties participating in a transaction. The role of the CA is to guarantee the identity of each party participating in the transaction.

Figure 2.1 shows the details contained within a digital certificate.

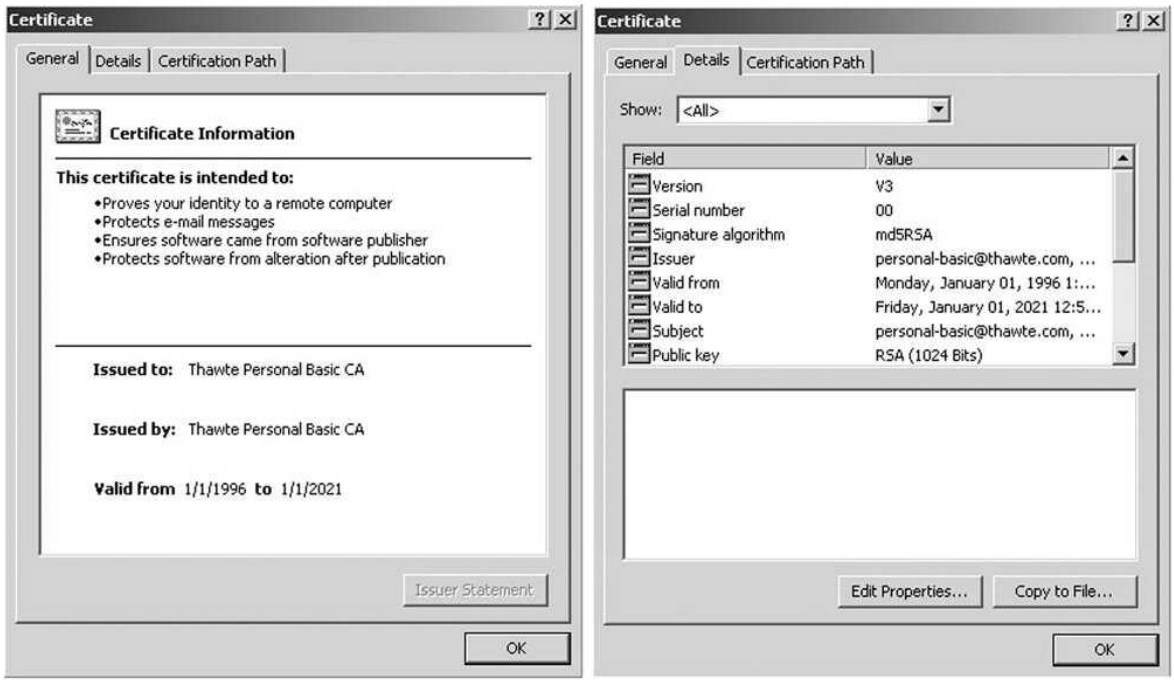


Figure 2.1: Certificate or Digital ID

The section at the left side on Figure 2.1 contains general information about the signature. The details of the certificate are displayed on the right side of the figure.

This digital ID is issued by Thawte Personal Basic CA. This signature is intended to prove the validity of the server's identity to a remote computer and can also be used to protect e-mail messages. The certificate ensures that the software is protected against alterations after publication. Typically, to check the parameters of a digital ID, a user can click the Details tab on the certificate. The parameters of the digital ID can include the following:

- Version number: V3
- Serial number: 00
- Signature algorithm: MD5RSA
- Name of the issuer: Thawte Personal Basic CA
- Expiration date: Friday, January 01, 2021
- Owner's name: Thawte Personal Basic CA
- Owner's public key: RSA (1024 bits)

All these fields are in compliance with the ITU-T X.509 specifications.

Let's go back to the book-ordering process through the Cisco Press website. The online user connects to the Cisco Press website using Internet Explorer. To start sending protected (encrypted) information, the web browser must obtain the proper certificate and be set up to use this certificate. From the moment the user visits the Cisco Press secure website, the Cisco Press web server automatically sends its certificate. Note that secure URLs add an s to http to become https. Figure 2-2 displays the certificate that is received from the server.

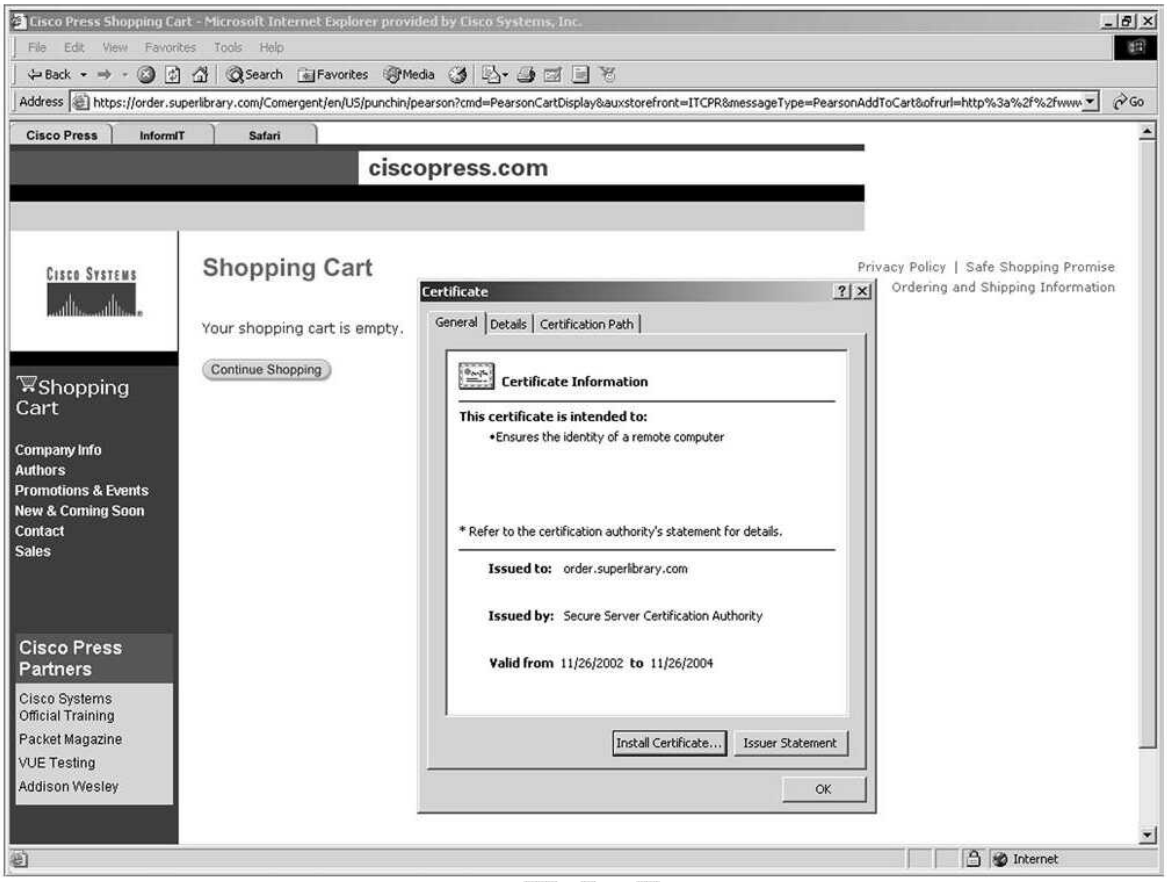


Figure 2.2: Secure Website

Once the exchange is successfully completed, the web browser displays a lock icon on the status bar of the application to indicate that a secure channel is established. This certificate guarantees the identity of the remote computer for the user. The certificate was issued by Secure Server Certification Authority for order.superlibrary.com and is valid until November 26, 2004.
Three types of certificates are available:

- Personal digital ID or personal certificate
- Server digital ID or website certificate
- Developers' digital ID

Software developers use developers' IDs. Internet Explorer and Netscape use only personal digital IDs and server digital IDs.

Personal certificates are used for sending personal information over the Internet to a website, whereby the web server requires verification of the user's identity. Personal certificates are most commonly used for the exchange of e-mails by individual users. Once the personal certificate is installed, the digital ID is bound to your e-mail address and can be used to digitally sign your e-mail and receive encrypted e-mails. Personal certificates are not seen during communication, which makes the process transparent to the user.

Website certificates enable and state that a specific web server is operating in a secure and authentic way. A web server ID or certificate unambiguously identifies and authenticates the web server and guarantees the encryption of any information passed between the web server and the individual user. For instance, when sending your personal information (credit card details) to an

online store, it is a good idea to first check the certificate of the store to ensure that your information is protected while in transit.

The different digital ID services, whether they take the form of a personal certificate or a website certificate, use key encryption techniques with two keys, namely a public key and a private key. Figure 2.3 illustrates the mechanism behind this encryption technique.

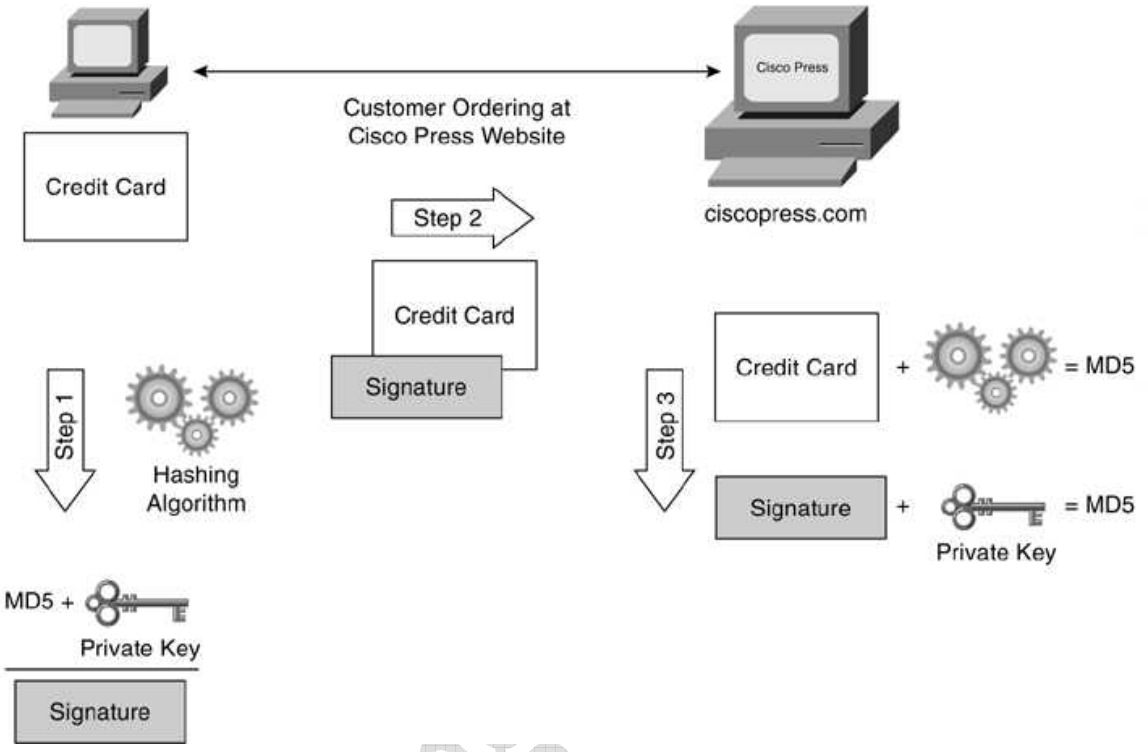


Figure 2.3: Digital ID Functionality

Only the public key is exchanged between the sender and receiver. Before actual transmission starts between two hosts, the sending host forwards its certificate, providing the public key, so the receiver can send encrypted data or information back. The information that is received back can be decrypted using the private key. The private key has two main functions. First, it makes a digital ID or signature unique, and second, it decrypts information in combination with the corresponding public key.

Let's take a closer look at this process by examining the steps shown in Figure 2.3.

Step 1. The online user passes credit card information through a hashing algorithm to produce the message digest (MD5). The message digest is then encrypted using the private key.

A message digest is a function that takes arbitrary-sized input data (referred to as a message) and generates a fixed-size output, called a digest (or hash).

Step 2. At this point, the signed data is sent to the web server.

Step 3. The server uses the same algorithm to create a message digest, decrypts the signature using the public key (added to the signature), and compares the two message digests. When the two message digests are equal, identity is checked and secure transmission can occur.

2.3. Intrusion Detection System

Digital IDs protect the integrity of your data end to end. In contrast, intrusion detection systems (IDSs) detect and prevent intrusions into your systems. IDSs are often referred to as intrusion protection systems or intrusion prevention systems.

Intrusion is when someone tries to break into, misuse, or exploit your system. Intrusion detection is the technology used to detect whether someone is trying to exploit a system.

Although the majority of intrusion attempts actually occur from within an organization and are usually perpetrated by insiders, the most common security measures protect the inside network from the outside world. Outside intruders are often referred to as crackers.

Mechanisms are required to continuously detect both inside and outside intrusions. IDSs have proved to be effective solutions for both inside and outside attacks. These systems run constantly in your network, notifying network security personnel when they detect an attempt considered suspicious. IDSs have two main components: IDS sensors and IDS management.

IDS sensors are software and hardware used to collect and analyze the network traffic. These sensors are available in two types, network IDS and host IDS.

A host IDS is a server-specific agent that runs on a server with a minimum of overhead to monitor the operating system and applications residing on the server, such as HTTP, SMTP, and FTP.

A network IDS can be embedded in a networking device, a standalone appliance, or a module to monitor the network traffic.

IDS management, on the other hand, acts as the collection point for alerts and performs configuration and deployment services in the network.

2.4. PC Card Based Solutions

To establish a network environment that is secured in depth, you can add PC card based solutions to digital IDs and IDSs. A couple of PC card based solutions are available to protect your data in today's challenging network environment. These PC card based solutions enable the network administrator to add security to the control of access, identities, software, file storing, e-mails, and so on. Security cards or smart cards, hardware keys and PC card encryption cards are most commonly used. The following sections discuss all three in a little more detail.

2.4.1. Security Cards

Security cards (often referred to as smart cards) are credit card sized plastic cards embedded with an integrated circuit chip (IC). Smart cards can be used for a broad range of applications and purposes that require security protection and authentication because all the information is stored on the card itself. Once the card is programmed, it no longer depends on external resources. This independence makes it highly resistant to attacks. Functional examples of smart cards are the following:

- Identification cards (including biometrics)
- Medical cards
- Credit and debit cards
- Access control cards (authentication)

All these applications require sensitive data to be stored in the card, such as biometrics, cryptographic keys, medical history, PIN codes, and so on.

Let us now focus a little more on smart card applications in the computer networking environment.

NOTE

Token-based authentication systems usually display numbers that change over time. The authentication systems synchronize with an authentication server on the network, and they may also

use a challenge/response scheme with the server. Tokens are based on something you know (a password or PIN) and something you have (an authenticat or the token).

Token-based authentication systems are increasing in popularity over software-only encryption packages mainly because of the enhancements and add-on functionality that token-based systems offer. Smart cards are seen as a rising trend in token-based authentication. Nowadays, most security functions reside on vulnerable servers. These functions can include boot integrity, file system integrity, public key encryption, key storage, and digital signatures, as you will see throughout the course of this book. By adding smart cards into your security design implementation, some mission-critical security functionality can be performed on the card itself, with significantly greater security protection and lower risk. On the other hand, smart cards are not cheap and can have potential management issues, including the need for replacement and reprogramming.

A good example of porting some of this functionality to the smart card is the protection of the boot sectors on a hard drive of a personal computer. Most users don't even worry about protecting these system areas, although they are exposed and vulnerable to computer virus infection. The basic idea is that during the boot sequence and after the user has been authenticated to the smart cards, the computer requires data from the smart card to complete the booting process. (The smart card is also password protected.) This guarantees system integrity even if the attacker gains physical access to the computer.

Smart card deployment is also used to assure file system integrity. In general for all computer systems, validity of files, such as executable programs, is checked against a checksum. It is in this context that smart cards can be an effective protection mechanism against viruses. The smart card stores the checksum of the executable program or plain data file. When opening or launching the file, the checksum on the card is verified. If the checksum differs, an alarm goes off. This is an efficient way of validating file integrity and works as a complementary solution to virus-scanning software applications, which are scanning for known viruses using well-defined signatures.

2.4.2. Hardware Keys

Hardware keys are best known as software protection elements. They are USB-based solutions. If your company is in the software development business, you are most likely aware of software pirates and hackers trying to gain free access to the software your company has developed in-house. Hardware keys protect software application code and are the first line of a good defense in tackling this threat. Users cannot launch applications without the hardware key that is plugged into the laptop or workstation.

Hardware keys are also used for authentication purposes to protect against unauthorized data access. Lost and stolen laptops endanger data confidentiality and data integrity. Installing hardware keys prevents thieves from breaking into corporate servers via the laptop because hardware keys need to be plugged into the laptop for authentication of the user. The hardware key is small and handy and can be easily attached to a key ring containing other personal keys. Advantages of these solutions include ease of implementation as well as low cost.

2.4.3. PC Encryption Cards

PC encryption cards are available for USB, LPT, COM, RS232, PCMCIA, and (E)ISA. These cards can be attached as peripherals or integrated in almost any computer device. Figure 2.4 shows the setup for data encryption using PC cards. Encryption can be accomplished locally and remotely on the file server.

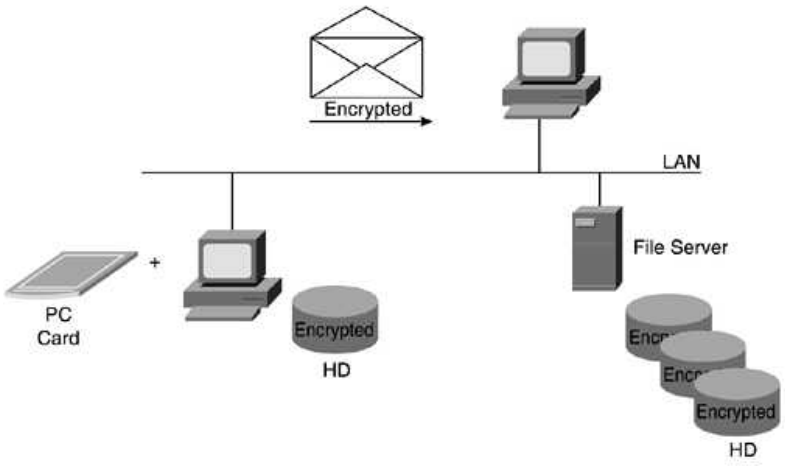


Figure 2.4: PC Encryption Cards

PC cardbased solutions using encryption cards provide secure file storage and file transmission over a LAN segment, as seen in Figure 2.4. This option can also be used to protect data within e-mails against unauthorized access during transmission. Moreover, PC cardbased solutions provide data authentication and data integrity. The user needs to install an OS driver, often a plug-in for an e-mail application (Microsoft Outlook, for instance), that uses the PC card as the encryption platform.

2.5. Physical Security

Although this book focuses mainly on the security issues of networks, physical security is also important. It is relatively easy to implement and maintain a tight security policy for your network security. Physical security is much more difficult to implement.

Physical security is defined here as the use of blueprints, standards, or models to protect networks. Physical security involves the identification and description of all the measures required to protect your facility. This process includes both internal and external security measures, disaster-recovery plans, and personnel training.

The implementation of a valid physical security plan can fall short for various reasons, the most important being budget constraints. A slight shift in focus is taking place with the recent effects and threats of global terrorism. This shift might trigger the necessary attention so that comprehensive physical security implementations become as common as encryption, firewalls, virtual private networks (VPNs), and others.

2.5.1. Outside and External Security

When implementing physical security at a company level, the first consideration is the location of your site. In reality, considering the change of a company's location might not be an option because budget limitations may force you to use an existing building. Given an existing site, however, you can make sure that the site meets a minimum set of requirements, which are defined by physical security blueprints or models.

Once a facility is built, multiple layers of security are required. The following list is an overview of available layers and options for external physical security:

- Electronic fence
- Electromagnetic IDS
- Camera systems
- Entrance security (smart cards, PIN code)
- Permanent guards

In many situations, the objective of achieving maximum external physical security, according to the specifications in the preceding list, is compromised because not all layers can be easily implemented.

2.5.2. Internal Security

The approach to implementing internal physical security is similar to the approach to implementing external physical security. Some of the external and internal measures overlap. For instance, camera systems can be installed all over a campus, with priority given to the entrances to mission-critical areas such as lab space, communication rooms, and server rooms. Just as the effectiveness of external security depends on layers of security, internal security is implemented in layers. For example, low-security areas may require only a pin code or card reader for entrance, and high-level security areas may require card readers in combination with biometrics for entrance. High-level security areas can also be equipped with smoke, temperature, and humidity sensors.

It is also important to think about the physical access to devices. Having terminals available to connect to console ports of routers and switches makes it possible to alter configurations fairly easily. In general, avoid console access to any platform in your labs, server rooms, and communication rooms. Console authentication should be configured if physical console access is required to assure that unauthorized console access is avoided.

2.5.3. Disaster-Recovery Plans

Even for the most protected and secure areas, a decent disaster-recovery plan needs to be defined. A disaster-recovery plan spells out measures that limit losses that can be incurred by disasters such as hurricanes, floods, and electrical failure. Disaster-recovery plans also outline how business practices are to be resumed after disaster. The possibility of things going wrong needs to be addressed upfront. For instance, uninterruptible power supplies (UPSs) are the de facto standard for countering power blackouts. In addition, implementation of multiple Internet connections is a must for connecting your site to a service provider's network. Having only a single connection creates a single point of failure. Furthermore, a central backup system is a mandatory service for all servers in the network.

The industry has developed three levels of disaster-recovery plans:

Hot site: This is the most sophisticated and expensive type of data replication routine. Data is replicated on two separate servers, one housed in the operational location and one at a different physical site. Data is updated on both systems simultaneously.

Warm site: With this solution, the data replication routine can occur from once every 24 hours to once a week. In the event of a disaster, the warm site would provide day-old data.

Cold site: This solution is the most cost effective because companies do not have to purchase duplicate machines. Data is sent either on tape or via the Internet and installed on shared hardware.

The ultimate disaster-recovery service is the implementation of a complete fail-over site. This is a drastic approach. When defining the disaster-recovery plan, companies need to consider not just the loss of data but also the loss of a complete workplace. This might sound ridiculous, but the cost of losing your complete workplace, data included, is nothing compared to installing a fail-over site.

2.5.4. Personnel Awareness

Developing a strong security policy helps to protect your resources only if all staff members are properly instructed on all facets and processes of the policy. Most companies have a system in place whereby all employees must sign a statement confirming that the policy was read and understood. This policy covers the multiple security situations that employees encounter during a day of work: laptop security, password policy, handling of sensitive information, access levels, photo IDs, PIN codes, and so on. A top-down approach is required if the policy is to be taken seriously. This means that the security policy needs support from the executive level downward.

The security policy can be experienced as cumbersome by many employees, but it can have multiple advantages as well. If a high-level manager asks you to bend the rules, you can point to the security policy that says that both of you will be disciplined if you acquiesce.

As far as physical security goes, many standards and blueprints exist, but implementation costs often require compromises. Only serious attacks, intrusions, loss, or the latest threats of global terrorism can make the implementation of the physical security policy a priority.

2.6. Encrypted Login

Similar to PC card based solutions and digital IDs, encrypted logins are critical in guaranteeing confidentiality, integrity, and authentication of data for remote connectivity across the Internet. Encrypted login sessions play a significant role in assuring that all three of these requirements are met.

2.6.1. Secure Shell Protocol

Secure Shell (SSH) login sessions can be used for securing remote Telnet sessions and remote logins. The SSH protocol is used to secure connections by encrypting data such as passwords, command-line entries, debug output, or even binary files. This section focuses solely on SSH as a protocol that provides a secure, remote connection to a Cisco IOS router.

Imagine an administrator logging in to the remote router with IP address 10.10.10.1. Figure 2.5 illustrates this remote login.

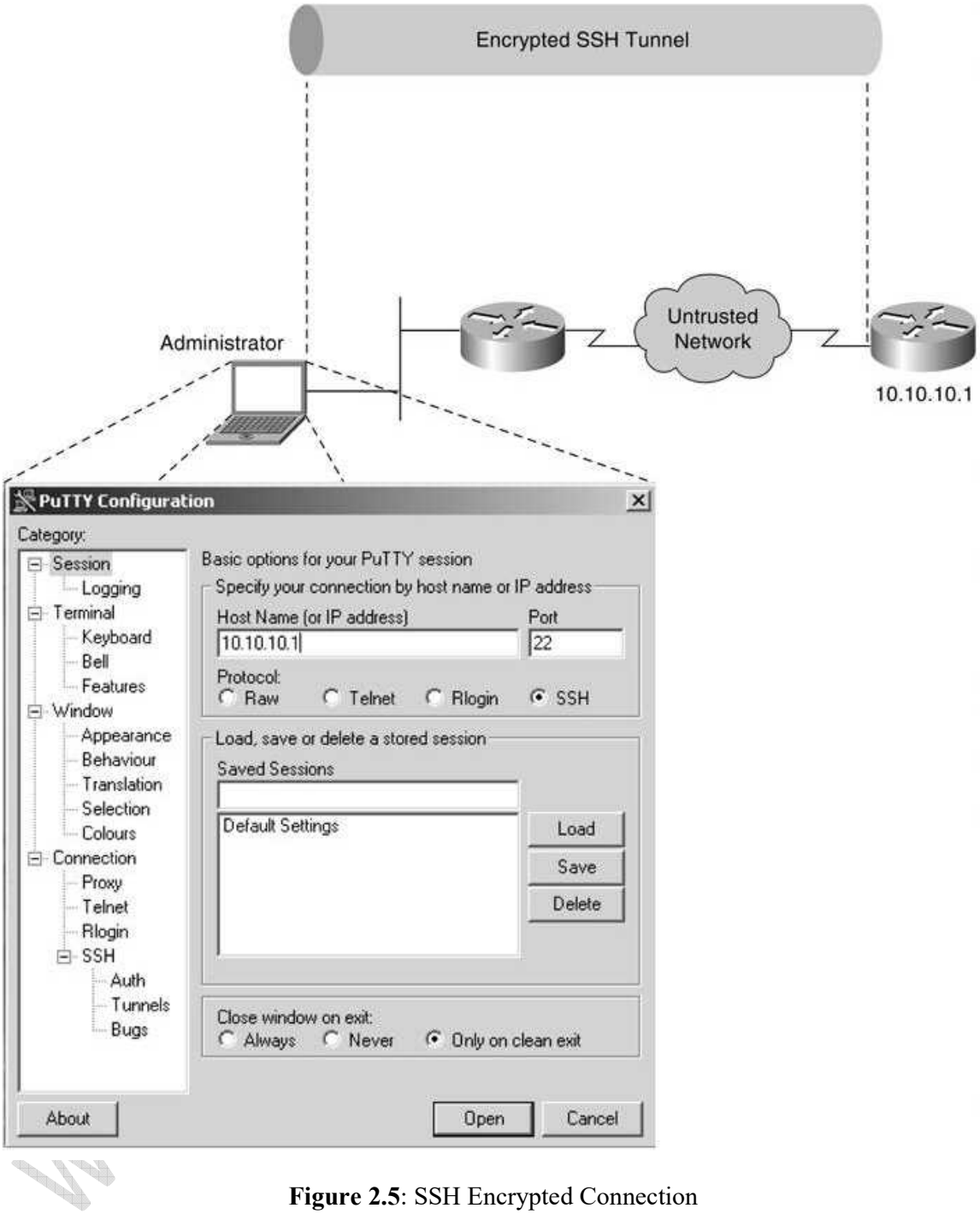


Figure 2.5: SSH Encrypted Connection

This is a client-server setup in which the Cisco IOS router is a SSH Server and the administrator's laptop is the SSH client. The SSH server in Cisco IOS works with publicly and commercially available SSH clients. A shareware application, PuTTY, is used just for this example. The connection between the SSH client (laptop) and the SSH server (Cisco IOS router) is similar to that of an inbound Telnet session, except that the connection is encrypted. Using authentication and encryption, the SSH client allows for secure communication over an insecure medium. There are two versions of SSH available, SSH Version 1 and SSH Version 2.

2.6.2. Kerberos Encrypted Login Sessions

A Kerberos Encrypted login session provides an alternative approach to SSH-encrypted login, whereby a trusted third-party authentication mechanism verifies the identity of the users. Kerberos is designed to ensure strong authentication in client-server scenarios by using secret key cryptography. SSH provides encrypted authentication as well as encrypted data transmission (sessions) end-to-end. Kerberos provides encrypted authentication only.

2.6.3. Secure Socket Layer (HTTP versus HTTPS)

HTTP is non-secure, and HTTPS is Secure Socket Layer (SSL) secured. As discussed in the first section of this chapter, digital IDs use HTTPS, whereby the data sent is encrypted and cannot be decrypted without the private key. In HTTP, the information is sent in plain text and is insecure. The main difference is this: HTTP has no encryption, and HTTPS uses the public/private key system for authentication.

SSL was originally developed by Netscape Communications to allow secure access of a browser to a web server. Nowadays, SSL has become the standard for web security. With the increasing number of high-availability, HTTPS-based transactions, the Cisco SSL products (content switches and standalone SSL appliances) simplify the support responsibilities for the website administrator. SSL-enabled websites provide a strong sense of confidentiality, message integrity, and server authentication to users who are using encrypted logins.

2.7. Firewalls

Numerous tools, techniques, systems, services, and processes are available to protect your data in today's challenging network environment. Firewalls are particularly important strategic elements at the core of the security policy implementation. Figure 2.6 shows a firewall as a device that separates different functional areas of a network. These functional areas are often referred to as secure areas. In general, these functional areas are private networks, public networks, and demilitarized zone (DMZ) networks.

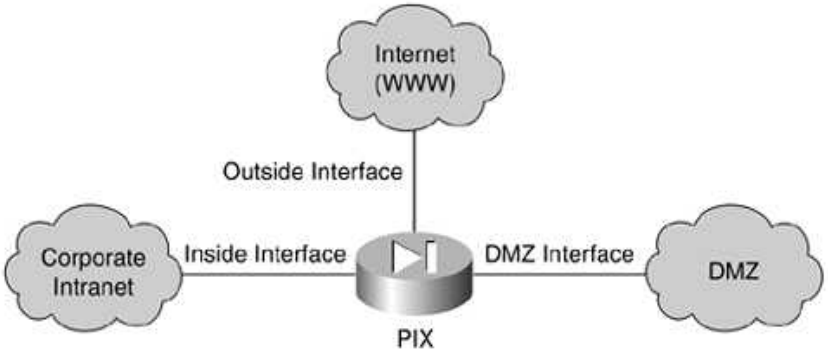


Figure 2.6: Firewall Placements

Cisco Press's Dictionary of Internetworking Terms and Acronyms defines a firewall as "a router or access server, or several routers or access servers, designed as a buffer between any connected public networks and private network. A firewall router uses access lists and other methods to ensure the security of the private network."

As shown in Figure 2.6, the inside interface of the PIX is connected to a private or corporate intranet. The outside interface is connected to the Internet (untrusted network). The DMZ is an isolated network hosting web servers and mail servers.

2.8. Reusable Passwords

User authentication for access control systems is accomplished using username and password combinations or PIN codes. These passwords are referred to as reusable passwords in security jargon. This system has been in use for many years and will probably continue for many years to come. Some alternatives to reusable passwords are discussed in the course of this chapter

and in other chapters of this book because the mechanism hasn't kept pace with the introduction of new features, tools, and techniques in the computing technologies industry.

2.8.1. Weaknesses

The list of disadvantages and weaknesses of reusable passwords is long. Statistics have proven that many users have a tendency to pick weak passwords. Also, experience tells us that users can easily violate the security rules defined in the password security policy. For instance, employees share passwords with colleagues for various reasons. Many passwords do not conform to the password security policy. Passwords can violate the following security policy requirements:

- Users select obvious passwords.
- Password length requirements are violated.
- Password lifetime requirements are violated.
- Use of characters and character classes are violated (uppercase, lowercase, numbers, punctuation).

The fact that passwords or PIN codes can be used more than once is an inherent weakness that cannot be solved without considering new technologies.

A few enhancements can be used to improve the security of reusable passwords. Developing and implementing standards and policies can result in a better understanding and awareness of the weaknesses inherent in reusable passwords. There has been a recent increase in commercially available alternative authentication mechanisms such as challenge/response and time-synchronized mechanisms, tokens, and biometrics.

2.8.2. Sample Password Policy

The following list is a sample password policy providing users of computer systems with the necessary minimum criteria for password-related information:

- Password length Eight characters or more
- Character classes Upper- and lowercase letters
- Characters Mix of numbers, symbols, and letters
- Grammar check No dictionary or jargon words
- Recurrence No use of the same character more than twice

2.9. Antivirus Software

A computer virus can be best described as a small program or piece of code that penetrates into the operating system, causing unexpected and negative events to occur. A well-known example is a virus, SoBig. Computer viruses reside in the active memory of the host and try to duplicate themselves by different means. This duplication mechanism can vary from copying files and broadcasting data on local-area network (LAN) segments to sending copies via e-mail or an Internet relay chat (IRC). Antivirus software applications are developed to scan the memory and hard disks of hosts for known viruses. If the application finds a virus (using a reference database with virus definitions), it informs the user. The user can decide what needs to happen next. Figure 2.7 illustrates the action decisions that can be made using McAfee Antivirus software applications.

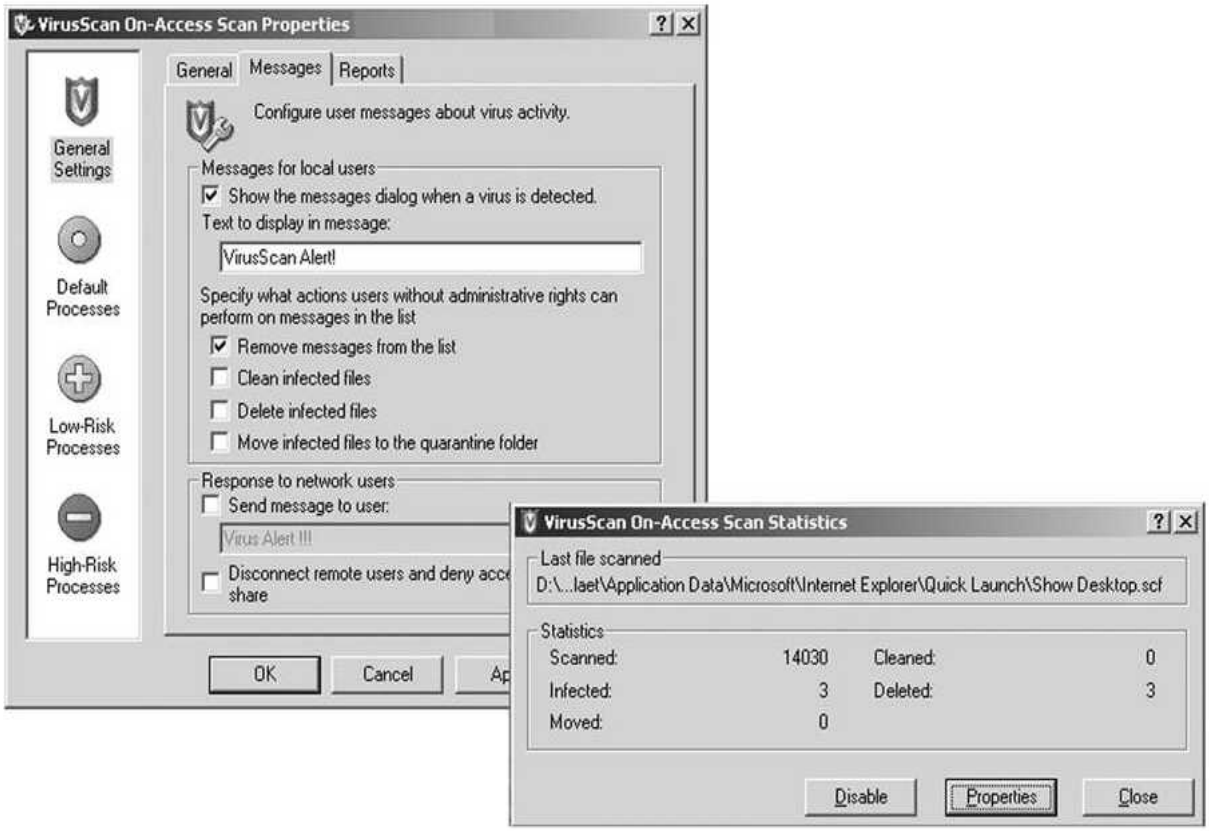


Figure 2.7: Antivirus Software Scan

The user can choose from three options: delete the file, clean the file, or place the file in a so-called quarantine folder.

Before making a decision on what antivirus software package to purchase, it is important to understand which solution best protects your organization and which antivirus software features match your needs. The following list can be used when making a comparison matrix for different solutions:

- Purchase price
- Ease of use
- Identification of viruses and worms: real-time scanning, manual, and scheduler
- Activity reporting mechanism
- Actions: deleting, cleaning, and quarantine
- Virus definition update mechanism: auto or manual definition updates
- Central management
- Operating system support
- Technical support

With the introduction of new viruses almost every day, it is hard to tell which antivirus package is best suited for your needs. Also, the installation of antivirus software should be seen as only part of your overall security solution and does not guarantee complete protection.

2.10. Encrypted Files

Another technique that can be used to protect and preserve the integrity of the data locally on your workstation is file encryption. The file encryption feature encrypts your data when it is

written to the disk. This data encryption process happens on-the-fly when data is saved and goes unnoticed by the users.

File encryption was introduced with NT File System for Windows NT (NTFS). Compared with FAT and FAT32, NTFS has a strong focus on security because an encryption file system (EFS) was one of the added security features. File encryption is linked to individual user accounts. Files encrypted by a user are accessible only from that user's account. Other users (apart from the administrator) have no access to these files because they are encrypted with individual keys. Special caution needs to be used for data recovery because related certificates with public and private keys need to be restored as well.Figure 3-8 illustrates how to enable this feature for Windows 2000.

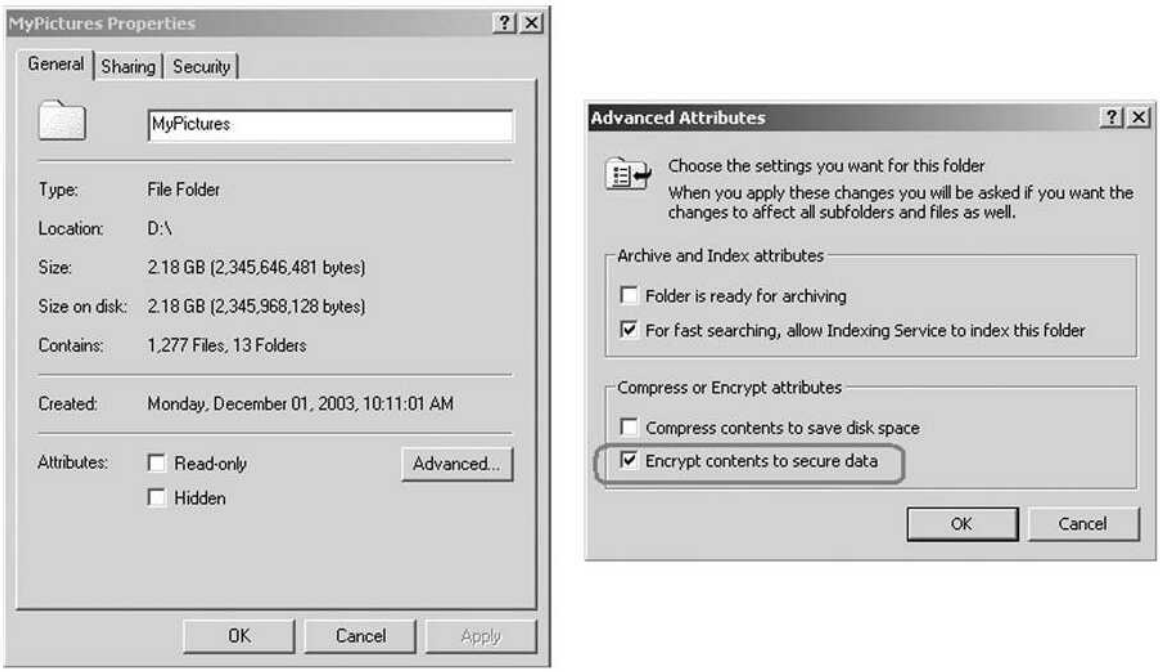


Figure 2.8: Enable NTFS File Encryption

To encrypt a file or a complete directory, you right-click on the icon. Select Properties from the options and click Advanced. This opens the Advanced Properties window. Select Encrypt contents to secure data.

NOTE

In the context of file encryption, it is worth mentioning file protection. File protection using passwords is an easy-to-implement security defense. It can be implemented in a number of ways, such as requiring passwords to open files to prevent unauthorized users from accessing the data or requiring passwords just to modify a file. This allows everyone to open the file, but only authorized users are permitted to make changes, and unauthorized access is prevented.

2.11. Biometrics

Biometrics is the science of measuring a unique physical characteristic about an individual as an identification mechanism. A number of widely used biometric technologies and techniques exist. These techniques are deployed in new network design to secure the network environment even better. The most common biometric technologies are fingerprint scanning and voice recognition. This section briefly touches on other technologies such as face recognition (iris and retina), typing biometrics, and signature recognition.

Biometric access methods for computer systems are gaining popularity because of governmental and corporate businesses' increased focus on security. Numerous commercial products

are already available, and the future will inevitably see all portable devices, access doors, and so on being biometrically protected. The integration of biometrics in your security policy will provide a solid foundation for developing a secure environment.

2.11.1. Fingerprint Scanning

Fingerprint scanning is probably the most widely used biometric technology. As everyone knows, the fingertips of each individual have unique characteristics. These characteristics vary from the geometry to the pattern and size of the ridges. Picture how the ridges of the fingertip generate a fingerprint. Fingerprint scanners can read the fingerprint and convert it into a digital representation. The digital copy is checked against an authorized copy stored on the central computer system.

Although this technology may seem sophisticated, it has a few drawbacks. For instance, the system can be cheated because it cannot determine if a fingerprint was made by a live user or was copied. If you are starting to deploy biometrics in your environment, consider commercially available computer keyboards with integrated fingerprint scanners. These are excellent and relatively cheap options.

2.11.2. Voice Recognition

Voice recognition, sometimes referred to as speech analysis, is based on vocal characteristics. Just as with fingerprints, each individual voice has unique characteristics. A few instruments and techniques are available most common is the microphone in combination with speech analysis applications. The purpose of all voice recognition systems is to depict the speech signal in some way and to capture and store its characteristics on a computer system. Again, these characteristics are checked against an authorized copy stored on the central computer system.

2.11.3. Typing Biometrics

Typing biometrics examine the characteristic typing techniques of computer users. Some known characteristics are as follows:

- Speed
- Patterns
- Force
- Keystroke duration
- Inter-keystroke latency (latency between the first and second keystroke)
- Error frequency

In general, typing biometric techniques are used when users type in their passwords during a login process. It is good practice when implementing this technology to set up a system in which deviation from the reference data in one or more of these characteristics requires further authentication or second-level authentication of the user by other authentication technologies.

2.11.4. Face Recognition

Just as with other recognition techniques, face recognition uses certain parameters and characteristics to reveal an individual's identity.

Since September 11, 2001, discussion on the subject of using biometrics has increased, specifically about face recognition at airports to identify known terrorists crossing borders. The U.S. Department of Defense is involved in the development of a facial recognition technology program called FERET.

2.11.5. Signature Recognition

Signature identification systems analyze individual signatures based on factors such as speed, acceleration, velocity, pen pressure, and stroke length.

Newer biometric measurements include techniques for DNA comparisons, which will be refined in the years to come.

2.12. Conclusion

Network security is an important concern that must be seriously deliberated. This chapter explained digital IDs and how they can protect the network. Intrusion protection and intrusion prevention techniques, as well as PC card-based solutions, can counter weaknesses with different encryption techniques to protect the network environment. Physical security of the site can be achieved using access control and biometric techniques. Antivirus software and firewalls are other technologies used to protect your network environment.

2.13. Important Questions:

1. List four parameters of a digital ID.
2. A host IDS can be embedded in a networking device, a standalone appliance, or a module monitoring the network traffic. True or False?
3. What processes are covered in physical security policies?
4. List two protocols that can be used for encrypted logins.
5. Which three functional areas can be connected to a firewall?
6. What is file encryption?
7. List four of the most common biometric technologies.

Chapter 03
CRYPTOGRAPHY

3.1. Introduction:

`

Cryptography is a science of encoding the information in unintelligible fashion to conceal information so that only intended user may be able to get access to the information. It provides the security to data from various attacks in the channel, i.e. active attacks, passive attacks. The main intention of cryptography is ensuring data security form unauthorized persons.

The data before encryption is called plain text and encrypted form of data is called cipher text. Any kind of data, text, number, image, audio or video, is manipulated by an encryption algorithm mostly using a key to convert it into scrambled form. This process is reversed on receiver side to get original text using some algorithm same or different key. Working of a crypto system is illustrated in Fig.3.1.

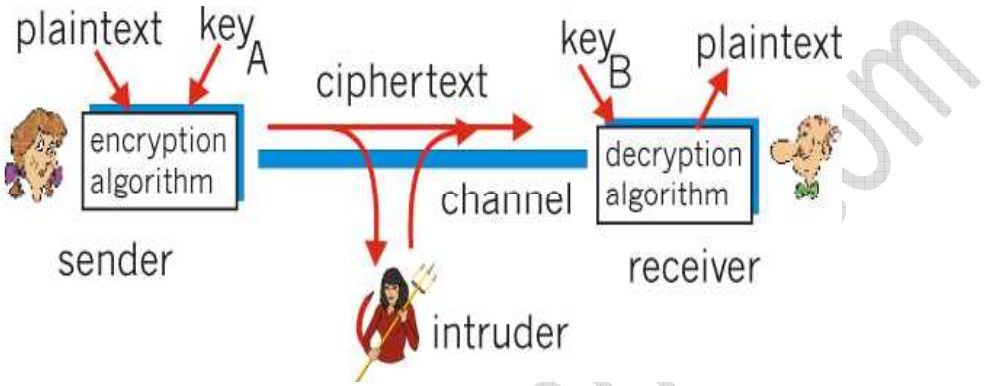


Figure 3.1: Cryptography process

An illustrative example is provided in Fig. 2. Here sender is intended to send, "Hello Alice" to his friend. His message is converted to cipher text by encryption algorithm using a specific key 'A'. Now the scrambled form of the message is "x0Ak3o$2Rj". This scrambled data is transmitted to the receiver, on receiver side data is decrypted using the same key 'A' and show to the Alice as the real message sent by the sender.

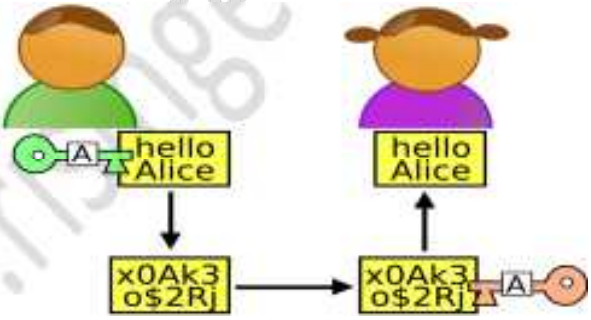


Figure 3.2: Illustrative example

Cryptographic techniques can be categoriesed in three broad categories. Symetric , asymetric , hashing. All or algorithms fall in these categories have their own strenghts and weaknesses.

3.1.1. Cryptanalysis

Cryptanalysis is the reverse of cryptography. It is the science of cracking codes, decoding secrets, and in general, breaking cryptographic protocols. It is a process of attempting to discover the plaintext or key. To design a robust encryption algorithm, one should use cryptanalysis to find and correct any weaknesses.

The various techniques in cryptanalysis that attempt to compromise cryptosystems are called attacks. The strategy used by the cryptanalyst depends on the nature of the encryption scheme and the information available to the cryptanalyst. A cryptanalyst starts from the decoded message. The cryptanalyst then tries to get this message back into its original form without knowing anything of that original message. This kind of attack is called a cipher-text-only attack. The data that a cryptanalyst needs for this attack is fairly easy to obtain, but it is very difficult to successfully recover the original message.

3.2. History of Cryptography:

3.2.1. Manual Systems

Cryptography dates as far back as 1900 B.C., when a scribe in Egypt first carved a derivation of the standard hieroglyphics on clay tablets. Early Indian texts such as the Kama Sutra used ciphers that consisted mostly of simple alphabetic substitutions often based on phonetics. This is somewhat similar to "pig latin" (igpay atinlay), in which the first letter is placed at the end of the word and is followed by the sound "ay."

3.2.1.1.Scytale:

The ancient Greeks and the Spartans in particular, are said to have used this cipher to communicate during military campaigns. Sender and recipient each had a cylinder (called a scytale) of exactly the same radius. The sender wound a narrow ribbon of parchment around his cylinder, and then wrote on it lengthwise. After the ribbon is unwound, the writing could be read only by a person who had a cylinder of exactly the same circumference.

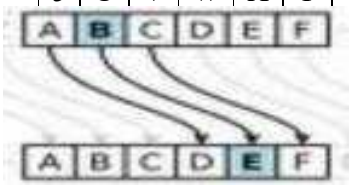
3.2.1.2.Polybius Square:

Another Greek method was developed by Polybius (now called the "Polybius Square"). Each letter is represented by its coordinates in the grid. For example, "BAT" becomes "12 11 45". Developed for telegraphy e.g. pairs of torches

	1	2	3	4	5
1	A	B	C	D	E
2	F	G	H	I	J
3	K	L	M	N	O
4	P	Q	R	S	T
5	U	V	W	X	Y

3.2.1.3.Caesar Cipher:

The Romans knew something of cryptography (e.g., the Caesar cipher and its variations). The method is named after Julius Caesar, who used it to communicate with his generals. The Caesar Cipher is an example of what is called a shift cipher. To encode a message, letters are replaced with a letter that is a fixed number of letters beyond the current letter.



3.2.1.4.Atbash cipher:

Later still, Hebrew scholars made use of simple mono alphabetic substitution ciphers (such as the Atbash cipher) beginning perhaps around 500 to 600 BC. The Atbash cipher is a very specific case of a substitution cipher where the letters of the alphabet are reversed. In other words, all As are replaced with Zs, all Bs are replaced with Ys, and so on.

3.2.2. Crypto Machines

Thomas Jefferson invented a wheel cipher in the 1790s that was used during World War II with only slight modification. The wheel cipher consisted of a set of wheels, each with random orderings of the letters of the alphabet.

In 1844, the development of cryptography was dramatically changed by the invention of the telegraph. Communication with the telegraph was by no means secure, so ciphers were needed to transmit secret information. Just as the telegraph changed cryptography, the radio changed it again in 1895. Now transmissions were open for anyone's inspection, and physical security was no longer possible.

During World War II, most German codes were predominantly based on the Enigma machine. A British cryptanalysis group first broke the Enigma code early in World War II. Some of

the first uses of computers were for decoding Enigma ciphers intercepted from the Germans. The sidebar on the Enigma machine is somewhat detailed, but it gives you an idea of the complexity of mechanical operations that were later replaced by computer processes.

The Enigma Machine

The Enigma machine was a simple cipher machine. It had several components such as a plug board, a light board, a keyboard, a set of rotors, and a reflector (half rotor). The first Enigma machine looked very similar to a typewriter. The machine had several variable settings that could affect the operation of the machine. First, the user had to select three rotors from a set of rotors. A rotor contained one-on-one mappings of all the letters. Another variable element to this machine was the plug board. The plug board allowed for pairs of letters to be remapped before the encryption process started and after it ended.

When a key was pressed, an electrical current was sent through the machine. The current first passed through the plug board, then through the three rotors, then through the reflector, which reversed the current back through the three rotors and then the plug board. Then the encrypted letter was lit on the display. After the display was lit, the rotors rotated. The operation of the rotors was similar to that of an odometer, where the rotor farthest to the right must complete one revolution before the middle rotor rotates one position and so on.

In order to decrypt a message, the receiver needed the encrypted message as well as knowledge of which rotors were used, the connections on the plug board, and the initial settings of the rotors. To decrypt a message, the receiver set up the machine to be identical to the way the sender initially set it up and then typed in the encrypted message. The output of typing in the encrypted message was the original message. Without the knowledge of the state of the machine when the original message was typed in, it was extremely difficult to decode a message.

3.2.3. Computers

By 1948, cryptographers started to use advanced mathematical techniques to calculate ciphers and to prevent computers from unscrambling the ciphers. Symmetric and asymmetric key algorithms were developed to this end. A symmetric key algorithm uses the same key to encrypt and decrypt a message, whereas an asymmetric key algorithm uses two different keys.

3.3. Cryptography from Muslim History (Medieval Cryptography):

Al- Kindi, wrote a book on cryptology, the "Risalah fi Istikhraj al-Mu'amma" (Manuscript for the Deciphering Cryptographic Messages), circa 850CE. This book apparently antedates Western European cryptography works by 300 years and predates writings on probability and statistics by Pascal and Fermat by nearly 800 years. He was a pioneer in cryptanalysis and cryptology, and devised new methods of breaking ciphers, including the frequency analysis method. In his book Al-Kindi described the first cryptanalysis techniques, including some for polyalphabetic ciphers, cipher classification, Arabic phonetics and syntax, and, most importantly, gave the first descriptions on frequency analysis. He also covered methods of encipherments, cryptanalysis of certain encipherments, and statistical analysis of letters and letter combinations in Arabic. **Cryptography in the Renaissance Period:**

Essentially all ciphers remained vulnerable to the cryptanalytic technique of frequency analysis until the development of the polyalphabetic cipher, and many remained so thereafter. The polyalphabetic cipher was most clearly explained by Leon Battista Alberti around the year 1467, for which he was called the "father of Western cryptology".

3.4. Key Concepts in Cryptography:

3.4.1. Confusion

Confusion pertains to making the relationship between the key and resulting ciphertext as complex as possible so the key cannot be uncovered from the ciphertext Each ciphertext value should depend upon several parts of the key, but this mapping between the key values and the ciphertext values should seem completely random to the observer.

3.4.2. Diffusion

Diffusion (transposition) means that a single plaintext bit has influence over several of the ciphertext bits. Changing a plaintext value should change many ciphertext values, not just one. In fact, in a strong block cipher, if one plaintext bit is changed, it will change every ciphertext bit with the probability of 50 percent. This means that if one plaintext bit changes, then about half of the ciphertext bits will change.

3.4.3. Block Cipher:

A block cipher divides a message into blocks of bits. These blocks are then put through mathematical functions, one block at a time. Suppose you need to encrypt a message you are sending to your friend and you are using a block cipher that uses 64 bits. Your message of 640 bits is chopped up into 10 individual blocks of 64 bits. Each block is put through a succession of mathematical formulas, and what you end up with is 10 blocks of encrypted text. You send this encrypted message to your friend. He has to have the same block cipher and key, and those 10 ciphertext blocks go back through the algorithm in the reverse sequence and end up in your plaintext message. A strong cipher contains the right level of two main attributes: confusion and diffusion. Confusion is commonly carried out through substitution, while diffusion is carried out by using transposition. For a cipher to be considered strong, it must contain both of these attributes, to ensure that reverse-engineering is basically impossible. The randomness of the key values and the complexity of the mathematical functions dictate the level of confusion and diffusion involved.

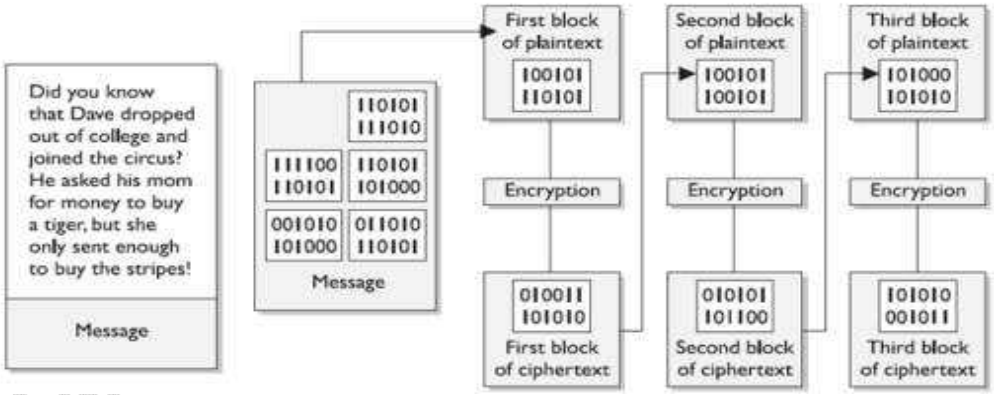


Figure 3.3: Block Cipher

3.4.3.1.Example Of A Block Cipher

- Block ciphers use diffusion and confusion in their methods. The Figure shows a conceptual example of a simplistic block cipher. It has four block inputs, and each block is made up of four bits
- The block algorithm has two layers of four-bit substitution boxes called *S-boxes.*
- *Each S-box contains a lookup table used by the algorithm as* instructions on how the bits should be encrypted

A message is divided into blocks of bits, and substitution and transposition functions are performed on those blocks

S-Boxes

The Figure 3.4 shows that the key dictates what S-boxes are to be used when scrambling the original message from readable plaintext to encrypted non-readable cipher text

Each S-box contains the different substitution methods that can be performed on each block. This example is simplistic most block ciphers work with blocks of 32, 64, or 128 bits in size, and many more S-boxes are usually involved

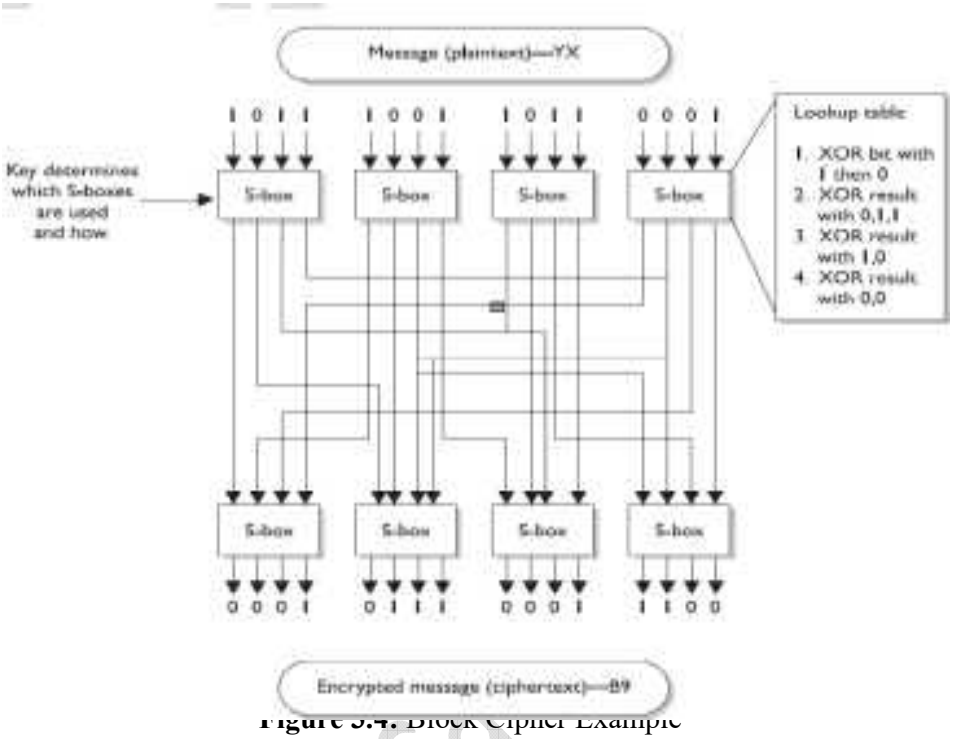


Figure 3.4: Block Cipher Example

3.4.4. Stream Ciphers

A stream cipher treats the message as a stream of bits and performs mathematical functions on each bit individually. A stream cipher does not divide a message into blocks. When using a stream cipher, a plaintext bit will be transformed into a different ciphertext bit each time it is encrypted. Stream ciphers use keystream generators, which produce a stream of bits that is XORed with the plaintext bits to produce ciphertext, as shown in the Figure 3.5.

With stream ciphers, the bits generated by the keystream generator are XORed with the bits of the plaintext message.

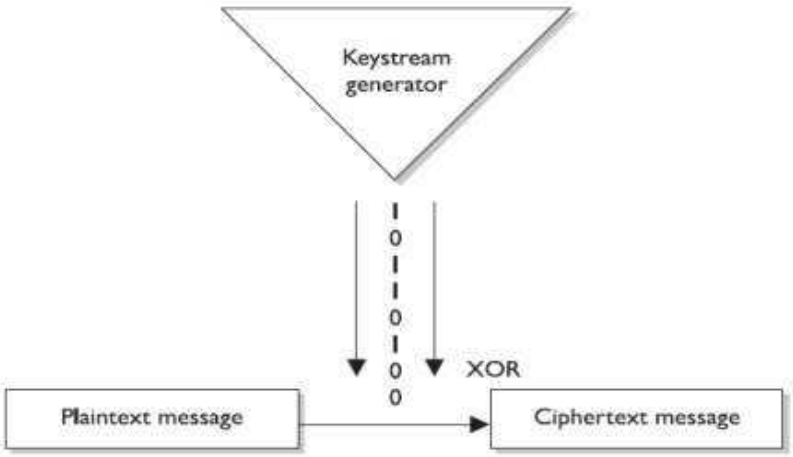


Figure 3.4: Stream Cipher

3.5. Modern-Day Techniques

3.5.1. Encryption

3.5.1.1. Symmetric Key Algorithms:

Symmetric encryption, also referred to as conventional encryption, secret-key, or single-key encryption, was the only type of encryption in use prior to the development of public-key encryption in the late 1970s.

In the symmetric key algorithms, both encryption and decryption process used identical key. In these types of cryptographic algorithms, security of key must be ensured up to any extent and the key must be kept private. If an intruder may get the key, he can easily decrypt the information. The symmetric key encryption further may take place into two approaches, Block cipher and stream cipher. The pros of using symmetric key encryption are that it requires small amount of computational resources and time relatively. Mostly keys used for this process are kept unrivaled or there may is a simple makeover between the two keys. The cons of algorithms in this category are that the key used for encryption must be kept private. So security of key is most challenging issue. Moreover, both sender and receiver must be agreed on a common private key.

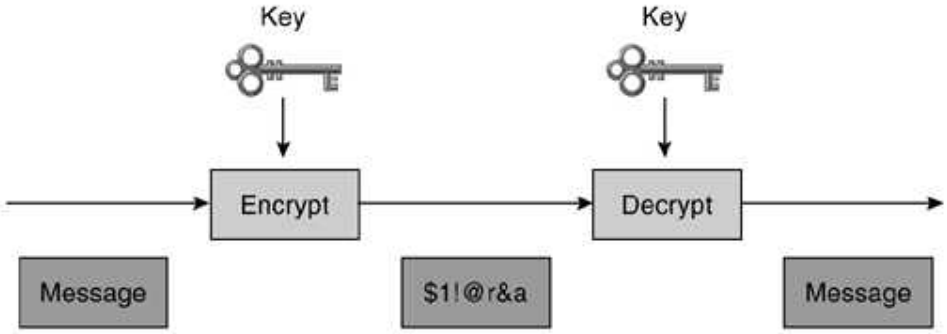


Figure 3.5: Symmetric Key Algorithm illustration

A **symmetric encryption** scheme has five ingredients.

1. **Plaintext:** This is the original message or data that is fed into the algorithm as input.
2. **Encryption algorithm:** The encryption algorithm performs various substitutions and transformations on the plaintext.
3. **Secret key:** The secret key is also input to the algorithm. The exact substitutions and transformations performed by the algorithm depend on the key.
4. **Ciphertext:** This is the scrambled message produced as output. It depends on the plaintext and the secret key. For a given message, two different keys will produce two different ciphertexts.
5. **Decryption algorithm:** This is essentially the encryption algorithm run in reverse. It takes the ciphertext and the same secret key and produces the original plaintext.

There are two requirements for secure use of symmetric encryption:

1. We need a strong encryption algorithm. At a minimum, we would like the algorithm to be such that an opponent who knows the algorithm and has access to one or more ciphertexts would be unable to decipher the ciphertext or figure out the key. This requirement is usually stated in a stronger form: The opponent should be unable to decrypt ciphertext or discover the key even if he or she is in possession of a number of ciphertexts together with the plaintext that produced each ciphertext.

2. Sender and receiver must have obtained copies of the secret key in a secure fashion and must keep the key secure. If someone can discover the key and knows the algorithm, all communication using this key is readable.

It is important to note that the security of symmetric encryption depends on the secrecy of the key, not the secrecy of the algorithm. That is, it is assumed that it is impractical to decrypt a message on the basis of the ciphertext *plus* knowledge of the encryption/decryption algorithm. In other words, we do not need to keep the algorithm secret; we need to keep only the key secret.

This feature of symmetric encryption is what makes it feasible for widespread use. The fact that the algorithm need not be kept secret means that manufacturers can and have developed low-cost chip implementations of data encryption algorithms. These chips are widely available and incorporated into a number of products. With the use of symmetric encryption, the principal security problem is maintaining the secrecy of the key.

3.5.1.1.1. Data Encryption Standard

The Data Encryption Standard (DES) has been the worldwide encryption standard for a long time. IBM developed DES in 1975, and it has held up remarkably well against years of cryptanalysis. DES is a symmetric encryption algorithm with a fixed key length of 56 bits. The algorithm is still good, but because of the short key length, it is susceptible to brute-force attacks that have sufficient resources.

How DES Works:

- DES is a symmetric block encryption algorithm. When 64-bit blocks of plaintext go in, 64-bit blocks of ciphertext come out.
- It is also a symmetric algorithm, meaning the same key is used for encryption and decryption.
- It uses a 64-bit key: 56 bits make up the true key, and 8 bits are used for parity.
- When the DES algorithm is applied to data, it divides the message into blocks and operates on them one at a time.
- The blocks are put through 16 rounds of transposition and substitution functions.
- The order and type of transposition and substitution functions depend on the value of the key used with the algorithm.
- The result is 64-bit blocks of ciphertext.

NSA announced in 1986 that, as of January 1988, the agency would no longer endorse DES and that DES-based products would no longer fall under compliance with Federal Standard 1027.

The NSA felt that because DES had been so popular for so long, it would surely be targeted for penetration and become useless as an official standard.

In 1998, the Electronic Frontier Foundation built a computer system for $250,000 that broke DES in three days by using a brute force attack against the keyspace. It contained 1,536 microprocessors running at 40MHz, which performed 60 million test decryptions per second per chip. Although most people do not have these types of systems to conduct such attacks, as Moore's Law holds true and microprocessors increase in processing power, this type of attack will become more feasible for the average attacker. This brought about 3DES, which provides stronger protection, as discussed later in the chapter. DES was later replaced by the Rijndael algorithm as the Advanced Encryption Standard (AES) by NIST.

3.5.1.1.2. Triple Data Encryption Standard:

3DES also known as Triple DES is application of DES three times to encrypt data. It uses two approaches; one approach uses three keys and other approach use two keys. The block of plaintext p is initially encrypted with a key k1, generate p', this encrypted block p' is again encrypted with second key k2, and generate k double encrypted cipher text p'', finally apply DES on p'' with third key k3, with resultant p'''. In this algorithm, the decryption process is reverse of

encryption, to decrypt a cipher text C, the system has to perform decryption process in this way p= DK3 (DK2 (DK1(C))). Sequence of these algorithms is illustrated in Figure 3.6.

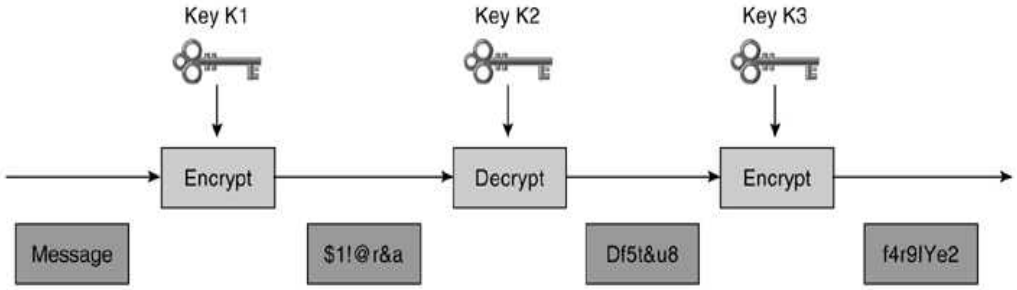


Figure 3.6: Triple DES

When a message is to be encrypted with 3DES, a method called EDE (encrypt decrypt encrypt) is used. The EDE method is described in the following list:

Step 1. The message is encrypted with the first 56-bit key, K1.

Step 2. The data is decrypted with a second 56-bit key, K2.

Step 3. The data is again encrypted with the third 56-bit key, K3.

The EDE procedure provides encryption with an effective key length of 168 bits. If keys K1 and K3 are equal (as in some implementations), a less secure encryption of 112 bits is achieved.

To decrypt the message, you must use the following procedure, which is the opposite of the EDE method:

Step 1. Decrypt the ciphertext with key K3.

Step 2. Encrypt the data with key K2.

Step 3. Finally, decrypt the data with key K1.

Encrypting the data three times with three different keys does not significantly increase security. The EDE method has to be used. Encrypting three times in a row with different 56-bit keys equals an effective 58-bit key length and not the full 128-bit, as expected.

3.5.1.1.3. Advance Encryption Standard (AES):

Rijndeal is a cryptographic algorithm designed by Vincent Rijman and Joan Daemen in Belgium NIST. It was chosen as Advanced Encryption Standard in Oct-2000. AES is a symmetric crypto algorithm. It shares an agreed upon private secret key in sender and receiver for encryption and decryption process. It uses the key lengths 128bit, 192 bits, and 256 bits and known as AES-128, AES-192 and AES-256 respectively.

Overview of AES

AES algorithm is based on S-P network. In this algorithm the, the input plaintext of 128 bit is arranged in a 4x4 matrix. This matrix is called state. AES is iterative in nature and take various iterations to process the state. Number of iterations based on length of key used, i.e. 10 iterations in case of 128 bit key, 12 iterations in case of 192 bit key and 14 iterations in case of 256 bit key[11].

A round function modifies the array at each round using four different transformations [11][12].

1. **SubByte** transformation: in this transformation each byte is substitute with another byte according to a substitution table (S-Box). S-Box is an invertible table.
2. **ShiftRows** transformation: in this step each row of the state matrix is shifted in cycle up to a many steps.
3. **MixColumns** transformation: this operation mix for bytes in a column.
4. **AddRoundKey**: this step takes XOR of each byte of the state table and round key.

Encryption and decryption takes place in various iterations. Number of iteration relies on the length of cipher key. In each iteration, some specific operations are performed. In both operations,

encryption and decryption, the first iteration perform AddRoundKey transformation on state matrix and only this iteration uses the secret key. Rest of all iterations performs all the four transformations where the final iteration doesn't perform MixColumns transformation.

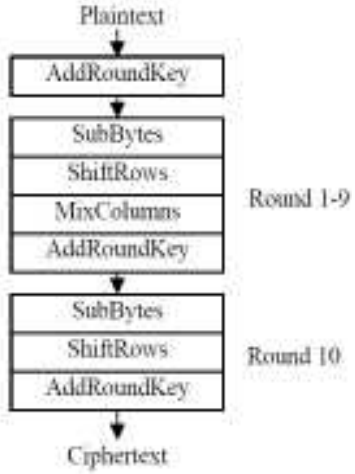


Figure 3.7: Complete Process of AES

AES is more reliable and secure algorithm. No major attack proved successful on it till now. In case of bruit force attacks, AES-128 uses the minimum length key, 128 bit. There are 2^{128} keys can be generated using 128 bits. Searching At the speed of 256 keys per second, it will take around 149 Trillion years to search all keys. So, exhaustive search is not feasible in this huge search space. Only sub channel attacks are proved a little successful against AES. Daniel Bernstein also presented a case study of such attacks on a server with OpenSSL AES.

3.5.1.2.Asymmetric Key Algorithms:

In asymmetric techniques both encryption and decryption processes use different keys. It is also called public key encryption. Due to this fact, these algorithms are relatively slow in processing and are impractical to used huge amount of data. The major pros of using public key encryption are that keys used in this technique are long and in this way increases the security level of data and also data can be transferred securely if there is no agreement between sender and receiver on a common private key. Cons of using asymmetric encryption technique are that it requires a lot of computational resources and relatively slow in processing.

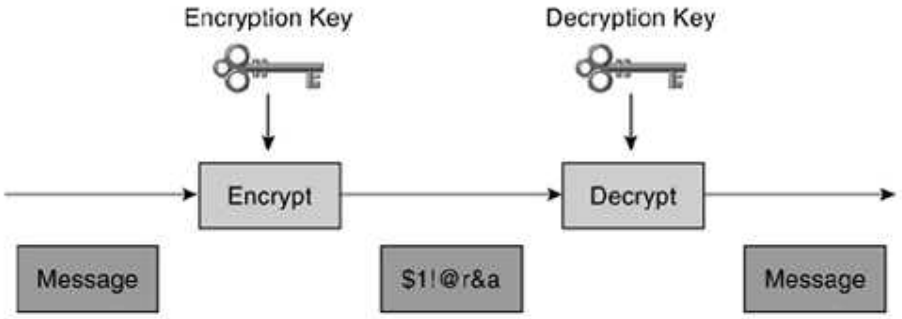


Figure 3.8: Asymmetric Key Algorithm

To understand the advantages of an asymmetric system, imagine two people, Alice and Bob, sending a secret message through the public mail.

In a symmetric key system, Alice first puts the secret message in a box and then padlocks the box using a lock to which she has a key. She then sends the box to Bob through regular mail. When

Bob receives the box, he uses an identical copy of Alice's key (which he has obtained previously) to open the box and read the message.

In an asymmetric key system, instead of opening the box when he receives it, Bob simply adds his own personal lock to the box and returns the box through public mail to Alice. Alice uses her key to remove her lock and returns the box to Bob, with Bob's lock still in place. Finally, Bob uses his key to remove his lock and reads the message from Alice.

The critical advantage in an asymmetric system is that Alice never needs to send a copy of her key to Bob. This reduces the possibility that a third party (for example, an unscrupulous postmaster) can copy the key while it is in transit to Bob, allowing that third party to spy on all future messages sent by Alice. In addition, if Bob is careless and allows someone else to copy his key, Alice's messages to Bob are compromised, but Alice's messages to other people remain secret.

Not all asymmetric algorithms operate in precisely this fashion. With the most common asymmetric algorithms, Alice and Bob each own two keys; one key cannot (as far as is known) be deduced from the other. These are called public key/private key algorithms because one key of the pair can be published without affecting the security of messages. In the preceding analogy, Bob might publish instructions on how to make a lock (a public key). But even if people followed the instructions and created a lock, it would be difficult for them to deduce from those instructions how to make a key that would open that lock (private key). To send a message to Bob, you have to use Bob's public key to encrypt the message, and Bob uses his private key to decrypt the message.

Asymmetric algorithms are designed so that the key for encryption is different from the key for decryption. The decryption key cannot be calculated from the encryption key (at least not in any reasonable amount of time) and vice versa. The usual key length for asymmetric algorithms ranges from 512 to 2048 bits.

Asymmetric algorithms are relatively slow (up to 1000 times slower than symmetric algorithms). Their design is based on computational problems such as factoring extremely large numbers or computing discrete logarithms of extremely large numbers.

3.5.1.2.1. Diffie-Hellman

Whitfield Diffie and Martin Hellman developed the Diffie-Hellman algorithm in 1976. Its security stems from the difficulty of calculating the discrete logarithms of huge numbers. The protocol allows two users to exchange a secret key over an insecure medium without any prior secrets.

The protocol has two system parameters, p and g. They are both public and may be used by everybody. Parameter p is a prime number, and parameter g (usually called a generator) is an integer that is smaller than p, but with the following property: For every number n between 1 and p 1 inclusive, there is a power k of g such that $n = g^k \bmod p$.

The following steps describe the Diffie-Hellman exchange:

Step 1. Alice and Bob agree on generator g and modulus p.

Step 2. Alice chooses a random number A and sends Bob its public value $A' = g^A \bmod p$.

Step 3. Bob chooses a random number B and sends Alice his public value $B' = g^B \bmod p$.

Step 4. Alice computes $k = (B')^A \bmod p$.

Step 5. Bob computes $k' = (A')^B \bmod p$.

Step 6. Both k and k' are equal to $g^{AB} \bmod p$.

Alice and Bob now have a shared secret (k = k'), and even if people have listened on the untrusted channel, there is no way they could compute the secret from the captured information (assuming that computing a discrete logarithm of A or B is practically unfeasible).

3.5.1.2.2. Rivest, Shamir, Adelman(RSA):

Rivest, Shamir, Adelman (RSA) was a patented public key algorithm invented by Ron Rivest, Adi Shamir, and Len Adelman in 1977. The patent expired in September 2000, and the algorithm is

now in the public domain. Compared to other algorithms, RSA is by far the easiest to understand and implement.

The RSA algorithm is very flexible and has a variable key length where, if necessary, speed can be traded for the level of security of the algorithm. The RSA keys are usually 512 to 2048 bits long. RSA has withstood years of extensive cryptanalysis. Although those years neither proved nor disproved RSA's security, they attest to a confidence level in the algorithm. RSA security is based on the difficulty of factoring very large numbers. If an easy method of factoring these large numbers were discovered, the effectiveness of RSA would be destroyed.

To generate an entity's RSA keys, you would follow these steps:

Step 1. Select two large prime numbers, p and q.

Step 2. Compute n using the following formula:
n = p x q

Step 3. Choose a huge prime e, with the constraint that e and (p 1)(q 1) are relatively prime. The public key is (e,n).

Step 4. Calculate the private key d:
e.d = mod (p 1)(q 1)
$d = e^{1}$ mod ((p 1)(q 1))

The numbers d and n are also relatively prime. The numbers e and n are the public key. The number d is the private key. The numbers p and q are no longer needed. They were used only to calculate the other values and can be discarded but never revealed.

3.5.2. Hashing Algorithms

Hashing is one of the mechanisms used for data integrity assurance. Hashing is based on a one-way mathematical function, which is relatively easy to compute but significantly harder to reverse. Breaking a glass is a good example of a one-way function. It is easy to smash a glass into thousands of pieces, but almost impossible to put all the tiny pieces back together to rebuild the original piece.

The hashing process shown in Figure 3.9 uses a hash function, which is a one-way function to input data to produce a fixed-length digest (fingerprint) of output data. The digest is cryptographically strong; that is, it is impossible to recover input data from its digest. If the input data changes just a little, the digest (fingerprint) changes substantially in what is called an avalanche effect.

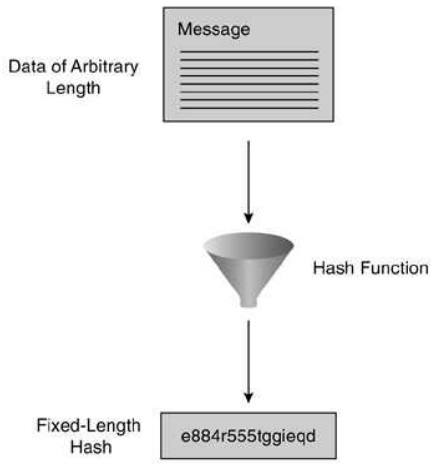


Figure 3.9: Hashing Function

The figure illustrates how hashing is performed. Data of arbitrary length is input to the hash function, and the result of the hash function is the fixed-length hash (for example, a digest or fingerprint).

Hashing only prevents the message from being changed accidentally (that is, by a communication error). There is nothing unique to the sender in the hashing procedure; therefore, anyone can compute a hash for any data, as long as she has the correct hash function.

Thus, hash functions are helpful to ensure that data was not changed accidentally, but they cannot ensure that data was not deliberately changed.

Some well-known hash functions are listed here and are discussed in the following section:

- Message Digest 5 (MD5) with 128-bit digest
- Secure Hash Algorithm 1 (SHA-1) with 160-bit digest

3.5.2.1. Message Digest 5

The Message Digest 5 (MD5) algorithm is a ubiquitous algorithm developed by Ron Rivest. It is used in a variety of Internet applications today.

As the name suggests, MD5 is a one-way function with which it is easy to compute the hash from the given input data, but it is unfeasible to compute input data given only a hash. MD5 is also collision resistant, which means that two messages with the same hash are very unlikely to occur.

MD5 is considered less secure than SHA-1 because MD5 has some weaknesses, the explanation of which is beyond the scope of this book. SHA-1 also uses a stronger, 160-bit digest, which makes MD5 the second choice as hash methods are concerned.

3.5.2.2.SHA-1

The NIST developed the Secure Hash Algorithm (SHA). SHA-1 is a revision to the SHA that was published in 1994. Its design is similar to MD5. The algorithm takes a message of less than 264 bits in length and produces a 160-bit message digest. This algorithm is slightly slower than MD5.

3.5.3. Secure Socket Layer and Transport Layer Security

Netscape originally developed Secure Socket Layer (SSL), but it is now accepted by the World Wide Web as the standard for authenticated and encrypted communication between clients and servers. The SSL protocol is application independent, allowing protocols such as HTTP, FTP, and Telnet to be layered on top of it transparently.

The SSL protocol is able to negotiate encryption keys and authenticate the server before data is exchanged by the higher-level application. The SSL protocol maintains the security and integrity of the transmission channel by using encryption, authentication, and message authentication codes.

The SSL Handshake Protocol consists of two phases: server authentication and optional client authentication. In the first phase, the server, in response to a client's request, sends its certificate and its cipher preferences. The client then generates a master key, which it encrypts with the server's public key, and transmits the encrypted master key to the server. The server recovers the master key and authenticates itself to the client by returning a message authenticated with the master key. Subsequent data is encrypted and authenticated with keys derived from this master key. In the optional second phase, the server sends a challenge to the client. On the challenge, the client authenticates itself to the server by returning the client's digital signature and its public-key certificate.

The Transport Layer Security (TLS) is based on SSL. It is an improved version of SSL, but the industry has not made the shift to this new standard yet. SSL is still the method supported by all web servers and web browsers.

3.5.4. Digital Certificates

Key management is often considered the most difficult task in designing and implementing cryptographic systems. Businesses can simplify some of the deployment and management issues that

are encountered with secured data communications by employing a Public Key Infrastructure (PKI). Because corporations often move security-sensitive communications across the Internet, an effective mechanism must be implemented to protect sensitive information from the threats presented on the Internet.

The three primary security vulnerabilities associated with communicating over a publicly accessible network are as follows:

- Identity theft Intruder gains illegitimate access by posing as an individual who actually can access secured resources.
- Eavesdropping Intruder "sniffs" the data transmission between two parties during communications over a public medium.
- Man-in-the-middle Intruder interrupts a dialogue and modifies the data between the two parties. In an extreme case, the intruder takes over the entire session.

3.5.4.1. Characteristics of Digital Certificates

PKI provides a hierarchical framework for managing the digital security attributes. Each PKI participant holds a digital certificate that has been issued by a CA. The certificate contains a number of attributes that are used when parties negotiate a secure connection. These attributes must include the certificate validity period, end-host identity information, encryption keys that will be used for secure communications, and the signature of the issuing CA. Optional attributes may be included, depending on the requirements and capability of the PKI.

A CA can be a trusted third party, such as VeriSign or Entrust, or a private (in-house) CA that you establish within your organization.

Digital signatures, enabled by public key cryptography, provide a means to digitally authenticate devices and individual users. In public key cryptography, such as the RSA encryption system, each user has a key-pair containing both a public key and a private key. The keys act as complements, and anything encrypted with one of the keys can be decrypted with the other. In simple terms, a signature is formed when data is encrypted with a user's private key. The receiver verifies the signature by decrypting the message with the sender's public key.

The fact that the message could be decrypted using the sender's public key means that the holder of the private key created the message. This process relies on the receiver having a copy of the sender's public key and knowing with a high degree of certainty that it really does belong to the sender and not to someone pretending to be the sender.

To validate the CA's signature, the receiver must know the CA's public key. Normally, this is handled out-of-band or through an operation performed during installation of the certificate. For instance, most web browsers are configured with the root certificates of several CAs by default.

3.5.4.2. Enrolling in a CA

The enrollment process of obtaining a certificate is shown in Figure 4-5. Enrollment is enacted between the end host desiring the certificate and the authority in the PKI that is responsible for providing certificates. The hosts that participate in a PKI must obtain a certificate, which they present to the parties with whom they communicate when they need a secured communications channel.

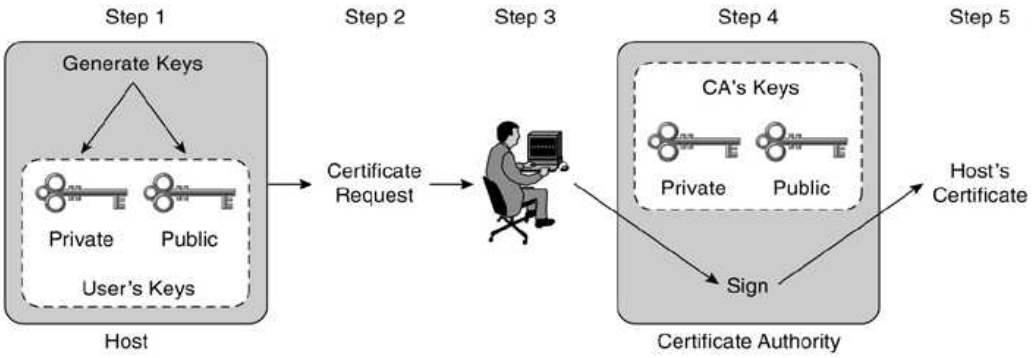


Figure 3.10: Enrollment Process in CA

The enrollment process is illustrated in Figure3.10 and described in the following list:

1. The end host generates a private-public key pair.
2. The end host generates a certificate request, which it forwards to the CA.
3. Manual human intervention is required to approve the enrollment request, which is received by the CA.
4. After the CA operator approves the request, the CA signs the certificate request with its private key and returns the completed certificate to the end host.
5. The end host writes the certificate into a nonvolatile storage area (PC hard disk or NVRAM on Cisco routers).

Chapter 04
Security Policies

4.1. Introduction:

If a company wants to adequately protect its network, it must implement a security policy. It is important to establish a good balance between the level of security and the ability of users to get to the information they need. The most secure PC is the one that is not connected to a network, but the problem with this approach is that nobody can access the data. This chapter provides guidelines for developing a security policy how to define it, develop it, adopt it, and enforce it with users. Cisco has developed a security wheel illustrate the process that a company has to undertake to have a proper security policy. With a security policy alone, you are nowhere. That policy needs to be implemented, monitored, tested, and improved all the time.

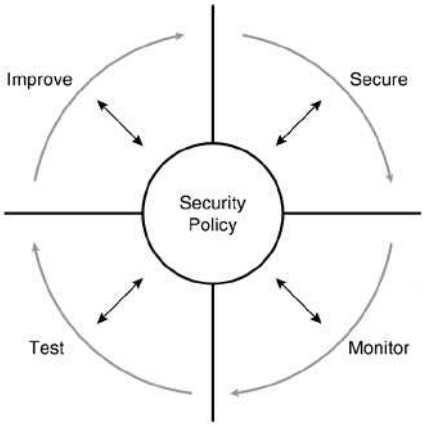


Figure 4.1: Security Wheel

Over the past years, Internet-enabled business has changed drastically. E-business applications such as e-commerce and remote access enable companies to streamline processes, lower operating costs, and increase customer satisfaction. Applications for e-commerce require mission-critical networks that accommodate voice, video, and data traffic. These networks must be scalable to support an increasing number of users as well as increases to capacity and performance. However, as networks grow to accommodate the applications that are available to increasing numbers of users, they become even more vulnerable to a wider range of security threats. To combat these threats, security technology must play a major role in today's networks.

The closed network shown in Figure 4.2 typically consists of a network designed and implemented in a corporate environment and provides connectivity only to known parties and sites without connection to public networks. Networks were designed that way in the past and were reasonably secure because of no outside connectivity.

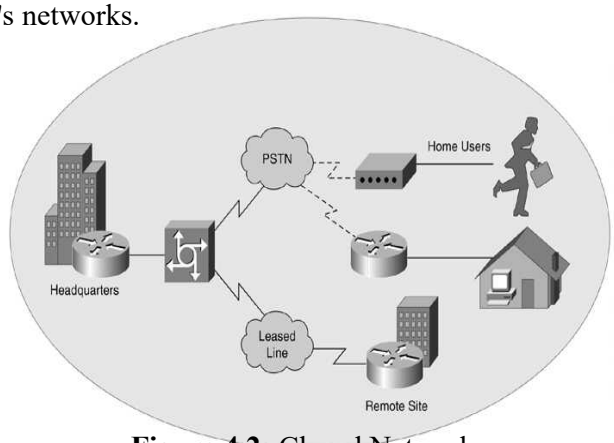


Figure 4.2: Closed Network

As shown in Figure 4.3, today's networks are designed with availability to the Internet and public networks. Most of today's networks have several access points to other networks both public and private; therefore, securing these networks has become fundamentally important. With the

development of large, open networks over the past 20 years, there has been a huge increase in security threats. Security threats have increased not only because hackers have discovered more vulnerabilities, but also because hacking tools have become easier to use and the technical knowledge simpler to learn.

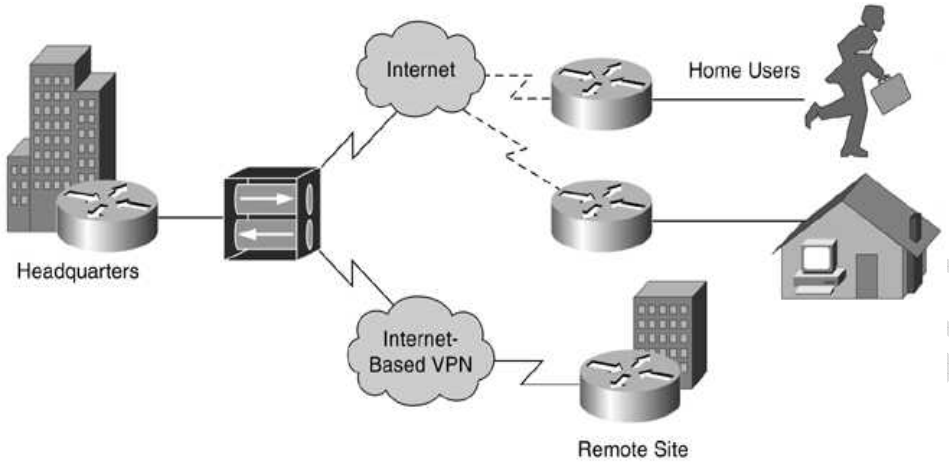


Figure 4.3: Modern Networks

Security has moved to the forefront of network implementation and management. Allowing open access to network resources and ensuring that the data and resources are as secure as possible is necessary for the survival of many businesses. The need for security is becoming more important because of the following:

- It is required for e-business. The importance of e-business and the need for private data to traverse public networks has increased the need for network security.
- It is required for communicating and doing business safely in potentially unsafe environments.

Networks require development and implementation of a corporate-wide security policy. Establishing a security policy should be the first step in migrating a network to a secure infrastructure.

4.2. Defining a Security Policy?

A security policy can be as simple as an acceptable use policy for the network resources, or it can be several hundred pages in length and detail every element of connectivity and associated policies. According to the Site Security Handbook (RFC 2196), "A security policy is a formal statement of rules by which people who are given access to an organization's technology and information assets must abide." It further states, "A security policy is essentially a document summarizing how the corporation will use and protect its computer and network resources." A security policy is actually the center of the security wheel that is explained in more detail later in this chapter.

4.3. Importance of a Security Policy

Security policies provide many benefits and are worth the time and effort needed to develop them. Security policies are important to organizations for a number of reasons, including the following:

- Create a baseline of your current security posture
- Set the framework for security implementation
- Define allowed and disallowed behavior

- Help determine necessary tools and procedures
- Communicate consensus and define roles
- Define how to handle security incidents

This leads directly to the next question: What should a good security policy contain? The following list is an overview of the key components or sections for a security policy:

- Statement of authority and scope identifies the sponsors of the security policy and the topics to be covered.
- Acceptable use policy Spells out what the company allows and does not allow regarding its information infrastructure.
- Identification and authentication policy Specifies what technologies and equipments are used to ensure that only authorized individuals have access to the organization's data.
- Internet access policy Defines the ethical and proper use of the organization's Internet access capabilities.
- Campus access policy Defines how on-campus users should use the data infrastructure.
- Remote access policy Describes how remote users should access the company's data infrastructure.
- Incident handling procedure Specifies how the organization creates an incident response team and the procedures the team uses during and after an accident occurs. A security policy has no use if no appropriate actions take place after an incident has happened.

Each company's security policy is unique and must meet the objectives of the company. Also note that the previous list is not definitive. The main purpose of a security policy is to inform users, staff, and management of their obligation to protect the organization's technology and information assets. The policy should state the mechanisms through which these requirements can be met. An acceptable use policy (AUP) can also be part of a security policy. It can tell the users what they can and cannot do on the network. A security policy should be as explicit as possible to avoid ambiguity or misunderstanding.

4.4. Development Process

All sites should have a comprehensive security plan. This plan should be at a higher level than more specific policies such as the one discussed in the example at the end of this chapter. The security plan should be crafted as a framework of broad guidelines into which specific policies fit. It is important to have this framework in place so that individual policies are consistent with the overall site security architecture. Having a strong policy on corporate access from home but weak restrictions on who is entering the building and using the PC in the lobby is inconsistent with the overall philosophy of strong security restrictions on data access.

Two diametrically opposed underlying philosophies can be adopted when defining a security plan: deny all and allow all. Both alternatives have strong and weak points, and the choice between them depends on the need of security for a particular site. The first option is to deny everything and then selectively enable services on a case-by-case basis. This model, which is called the deny all model, is generally more secure than the allow all model. Successfully implementing the deny all model is, however, more work intensive.

The other model, which is referred to as allow all, is much easier to implement, but it is generally less secure than the deny all model. To implement it, simply turn on all services (this is usually the default on a host system) and allow all protocols to travel across network boundaries (this is usually the default at the router level) on a host system. As security holes become apparent, they are restricted or patched at either the host or the network level. Both models can be used at the same

time. For example, the policy may be to use the allow all model when setting up workstations for general use but to use the deny all model when setting up information servers.

To craft an effective security policy, it is important to appoint a development team. For a security policy to be appropriate and effective, it needs to have the acceptance and support of all levels of employees within the organization. It is important that corporate management fully supports the security policy process; otherwise, there is little chance that the process will have the intended impact. When creating and reviewing a security policy, the following individuals and groups should be involved:

- Site security administrator
- Information technology technical staff
- Administrators of large user groups
- Security incident response team
- Representatives of the user groups affected by the policy
- Responsible management
- Human resources (HR)

4.5. Incident Handling Process

In the past when developing a security policy, incident handling was often overlooked. The result of that approach was that when an attack was in progress, many decisions were made in haste. Hastily made decisions actually made it more difficult to track down the source of the incident, collect evidence to be used in prosecutions, prepare for the recovery of the system, and protect the valuable data contained on those systems.

One of the most important, but often overlooked, benefits for efficient incident handling is economic. Having both technical and managerial personnel respond to an incident requires considerable resources. If employees are trained to handle incidents efficiently, less staff time is required when an incident occurs. Another benefit is related to public relations. If news comes out about security incidents, an organization's stature among current and potential clients can be damaged. Efficient incident handling minimizes the potential for negative exposure.

As in any set of preplanned procedures, attention must be paid to a set of goals for handling an incident. These goals are prioritized differently depending on the organization. The following list identifies objectives for dealing with incidents:

- Determine what happened
- Plan how to avoid a repeat attack
- Avoid escalation and further incidents
- Assess the impact and damage of the incident
- Recover from the incident
- Update policies and procedures as needed
- Identify the perpetrators

Depending on the nature of the incident, there might be a conflict of priorities between analyzing the original source of the problem and restoring systems and services. Major goals such as assuring the integrity of critical systems may be the reason for not analyzing an incident. This is an important management decision, but everyone involved must be aware that without analysis, the same incident can happen again.

4.6. Security Wheel

Cisco understands the importance of network security and its implications for the critical infrastructures on which developed nations depend. After setting appropriate policies, an organization must methodically consider security as part of normal network operations. This could be as simple as configuring routers not to accept unauthorized addresses or services, or as complex as installing firewalls, intrusion detection systems (IDSs), centralized authentication servers, and

encrypted virtual private networks (VPNs). After developing a security policy, you can secure your network using a variety of products. Before you can secure your network, however, you need to combine your understanding of users, the assets needing protection, and the network topology. The process of developing and securing your network can be illustrated in a diagram like Figure 4.1, called a security wheel.

Figure 4.1 shows that network security is a continuous process built around a security policy. Securing your network is like a never-ending story. Security improvements are always necessary. Hackers continually find new ways to attack your network. In the Secure phase shown in the figure, the person or department responsible for an organization's security implements security solutions to stop or prevent unauthorized access and to protect information by using the following methods:

- Authentication This method is the recognition and the mapping to the policy of each individual user's identity, location, and the exact time logged on to the system. Authentication also encompasses the authorization of network services granted to users and what functions they are authorized to perform on the network.
- Encryption Encryption is a method for ensuring the confidentiality, integrity, and authenticity of data communications across a network. There are several encryption methods available, and some of them, such as DES, 3DES, and AES.
- Firewalls A firewall is a set of related services, located at a network gateway, that protects the resources of a private network from users from other networks. Firewalls can also be standalone devices or can be configured on most routers.
- Vulnerability patching This method entails the identification and patching of possible security holes that could compromise a network and the information available on that network.

After a network is secure, it has to be monitored to ensure that it stays secure. Network vulnerability scanners can proactively identify areas of weakness, and IDSs can monitor and respond to security events as they occur. Using these security monitoring solutions, organizations can obtain unprecedented visibility into the network data stream and the security posture of the network.

After the monitoring phase comes the testing phase. Testing security is as important as monitoring it. Without testing the security solutions in place, it is impossible to know about existing or new attacks. The hacker community is an ever-changing environment. An organization can perform the testing itself, or it can be outsourced to a third party such as the Cisco Advanced Services for Network Security (ASNS) group. Monitoring and testing provides the data necessary to improve network security. Administrators and engineers should use the information from the monitoring and testing phase to make improvements to the security implementation. They should also adjust the security policy as vulnerabilities and risks are identified.

.

4.7. Sample Security Policy

This is a portion of a sample security policy for a VPN. It includes all the points that a good security policy must contain.

Purpose

The purpose of this policy is to provide guidelines for remote access IPSec connections to the XYZ corporate network.

Scope

The policy applies to all XYZ employees, contractors, consultants, temporaries, and other workers, including all personnel affiliated with third parties who are using VPNs to access the XYZ corporate network. The policy applies to implementations of VPN that are established though a VPN concentrator.

Policy

Employees and authorized third parties (customers, vendors, and so on) who are approved by XYZ may use the benefits of VPNs, which constitute a company-managed service. This means that the user is not responsible for selecting an Internet service provider (ISP). XYZ will coordinate the installation and will pay associated fees. No equipment other than that ordered by XYZ can be used for this purpose. Further details can be found in the Remote Access Policy.

The following list identifies some additional guidelines:

- It is the responsibility of employees with VPN privileges to ensure that unauthorized users are not allowed access to XYZ's internal networks.
- VPN access is controlled by using a one-time password authentication with a token device. While connected to the corporate network, no other connections can be established.
- When actively connected to the corporate network, VPNs force all traffic to and from the PC over the VPN tunnel. All other traffic is dropped.
- Split tunneling is not permitted. Only one network connection is allowed.
- VPN gateways are set up and managed by XYZ network operational groups.
- All computers connected to XYZ internal networks via VPN or any other technology must use the most up-to-date antivirus software that is the corporate standard.
- VPN users are automatically disconnected from XYZ's network after 15 minutes of inactivity. The user has to log on again to reconnect to the network.
- The VPN concentrator is limited to an absolute connection time of 12 hours.

Enforcement

Any employee found to have violated this policy may be subject to disciplinary action, up to and including termination of employment.

Definitions

In this security policy, the following definitions apply:

- **VPN concentrator** A device in which VPN connections are terminated. This device is sometimes also called the IPSec concentrator.
- **InfoSec** A term used to refer to the team of people responsible for network and information security.
- **Split tunneling** The term used to describe a multiple-branch networking path. A tunnel is split. When some network traffic is sent to the VPN concentrator and other traffic is sent directly to the remote location without passing through the VPN concentrator.

4.8. Conclusion

Now that you know how to write a security policy, you can start implementing it in your own world. As discussed in this chapter, you should keep in mind that a security policy is not a fixed document. It needs to be updated on a regular basis to meet all the new requirements.

4.9. Important Questions:

1. What is the difference between a closed network and an open network?
2. Define a security policy.
3. Name three reasons why a company should have a security policy.
4. Name at least four key components that a good security policy should contain.
5. Name the two philosophies that can be adopted when defining a security plan.
6. Which individuals should be involved when creating a security policy?
7. Give the four stages of the security wheel.
8. Which security solutions can be implemented to stop or prevent unauthorized access and to protect information?
9. Explain the monitoring phase of the security wheel.
10. Write a security policy (similar to the VPN policy) for password protection.

Chapter 05
Secure Design

5.1. Introduction:

The goal of network security is to protect networks (including equipment, servers, content, and applications) against attacks, with the intent of ensuring data and system availability, confidentiality, and integrity. This chapter briefly covers the basics of a secure network design, taking that goal into consideration.

During the initial design phase of a network, the network architects identify the risk of attacks as well as the costs of repairing damage from attacks for all the network equipment, applications, and services. Cost-benefit analysis, Return on Investment, and Total Cost of Ownership are some of the techniques at hand for making these decisions.

As discussed in Chapter 5, "Security Policies," the roadmap for the implementation of network security and the driver behind the network security design process is the security policy. The security policy, which ideally is designed by both the network design and IT security teams, addresses security requirements and implementation guidelines. The security requirements for each process and service need to be defined before the network is divided into modules. Each module can then be treated separately and assigned a different security role.

Cisco has developed a comprehensive blueprint using this modular approach called Security Architecture for Enterprises (SAFE). The objective of SAFE is to have multiple layers of security so that intruders have limited access to certain parts of the network. This blueprint serves as a guide to network designers who are considering the security requirements of their network.

SAFE takes a defense-in-depth approach to network security design. This methodology focuses on expected threats and methods to mitigate them, resulting in a layered approach to security. With a layered approach, the failure of one security system is not likely to lead to the compromise of the network resources. More information on SAFE can be found in Appendix A, "SAFE Blueprint."

This chapter starts by delving into network design principles and methodologies so that you can gain a basic understanding of these network design concepts.

5.2. Network Design Principles

The fundamental principles of network design call for dividing the network into manageable blocks. This division ensures that the network can function within the specifications, performance, and scale limits of the required applications, protocols, and network services.

The network infrastructure itself is an important component in the design process because it transports the application and network-management traffic. The designed network infrastructure must meet at least three high-level goals:

- It should provide timely, reliable, and secure data transport.
- It should be adaptable to satisfy ever-changing application demands.
- The cost of future growth needed for business or information expansion should be appropriate to the extent of the required changes.

Building a network infrastructure requires considerable planning, designing, modeling, and, most important, information gathering. Network designers have many technologies to consider. The functionality of the selected technology and networking equipment is important because the equipment must conform to standards to provide interoperability and must be able to perform the tasks required by the network architecture.

The network architecture, an intermediate network design, provides a blueprint for the detailed design activities required to realize a functioning network infrastructure. When designing networks, it is important to look at the resources you have available to implement the new network

architecture. You must also be sensitive to the quantity and quality of the resources available to operate and manage the network.

5.2.1. Top-Down Design Practices

One of the basic requirements for a successful implementation and strategic use of a computer network and the Internet is the engagement of the top executives, particularly an organization's CEO, during the design phase. Strategic and secure use of the Internet to extend the organization's reach outward to customers, clients, vendors, and partners cannot really become a core part of an organization's business philosophy until all the top executives assume an active leadership role in the process. Top executive support speeds the development of an organization's Internet capabilities; when the company's CEO, CIO, or CTO recognizes that the efficiencies enabled by the Internet are the key to future growth and survival, cultural transitions and adoption rates are bound to happen faster.

It is good practice to perform a periodic executive review and to restate or revise an organization's goals. Given the effort required to gather input from the various constituencies and the value of executive time, many organizations undertake executive review annually. For instance, the leadership team of Cisco selected "Leadership in Internet capabilities in all functions" as one of its top three goals. Every group throughout the company identified areas in which the Internet could impact its business sector, defined how it could become one of the best in those sectors, and regularly reported progress on those plans. In other words, the Cisco Internet strategy was integrated with each group's business strategies, and each group was required to develop measurable and reportable results. Getting executive support not only aids in the allocation of necessary resources, but also sends the right message throughout the company. At the end of the day, the entire company needs to be involved in promoting secured network-enabled business initiatives as part of the overall business strategy.

5.2.2. Requirements and Constraints

A secure network design is an exercise in meeting new and old requirements while also working with certain constraints. These constraints can be technological, social, political, or economic.

5.2.2.1.Technological Constraints

The impact of technological developments is obvious. Technological developments are used to implement the latest global network business models and network virtual organizations. In conjunction, they are responsible for supporting the changing needs of consumers and society in general.

Recent technological developments are the reason that Internet traffic keeps increasing at a rapid pace. CPU processing speed takes approximately 18 months to double. The increase in Internet traffic and the inability of most organizations to augment capital equipment budgets to support these growth rates mean that CPU resources are a design constraint that you must address through network design and device configuration. Typically, the computation (processing) limitations that apply to network design are associated with processing routing-table calculations, encrypting and decrypting secured packets, accounting, enforcing access lists, or just forwarding packets.

As with device processing limitations, device memory size plays a significant role during the design phase, more or less for the same reasons.

Other resource considerations that can affect network design include configurable buffer capacity, device port density, interface bandwidth, and backplane capacity constraints. In general, greater capacity increases the cost of the implementation. (Greater capacity might increase the cost of a single device but lower the cost of the entire implementation because fewer devices are needed.)

Another technological constraint involves ensuring that appropriate ventilation, air-conditioning, and other environmental requirements are met in the operations and laboratory facilities used to house the equipment.

5.2.2.2.Social Constraints

Manpower, or labor in general, is a clear concern in any network design. The more often a task must be executed (for instance, the amount of effort and skill required to connect a new user to the network or to expand the capacity of the network infrastructure), the more the design should focus on making that particular task simple and efficient to manage. Including network-management services in the design can mitigate some of the labor concerns through the automation of monitoring and reporting functions. This automation should reduce the quantity of highly skilled employees required for the ongoing operation of the network.

5.2.2.3.Political Constraints

Political concerns include the compulsory use of standards and installed applications that are difficult to understand, implement, and use. These political concerns are internal company politics and not necessarily driven by governmental policy.

Some organizations might have a prearranged single-vendor partnership agreement, whereas other team members desire a multivendor type of environment. These partnership arrangements are often necessary to meet the business requirements of the company. By selecting a single partner, an organization can meet the business challenge of building a network with an integrated, intelligent design that accommodates business growth. The design should make it easy and cost effective to add new features, and technology can maximize the total value of network ownership.

5.2.2.4.Economic Constraints

Economic constraints play a major role for all network designers. Doing more with less is a common requirement, partially enabled by advances in semiconductor technology. Even when there is a mandate to "achieve the best possible service at the lowest possible cost," there are design consequences. Common areas of design compromise for minimizing network acquisition and operations costs include wide area network (WAN) bandwidth, quality-of-service (QoS) guarantees, availability, security, and manageability. Other requirements with a lower priority or less visibility are deferred to later implementation phases or cancelled.

5.2.3. Design Activities, Tools, and Techniques

During the network-design process, tools are available to facilitate some of the activities. Some of the activities supported by tools include network auditing, traffic analysis, and network simulation. The choice of tools is determined by the value of the network investment and the consequences of network failure.

Having tools available to support every stage of the design process helps to:

- **Reduce risk** The risk of adding new equipment in the network
- **Increase understanding** How certain components work in your environment
- **Improve responsiveness to design opportunities** Quickly obtain technical analysis and business cases

5.2.3.1.Auditing and Analyzing an Existing Network

Network audit tools help you to generate specific reports on certain parts of your network and to analyze how these segments of the network are performing. The network audit process should provide detailed recommendations to address the challenges, opportunities, and problems identified in the audit. The audit also help the network-engineering team proactively identify and resolve potential network troubles before major problems are encountered.

Following is a list of reports that are often generated as part of a network audit:

1. Performance
2. Configuration
3. Software
4. Hardware

In general, a network audit identifies specific opportunities to improve network utilization, availability, and stability, resulting in a reduced operation cost and a maximum return on the investment in the network infrastructure.

Network traffic analysis collects and analyzes data, which allows the network designer to balance the network load, troubleshoot and resolve network problems, optimize network performance, and plan future network growth. Traffic analysis is often performed as part of a network audit to generate performance reports.

The analysis tools help engineers and network designers better understand traffic patterns in the network. Many analysis-tool suites are on the market. Some provide only basic calculations. Others give extensive detail, including a complex analysis of traffic patterns, capacity availability, delay, and operational stability. Some tools allow the designer to rerun the analysis as the design is developed. Traffic analysis conducted during deployment allows timely adjustment of the design based on issues encountered at various locations or times.

5.2.3.2.Simulating Network Traffic

Network simulation has at least two distinct realizations. The first models the network using software to emulate the traffic sources and sinks (drop offs), network devices, and the links that connect them. By varying model parameters, the designer can approximate the impact of more or less traffic demand or network resources. Although simulation software is expensive, for a large network it is far less expensive than building a flawed design. The second kind of simulation uses special hardware and software to generate traffic for injection into a live network for subsequent traffic analysis.

This testing activity is useful for

- Validating and adequately testing QoS
- Testing latency
- Checking adaptive protocols
- Testing multicasting

Traffic generation is also appropriate for estimating how the existing network responds as you add new applications and services. Dynamic bandwidth utilization and latency are relatively difficult to estimate compared to simple traffic delivery. Loss is relatively obvious. You can use adaptive protocols and applications with traffic generators to validate the expected behaviors.

5.2.3.3.Defense in Depth

As the risks and challenges related to network security grow, organizations should take a systematic and multitiered approach to planning and deploying secure network infrastructures. Defense in depth is a practical strategy for achieving efficient security solutions by establishing multiple overlapping layers and countermeasures. This strategy ensures that even when an intruder or attacker is able to penetrate a company's network, other security systems (the second line of defense) detect and prevent the attack before unauthorized access takes place.

The defense in depth strategy rests on several principles:

- Layered defenses First, second, and so on lines of defense
- Defenses residing in multiple locations At network boundaries, in different security zones, on servers, in applications, and so on
- Robust defenses Balance between protection capabilities and cost, a stronger defense at network boundaries than on servers, and so on

The SAFE Blueprint for network security from Cisco offers a defense-in-depth, modular approach to security that can evolve and change to meet the needs of different organizations.

5.3. Network Design Methodology

As network expectations have changed, so have design principles. Enterprises no longer rely on a single vendor, technology, or protocol. The design strategy has changed dramatically to include security and scalability as primary criteria. Security has a large impact on network design.

There is greater redundancy in network designs. Since the events of September 11, 2001, business continuity has become a priority. Organizations are focusing on increased levels of redundancy, and disaster-recovery planning is becoming a necessity. Redundancy takes many forms, including separate power sources, multiple WAN carriers, alternate cable routes, and redundant hardware. Network connectivity and services are critical components of enterprise operations. The cost of downtime is increasing at a phenomenal rate.

Enterprises are no longer locked into using a single vendor, technology, or protocol; many technologies have standardized. But designing a network is still not a trivial factor. Assessing the design criteria enables you to understand the network and what it was meant to do. Network designs must easily adapt to implement the next generation of technology. Many network designers are planning for IP telephony; these network design plans are not just for new networks but are improvements on existing ones. Properly planning networks based on sound architecture makes necessary network redesigns easier at a later stage.

5.3.1. Stages of the Network

Design is just one component of a network life cycle. Planning, design, implementation, operation, and optimization (PDIOO) are the stages of the network life cycle. Each stage builds on its predecessor to create a sound network that maintains its effectiveness despite changing business needs. You can apply the PDIOO methodology to all technologies. During the PDIOO process, you define key deliverables and associated actions with a direct correlation to the added value and benefit for the client's network. For example, understanding business goals, usage characteristics, and network requirements helps you avoid unnecessary upgrades and network redesigns, thereby reducing the time it takes to introduce new services in the network.

5.3.1.1.Planning Phase

During the planning stage, you can test the logic of your future design for flaws. Planning helps you avoid replicating a logical mistake in a network design that you might use as a template across a number of locations. The planning stage focuses on technical as well as financial criteria and takes into account all the requirements and constraints that were discussed in the previous section. During this phase, it is important to identify all the stakeholders in order to make this process a success. The stakeholders are people or organizations who have a vested interest in the environment, performance, and outcome of the project.

5.3.1.2.Design Phase

After completing the planning stage, you have enough information to develop a network design. If a network is already in place, use this phase to review and validate the network design as it is currently implemented. At this stage, you choose products, protocols, and features based on criteria defined in the planning stage. You develop network diagrams to illustrate what changes will occur in the network to achieve the desired results. The more detailed the network diagram and plan, the better you can anticipate the challenges during implementation.

5.3.1.3.Implementation Phase

The implementation stage provides detailed, customized deliverables to help avoid risks and meet expectations. A sound implementation plan ensures smooth deployment even when issues arise. Communicating the implementation plan to all stakeholders provides you with an opportunity to assess the viability of the plan. It is better to find mistakes on the drawing board than during implementation.

Good processes, such as change control, can effectively handle issues that occur during deployment. Change control provides flexibility because it is impossible to plan for every contingency, especially if the implementation has a long duration.

5.3.1.4.Operation Phase

The operation phase, also known as the operational-support phase, is designed to protect your network investment and help your staff prevent problems, maximize system utility, and accelerate problem resolution.

5.3.1.5.Optimization Phase

The last step in the PDIOO process is the optimization of the network. A sound design still requires optimization and tweaking to reach its full potential. The optimization of the network can be as simple as hardening servers against security threats or adding QoS to the network for latency-sensitive traffic.

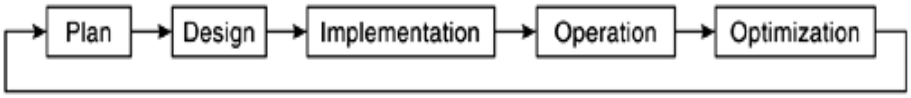


Figure 5.1: Stages of the PDIOO Process

Optimization can even lead to a redesign of the network, so the cycle would begin again.

5.4. Return on Investment

A strategic part of the network design process is a tracking mechanism to measure the profit for a specific investment. A company's management team uses this simple tool as a financial metric to make business investment decisions and to measure a company's performance over time.

Return on Investment (ROI) is often calculated and defined in percentage terms. It results in the return a customer can expect from the investment made. The ROI is calculated by dividing the profit (return) by the total investment cost. Sometimes the ROI is also specified as a ratio or break-even number. The latter has a time ratio in the calculation and results in the exact timeframe until the investment is returned. Most customers in today's business environment try to understand or require a value justification, which is where the ROI calculation plays a significant role. In a value justification, the network designer is requested to prove the value of the proposal.

5.5. Physical Security Issues

It is relatively easy to implement and maintain a tight security policy for your network security. Physical security, on the other hand which can also be defined using a blueprint, standards, or even models is much more difficult to implement in the real world. The implementation can fall short for various reasons, most important being budget constraints. A slight shift in focus is taking place because of the recent effects and threats of global terrorism. This shift might trigger increased attention to the physical security that is necessary for the implementation of comprehensive physical security measures. Such implementations will become as common as encryption, firewalls, VPNs, and others.

Physical security is defined as the process of identifying and describing all the measures necessary to protect your facility. This process includes internal and external security measures, disaster-recovery plans, and personnel training.

5.5.1. Securing the Perimeter

When implementing physical security at a company level, the first consideration is the location of your site. In reality, this step might not be an option because a limited budget can force you to use an existing building. A site must meet a minimum set of requirements, which are defined by physical security blueprints or models.

Once the facility is built, multiple layers of security are required. The following list is an overview of available layers and options for external physical security:

- Electronic fence
- Electromagnetic intrusion detection system
- Camera systems
- Entrance security (smart cards, PIN codes)
- Permanent guards

Achieving maximum external physical security according to these specifications is compromised in many situations because not all layers can be easily implemented.

5.5.2. Internal Security

Internal physical security techniques can be defined by following a layered model approach. Some areas protected by both the external and internal measures overlap. For instance, camera systems can be installed all over the campus and as entrance security for mission-critical areas such as lab space, communication rooms, and server rooms. Just as with external security, internal security is layered. Entrance to low-security areas requires only a PIN code or card reader, and entrance to high-security areas requires card readers in combination with biometrics. High-level security areas can also be equipped with smoke, temperature, and humidity sensors.

5.5.3. Personnel Training

Developing a strong security policy helps to protect your resources only if all staff members are properly instructed on all facets and processes of the policy. Most companies have a system in place whereby all employees need to sign a statement confirming that they have read and understood the security policy. The policy should cover all issues the employees encounter in their day-to-day work, such as laptop security, password policy, handling of sensitive information, access levels, tailgating, countermeasures, photo IDs, PIN codes, and security information delivered via newsletters and posters. A top-down approach is required if the policy is to be taken seriously. This means that the security policy should be issued and supported from an executive level downward.

As far as physical security goes, many standards and blueprints exist, but implementation costs require compromises. Only serious attacks, intrusions, losses, or the latest threats of global terrorism can change the mindset that allows unreasonable compromises to physical security standards and the complete implementation of the physical security policy measures.

5.5.4. Survivability and Recovery

Even for the most protected and secure areas, a strong disaster-recovery plan needs to be defined. The possibility of things going wrong should be addressed upfront. For instance, uninterruptible power supplies (UPSs) are the de facto standard for countering power blackouts. When connecting your site to a service provider's network, only one connection creates a single point of failure. A central backup system is a mandatory service for all servers in the network.

Another disaster-recovery service is the implementation of a complete fail-over site. This is a drastic approach, but companies need to consider the loss of not just data but of their complete workplace when defining disaster-recovery plans. The cost of losing your complete workplace, data included, is nothing compared to the cost of installing a fail-over site.

5.6. Switches and Hubs

This section concentrates on switches and hubs. Many other networking devices are available, but switches and hubs are used here as an example of the network security design process. (Other devices are covered in other chapters.)

Bridged networks, with thousands of users connected, used to be large and flat (having no hierarchy), but that kind of network has almost disappeared. With the introduction of routers and switches, networks are subnetted (divided into subnets) into manageable sizes to limit broadcast

domains and to manage functional workgroups. Newer, multilayer switches perform routing and other high-level security network functions at speeds formerly attainable only with large switched networks. Before delving into some of the available security features on switches that need to be considered when designing a network, you should understand the basics of hubs and switches.

Both hubs and switches are networking devices used to interconnect workstations and servers. Externally they look similar, although from an operational standpoint some remarkable differences do exist.

Hubs share all available bandwidth among all connected devices, meaning that they distribute all the data received on one port to all the network devices they are connected to on the other ports. This is a highly inefficient use of network bandwidth. However, minimum processing delay is an advantage.

Switches, on the other hand, are smarter devices. Traffic-flow decisions are made based on tables. Traffic is analyzed and forwarding decisions are made using destination addresses. Only one port receives the traffic. The tables (containing MAC addresses) are populated by the switch, which knows each host and which port it resides on, with the exception of broadcasts.

Because of the simplicity of hubs and their limited feature set, they don't need to be discussed in depth here. This section concentrates only on switches and covers some of the added security features in these devices that can counter most attacks. Table 6-1 lists some of the features and mitigation techniques.

Table 5.1: Sample Switch Security Features

Feature	Mitigation Technique
Port security	Prevents MAC flooding attacks
Dynamic Host Configuration Protocol (DHCP) Option 82 and DHCP snooping	Secures DHCP transactions
Dynamic Address Resolution Protocol (ARP) inspection (DAI)	Prevents man-in-the-middle attacks
IP Source Guard	Prevents IP spoofing
802.1x enhancements	Implements authentication and guest virtual local-area network (VLAN) concept
Layer 2-4 access control lists (ACLs) including port-based access control list (PACL)	In isolated networks, limits IP addresses per customers on a port

Table 5.1 is just an example of some of the security features available on the current switches. Network security engineers can configure a rich set of switching security features to control security threats from their inception, wherever they occur in the network.

5.7. Conclusion

When designing a secure network, some goals need to be taken into consideration. The goal of network security is to protect networks against attacks, with the intent of ensuring data and system availability, confidentiality, and integrity. A good network design meets all these requirements. This chapter covered the basics of network design, network design principles, network design methodology, PDIOO, and physical security issues.

Chapter 06
Web Security

6.1. Introduction:

Is web security a worrisome topic? You bet it is. The many things to worry about include security risks to the operating systems, risks to the web servers, and even blunders by innocent users of web browsers. There are also access problems: who is authorized to access what, when can resources be accessed, and what should access privileges include. Webmasters can restrict access by using certificates, addresses, and credentials or by using a mechanism called Discretionary Access Control (DAC).

6.2. Hardening

When you install a new operating system, your security settings are all set to their default values. The same goes for installing a new web server or a browser. These settings need to be changed to harden the system against attacks or unauthorized access.

6.2.1. File Systems

When you install Windows, all versions have one thing in common: weak security. The obvious example is that after logging in, all users have full control (all permissions) at the root of every drive and at most of the drives' subdirectories and files. NT4 was the first Windows operating system to introduce a distinction between rights and permissions. A right allows the user to access the resources of the operating system itself, such as shutting down the system. A permission allows the user to access the file system's resources, such as reading and writing files. NT4 was also the first Windows product with DAC, which is discussed in more detail later in this chapter.

The Windows default for permissions is for the Everyone group to have full control from the root of each drive down. For a single user station, this is okay, but for a web server or file server, this is not acceptable. If you do not change the permissions, any user who logs in, no matter how, has full control. The easiest way to adjust these permissions is by using Windows Explorer as follows:

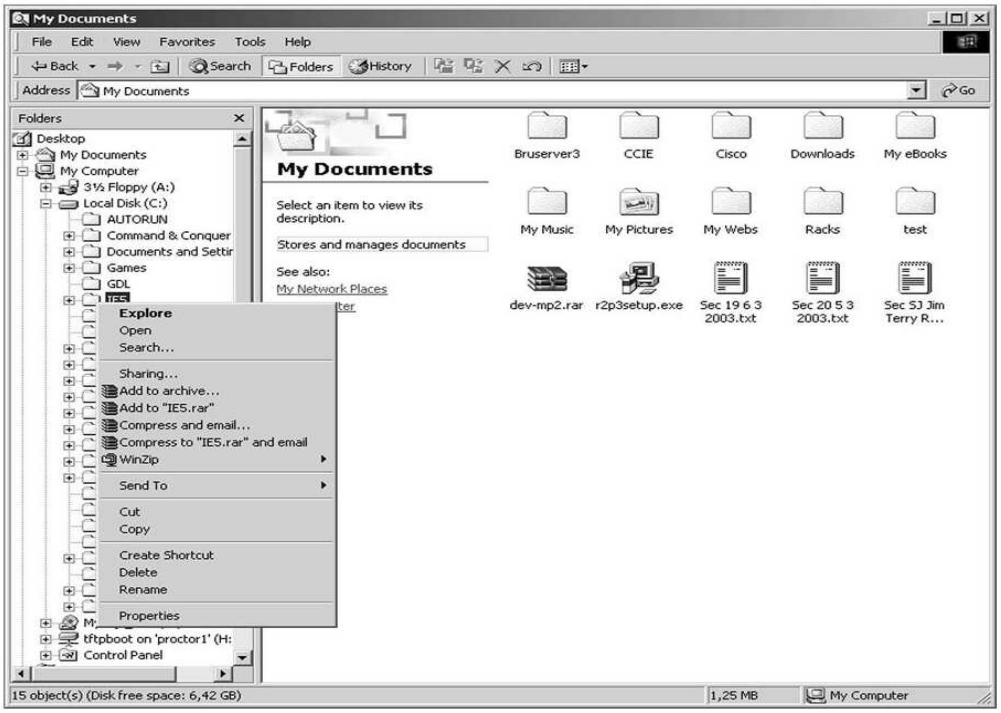


Figure 6.1: Windows Explorer

Step 1: Right-click the folder for which you want to change the permission. The pull-down choices are displayed in Figure 6.1.
Step 2. Select Properties from the pull-down choices. The screen shown in Figure 6.2 displays this option.

Figure 6.2: Properties Page
Step 3. Click the Security tab. The screen shown in Figure 7-3 displays this tab.

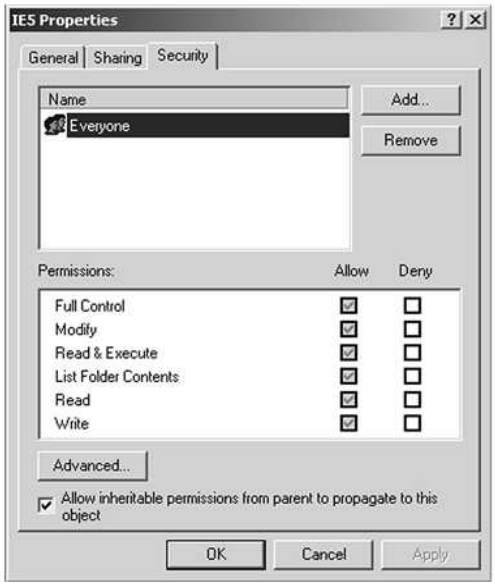


Figure 6.3: Security Tab

In Figure 6.3, you can see the default for Windows security. Every user logged in to the system has Full Control. This leaves the system wide open to any kind of unauthorized access. Therefore, you need to change those permissions. The case study in this chapter gives you an example of how to change these permissions.

There is much more to securing a web server than hardening the file system. Other things you need to do are

- Set account policies
- Edit group rights

`

- Rename critical accounts
- Turn on auditing
- Remove or disable unnecessary services

The last item in the list of tasks for securing the web server is removing or disabling unnecessary services. When you start your PC, many services run in the background. Disable all services that you do not need. Table 7-1 lists the services that you can disable. This is not a complete list, so be careful when disabling these services. Some services might be needed for operation.

Table 6.1. Services

Service Name	Description
ClipBook Viewer	Enables the ClipBook Viewer to create and share pages of data to be viewed by remote computers
Computer Browser	Maintains an up-to-date list of computers on your network and supplies the list to programs that request it
DHCP Client	Manages network configuration by registering and updating IP addresses and Domain Name Server (DNS) names for this computer
DHCP Server	Allocates IP addresses and allows the advanced configuration of network settings
DNS Server	Enables DNS name resolution
Fax Service	Enables you to send and receive faxes
File Server for Macintosh	Enables Macintosh users to store and access files on this Windows server machine
Gateway Service for Netware	Provides access to file and print resources on NetWare networks
Internet Connection Sharing	Provides NAT, addressing, and name resolution services for all computers on your home network
NetMeeting Remote Desktop Sharing	Allows authorized users to remotely access your Windows desktop
Print Server for Macintosh	Enables Macintosh clients to route printing to a print spooler located on a computer running Windows 2000 server
Print Spooler	Queues and manages print jobs
Remote Access Auto Connection Manager	Brings up a dialog box that offers to make a dialup connection to a remote computer when no network access exists
RPC Locator	Provides the name service for RPC clients
Remote Registry Service	Allows remote Registry manipulation
Routing and Remote Access	Offers routing services in local area and WAN environments
Run As Service	Allows you to run specific tools and programs with different permissions than your current logon provides
SAP Agent	Advertises network services on an IPX network
SMTP	Transports e-mail across the network
Simple TCP/IP Services	Implements support for Echo, Discard, Character Generator (CharGen), Daytime, and Quote of the day (QOTD)
Smart Card	Manages and controls access to a smart card
TCP/IP Print Server	Enables TCP/IP-based printing
Telephony	Provides Telephone API (TAPI) support for programs that control telephony devices
Telnet	Allows a remote user to log on to the system and run console programs using the command line
Windows Time Service	Sets the computer clock

DAC is a means of restricting access to information based on the identity of users and membership in certain groups. Access decisions are typically based on the authorizations granted to a user based on the credentials presented at the time of authentication (username, password, hardware/software token and so on). In most typical DAC models, owners of information or resources can change permissions at their discretion (thus the name). DAC's drawback is that

administrators cannot centrally manage these permissions on files and information stored on the web server. A DAC access control model often exhibits one or more of the following attributes:

- Data owners can transfer ownership of information to other users.
- Data owners can determine the type of access given to other users (read, write, copy, and so on).
- Repetitive authorization fails to access the same resource, or an object generates an alarm and restricts the user's access if auditing is turned on.
- Special add-on or plug-in software must be applied to an HTTP client to prevent indiscriminant copying by users (cutting and pasting of information).
- Users who do not have access to information should not be able to determine its characteristics (file size, filename, directory path, and so on).

6.2.2. Web Servers

A freshly installed web server is a completely defenseless platform. Before you can start using it as a web server, you need to secure it. This section shows you how. After the web server is installed, you can take several steps to secure it: You can prevent access to the server, and you can enable logging to monitor events on your web server.

6.2.2.1.Logging

Logging is an essential part of maintaining a secure web environment. To enable logging, open Internet Information Services in the Administrative tools menu, expand the tree, right-click Default Web Site, and choose Properties. Then click the Web Site tab to see the screen shown in Figure 6.4.

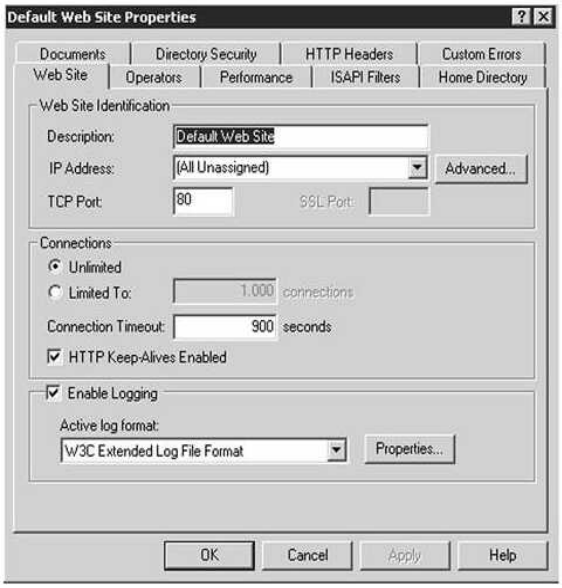


Figure 6.4: Default Web Site Properties

Near the bottom of the page, make sure that the Enable Logging check box is enabled. Internet Information Services (IIS) supports four log file formats, each with varying types and quantities of data collected. The default, W3C Extended Log File Format, is the most detailed. Now you can click Properties to bring up the screen in Figure 6.5.

Figure 6.5: Extended Logging Properties

In Figure 6.5, you can see that, by default, a new log file will be created every day. The default log file directory is %WinDir%\System32\LogFiles; however, you should change this to point to somewhere else preferably to another server. Log files should preferably be archived offline. Intruders usually hide their tracks by altering or deleting the log file. If intruders take control of your PC, a log in this location is vulnerable.

6.2.2.2.Restricting Access

You can restrict access to a website or to a specific folder of a website on a user-by-user basis or based on IP addresses. To configure access for user authentication, start the Internet Service Manager. Right-click the folder you want to use for basic authentication, which brings up a screen similar to Figure 7-6.

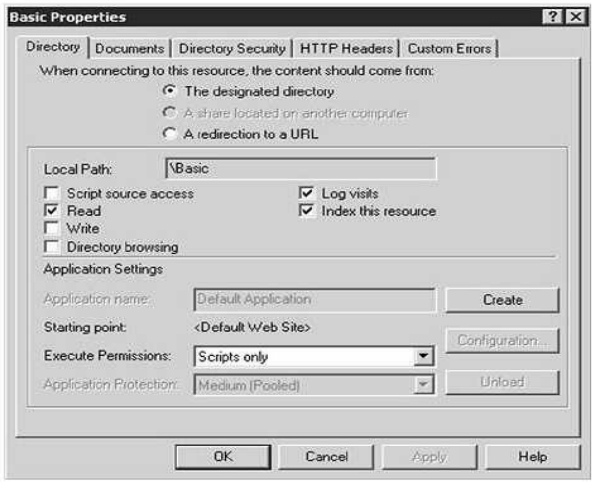


Figure 6.6: Folder Properties

On that screen, select the Directory Security tab. This brings you to a screen like the one in Figure 6.7, where you can edit the authentication method, IP address, or domain name restrictions.

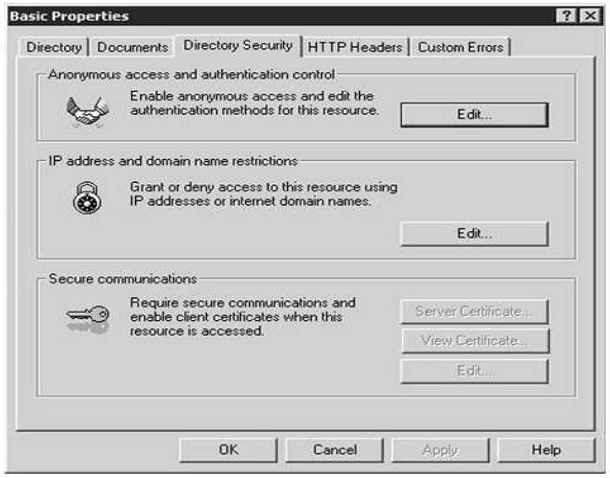


Figure 6.7: Directory Security

Click Edit for the anonymous access and authentication control to select the authentication method you want to use for that folder, as shown in Figure 6.8.

Figure 6.8: Authentication Methods

On the Authentication Methods screen, you can check boxes to indicate that anonymous access is allowed or to select basic authentication, for which the password is sent in clear text. You can also select to have integrated Windows authentication. To use integrated Windows authentication, add all the different users in Windows because IIS uses integrated Windows authentication to grant access to the website.

Access can also be controlled based on a PC's IP addresses. You can set specific addresses, address ranges, or DNS names from which access is either allowed or denied. After you click Edit IP addresses and domain name restrictions, you see a page, as shown in Figure 6.9.

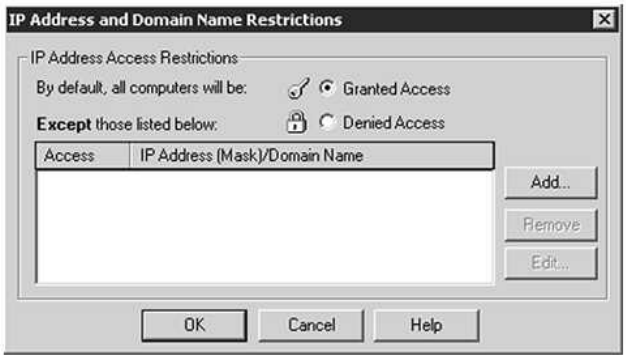


Figure 6.9: Authentication Methods

This dialog box needs careful reading. It either grants (the default) or denies access to all addresses except those you add manually. When you click Add, you see a screen as shown in Figure 6.10.

Figure 6.10: Deny IP Addresses

If you want to deny only one particular address, select Single computer; however, you can also restrict access to a group of computers or to a domain name. You can repeat these steps to exclude more than one domain or range.

6.2.3. Browsers

We all use browsers these days, and most of us run third-party plug-ins. This is not necessarily dangerous, but it is always better to keep in mind that malicious people can write plug-ins, too. The most popular scripting languages used for writing plug-ins today are the following:

- Java
- JavaScript
- VBScript
- ActiveX

Be very careful when installing plug-ins, just as you should be when downloading any software program from the Internet.

6.2.3.1.Security Zones

Because most people using the Internet today use Microsoft Internet Explorer to browse web pages, this chapter covers only that program. Internet Explorer has four zones of security. When you access a resource on another machine, the other machine's zone relative to yours is determined, and the restrictions placed on that zone control the interaction with that resource. Users can set the security policy on their computer. The four zones are as follows:

Internet Contains all websites that are not placed in another zone.

Local Internet Contains all the websites that are on your company's intranet. Here, you find all sites that have the same domain name as the one your PC is using.

Trusted sites Contains websites that you trust not to damage your data. If you want to have trusted sites, you need to add them manually.

Restricted This zone contains websites that you do not trust because they could potentially damage your data. This is also a list created manually.

To change the settings for these four zones in Internet Explorer, choose Tools > Internet Options. On the page that appears, select the Security tab, and you see a page as shown in Figure 6.11.

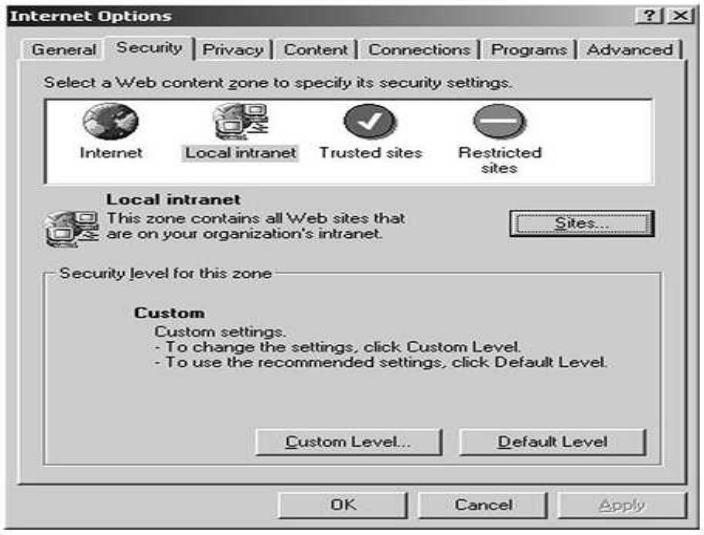


Figure 6.11: Security Setting Page

There are four predefined security levels. In addition, you have the ability to customize the settings for any or all the zones. Of the web content zones shown in Figure 6.11, the Internet zone is the one you need handle most carefully. The default setting here is Medium, which is not so secure for the World Wide Web. Table 7-2 lists all the security levels with a brief explanation of their purposes.

Level	Description
High	• This is the safest way to browse but also the least functional. • Less secure features are disabled. • Cookies are disabled. (Some websites do not work.) • This is appropriate for sites that might have harmful content.
Medium	• Browsing is safe and still functional. • Prompts before downloading potential unsafe content. • Unsigned ActiveX controls are not downloaded. • This is appropriate for most Internet sites.
Medium-low	• This is the same as Medium without prompts. • Most content is run without prompts. • Unsigned ActiveX controls are not downloaded. • This is appropriate for sites on your local network (intranet).
Low	• Minimal safeguards and warning prompts are provided. • Most content is downloaded and run without prompts. • All active content can run. • Appropriate for sites that you absolutely trust.

Because you cannot set the security level for the Internet zone to High, you must change the custom level. After you click the Custom Level button, you see a screen similar to that in Figure 6.12.

Following items can be changed in this window:

- ActiveX controls and plug-ins
- Cookies
- Downloads
- Microsoft VM
- Miscellaneous
- Scripting
- User authentication

Figure 6.12 shows Scripting. On this screen, you first change the custom settings on the bottom of the screen from Medium to High. At this point, you receive a warning asking if you are sure that you want to make this change. After you click Yes, you can take another look at the scripting options, as shown in Figure 6.13.

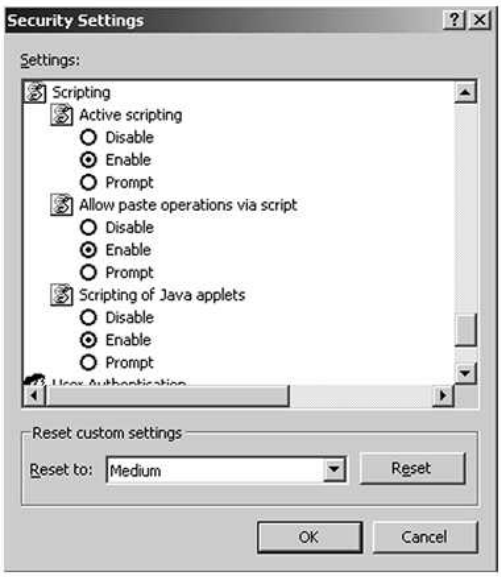


Figure 6.12: Scripting Options

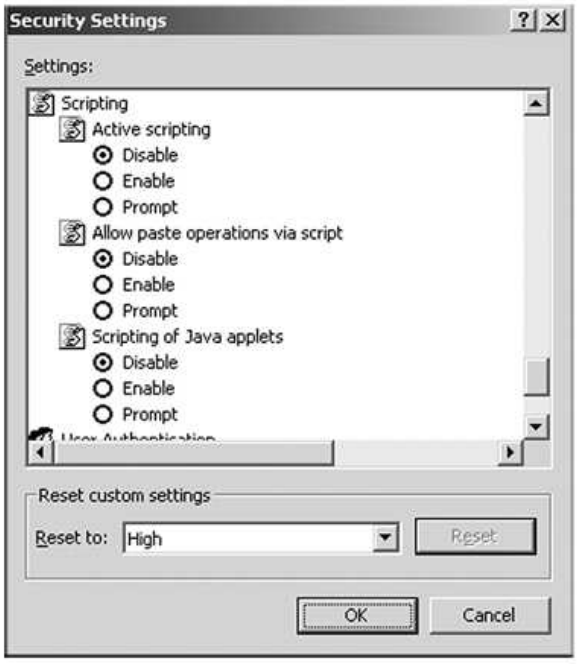


Figure 6.13: High Security Settings

some settings have changed. All the scripting items have been disabled. Be sure to check your browser every time you install a new version to ensure that these settings are correct. By disabling some features, such as ActiveX, you can occasionally cause a web page to generate an error. Most of the time, it is better to have that error than to let ActiveX run, but in some cases, you know the ActiveX controls can be trusted, and you need to let them work. You can do this by making the site a trusted site and by setting trusted site security so that ActiveX can run. To do that, you need to go back to the Security page of the Internet Options. After you click Trusted sites, you see a page as shown in Figure 6.14.

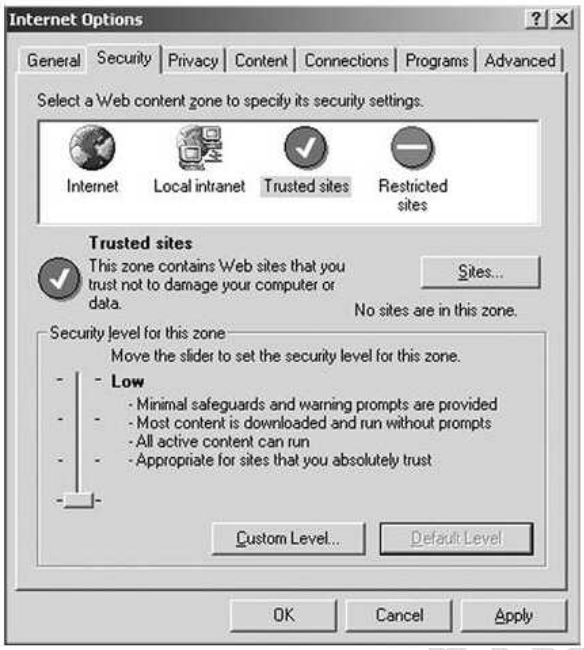


Figure 6.14: Security Setting Page

The default security for a trusted site is Low. You can set security to Medium-low or Medium to increase security. On that same page, you also need to add the site you trust. To do that, click Sites, which brings you to a screen as shown in Figure 6.15.

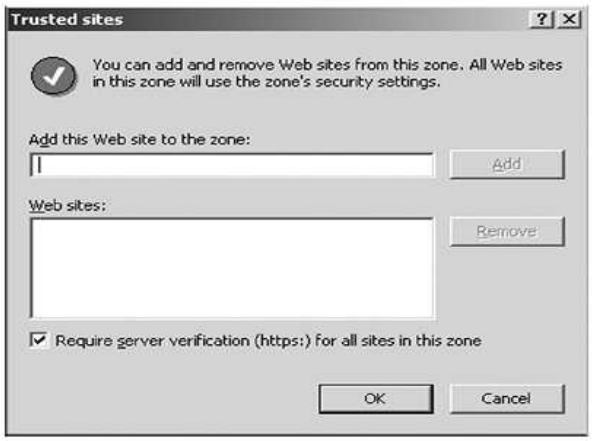


Figure 6.15: Trusted Sites

On that screen, you need to clear the check box requiring HTTPS, type in the domain of the site you trust, and then click Add. At this point, if you try to reload the page with the ActiveX content, it works and the content is visible.

6.2.3.2.Cookies

As you might already know, HTTP is a stateless protocol. Every time you visit a website, it looks as if that visit to the website is your first because HTTP does not keep track of your web history. To simulate a stateful environment, the HTTP protocol includes features such as cookies. There are two types of cookies:

Session cookie This cookie is created to keep track of what you buy when, for example, you visit an e-commerce website where you use a shopping cart. After you check out from that website, the session cookie is deleted from your browser memory.

Persistent cookie When you go to a website and see a personalized welcome message, you know that a persistent cookie is on your PC. These cookies contain information about you and your account. Often, that information is a key that is related only to a database with your profile.

You can manage cookies in several ways. You can delete all your cookies, or you can configure your browser to not accept cookies at any time. This would make browsing the Internet rather difficult because many sites need cookies to function properly. A better solution would be to force all your cookies to be session cookies. You can do this by making the folder where the cookies are stored read-only. Your browser will accept them but will be unable to save them to disk.

6.3. Conclusion

In this chapter, you learned that some trivial things, such as computer file systems, or common actions, such as browsing the Internet, are highly vulnerable to intruders. Every freshly installed system is like a house with all doors and windows open. It is the user's duty to close all these doors and windows.

6.4. Important Questions:

1: What is the difference between a right and a permission?
2: What can be done on a web server to make it more secure against intruders?
3: What is DAC?
4: How can you enable logging on your IIS web server?
5: What two methods restrict access to an IIS web server?
6: List three popular scripting languages used on web servers that are executed by browsers when visiting the site.
7: Describe the four security zones that are available in Internet Explorer.
8: Briefly describe the four predefined security levels in Internet Explorer.

9: What is the difference between session cookies and persistent cookies?
10: What is the best way to handle cookies?

Chapter 07
Router Security

7.1. Basic Router Security

Router security is the process to protect the router itself from being accessed by unauthorized persons. For example, a router could be configured to protect the network behind it, but an intruder could access the router easily because of the weak passwords that were used or some services the administrator forgot to turn off. In this case, the network behind that router is no longer safe because the intruder can easily change the router's configuration to gain access to the network behind it.

7.1.1. Administrative Access

Configuring administrative access is an extremely important security task. Otherwise, an unauthorized person could alter the routing parameters, change access lists, and gain access to other systems in the network. To perform basic router configuration tasks, access via a console is required. A console is a terminal that is connected to a router console port and can be either a dumb terminal or a PC running terminal emulation software. Consoles are just one way administrators obtain access to routers. Access can also be gained by Telnet, Hypertext Transfer Protocol (HTTP), and Simple Network Management Protocol (SNMP) if these services are turned on.

The first step in securing administrative access is to configure secure system passwords. These passwords can be stored either on the router itself or remotely on an authentication, authorization, and accounting (AAA) server. This chapter covers only the configuration of local passwords. Passwords should be as strong as possible. Never use existing words, birthdays, or names that are easy to guess. Most companies have creation rules for passwords in their security policies, such as how often a password must change and which characters have to be used in passwords.

There are two commands available to configure a password on a Cisco router.

enable password password

enable secret secret

If both commands are configured, the password is ignored and only the secret is used. Using enable secret is more secure than using enable password because enable secret hashes the password in the router configuration file. To hash the password, it uses a strong hashing algorithm based on MD5. When looking at the configuration file after using the enable secret command, you see only the hash and not the password anymore.

If you forget the enable secret or password, you will not be able to configure the router anymore. The only solution is to use the password-recovery procedure.

Also, Cisco routers support multiple Telnet sessions, up to five simultaneous sessions by default but more can be added. Each session is serviced by a logical virtual type terminal (VTY) line. By default, Cisco routers do not have any user-level password configured for these VTY lines. If an administrator does not configure a password on the VTY lines, no access to the router is available via Telnet, and you encounter an error message similar to Example 8-2.

Some routers also have an auxiliary port that is sometimes used by administrators to remotely configure and monitor the router using a dialup modem connection. Setting a password on this port is one of several steps that have to occur when configuring this port for remote dialup. This process is beyond the scope of this book.

By default, an administrative interface stays active for 10 minutes after the last session activity. After that, the interface times out and logs out. It is recommended that you fine-tune these timers. They can be configured by using the exec-timeout command in line configuration mode for each of the line types used. You can specify how long a user can be inactive by the minutes and the seconds after the exec-timeout command

The console port has an exec-timeout of 0 0, which means that it never times out. You have to be careful when using this timeout. All router passwords are stored in clear-text form by default, as you can see in Example 8-4, with the exception of the enable secret. These passwords can also be

seen by a network monitor if your configuration file traverses the Internet. By using the service password-encryption command, all passwords are encrypted using a proprietary Cisco algorithm indicated by the number 7 when viewing the configuration file, as seen in Example 8-5. This method is not as safe as MD5, which is used for the enable secret, but it makes it harder for the intruder to gain access to the router.

Another useful feature that can be used is the banner. The banner does not protect the router from intruders, but by using it, you can warn intruders that the device is for authorized people only.

To enter a banner in configuration mode, use the following command:

banner {exec | incoming | login | motd | slip-ppp} d message d

Table 7.1: Banner Command

Command	**Description**
banner exec	Specifies a message to be displayed when an EXEC process is created (a line is activated or an incoming connection is made to a VTY line).
banner incoming	Specifies a message used when you have an incoming connection to a line from a host on the network.
banner login	Specifies a message to be displayed before the username and password login prompts.
banner motd	Specifies and enables a message-of-the-day (MOTD) banner.
banner slip-ppp	Specifies and enables a banner to be displayed when a Serial Line Interface Protocol (SLIP) or PPP connection is made.
d	Represents a delimiting character of your choice (for example, a pound sign #). You cannot use the delimiting character in the banner message.
message	Represents message text. There are some tokens available to use in the message text: • $(hostname): Displays the hostname for the router • $(domain): Displays the domain name for the router • $(line): Displays the VTY line number • $(line-desc): Displays the description attached to the line

7.1.2. Services

Cisco routers run several services that may or may not be required in certain networks. Network security can be greatly improved by turning them off or at least restricting access to them. One of the most basic rules of router security is to run only the services that are really necessary and no more. Leaving unused network services enabled increases the possibility of those services being used maliciously. The services in the list that follows are all enabled by default on a router.

By default, the services that are enabled on a router differ based on the Cisco IOS version that router is running. For this example, Cisco IOS version 12.2 was used.

1. **BOOTP server** This service allows a router to act as a BOOTP server for other routers. This is rarely required and should be disabled. Use the following command to disable this service:
   Brussels(config)#no ip bootp server
2. **Cisco Discovery Protocol (CDP)** This is primarily used to obtain protocol addresses of neighboring devices and the platforms on which they are used. CDP is media- and protocol-independent and runs on all Cisco equipment, including routers, switches, and access servers. Use the following commands to disable CDP:
   Brussels(config)#no cdp run
   Brussels(config-if)#no cdp enable
   The first command is used to disable CDP globally, and the second command is used to disable it on a per interface basis.

3. **DNS lookup** By default, Cisco routers broadcast name requests to 255.255.255.255. If the DNS service is used, make sure that the proper DNS server address is configured. Use the following command to turn off the DNS service:
   Brussels(config)#no ip domain-lookup
4. **HTTP server** The default setting for this device depends on the platform. This service enables a network administrator to modify the configuration using a web browser. You should disable this service if not in use by using the following command:
   Brussels(config)#no ip http server
5. **IP redirect** This feature enables the sending of redirect packets if the router is forced to resend a packet through the same interface on which it was received. This can be used to map the network and should be turned off on interfaces to untrusted networks. This can be disabled using following command:
   Brussels(config-if)#no ip redirects

This is only a selection of the many services that run on a router. Make sure that you use only what you need to run a network and that everything else is turned off.

7.2. Router Security to Protect the Network

All the topics discussed to this point in the chapter have covered the different steps that an administrator needs to take to protect the router itself. The next step you need to learn is how to configure the router to protect the network behind it. This can be done by using access lists or enhanced access lists, such as dynamic or time-based access lists. If a device is running a security image, those networks can also be protected by using Context-Based Access Control (CBAC).

7.2.1. Access Lists

On a router, access lists are used as packet filters to decide which packets can go across a certain interface. Packets that are allowed on an interface are called permitted packets and packets that are not allowed are called denied packets. Access lists can consist of one or more statements that determine what data is permitted and denied on an interface. The statements are known as Access Control Entries (ACE). It is important to use well-written access lists to restrict access because Cisco router security is highly dependent on them for filtering packets as they travel across the network.

A router can identify an access list by either a name or a number. Table 7.2 lists some of the commonly used access list numbers and their associated types.

Table 7.2: Access List Numbers

Access List Number	Type
199	IP standard access list
100199	IP extended access list
800899	IPX standard access list
10001099	IPX SAP access list
13001999	IP standard access list (expanded range)
20002699	IP extended access list (expanded range)

Starting with Cisco IOS version 11.2, access lists can be identified by a name rather than just by a number. By using named access lists, you can identify an access list more easily than if you are using numbered access lists alone. The command syntax for named access lists is also slightly different. As stated in Table 8-2, there are two types of IP access lists:

Standard IP access lists: This type can filter IP packets based on the source address only.

Extended IP access lists: This type can filter IP packets based on several attributes, including the following:

- Source IP address
- Destination IP address

- Source TCP or UDP port
- Destination TCP or UDP port
- Protocol

The command syntax for a standard numbered access list is as follows:
access-list access-list-number {deny | permit} source [source-wildcard]
Table 7.3 describes the commands you can use when configuring a numbered access list.

Table 7.3: Numbered Access List Command

Command	Description
access-list-number	Serves dual purposes: • It is the number of the access list. • It specifies that this is a standard IP access list.
Deny	Drops all packets matching the specific source address.
Permit	Allows all packets matching the specific source address to flow through the interface.
Source	Specifies the IP address of a host or group of hosts (if a wildcard mask is specified).
source-wildcard	The wildcard mask is applied to the source group of hosts whose packets are to be examined.

Access lists must be applied to a router interface to take effect. When an access list is applied to an interface, you also have to configure the direction of the data flow, as shown in Figure 7.1.

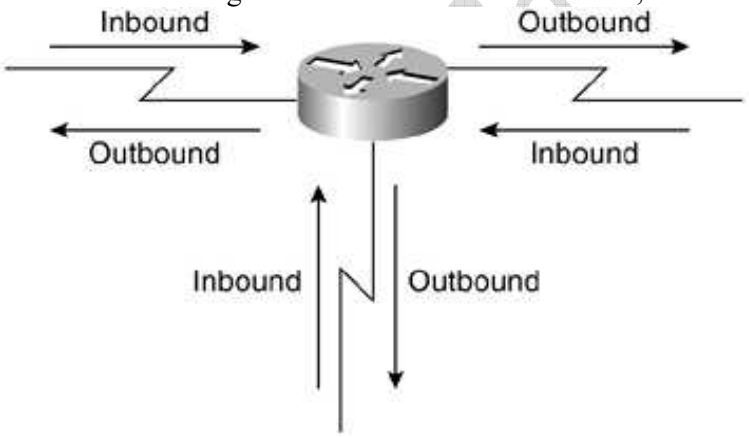


Figure 7.1: Access List Direction

As you can see in Figure 7.1, there are two directions:

- **Inbound** The access list is applied to packets flowing toward the router interface.
- **Outbound** The access list is applied to packets flowing away from the router interface.

The interface command to apply an access list to an interface is as follows:
ip access-group {access-list-number | access-list-name} { in | out }

Table 7.4 describes the keywords you can use when assigning the access list to an interface.

Table 7.4: Access Group Keywords

Keyword	**Description**
access-list-number	Number of the IP standard or extended numbered access list
access-list-name	Name of the IP standard or extended named access list
In	Filters on inbound packets
Out	Filters on outbound packets

7.2.2. Enhanced Access Lists

Several types of enhanced access lists can be configured on a router. So far, only standard and extended access lists have been discussed in this chapter. Enhanced access lists were designed to secure routers and their networks better. They all have special features, and selection depends on your particular needs for security. The following types of access lists are available:

- Dynamic access lists
- Time-based access lists
- Reflexive access lists

7.2.2.1. Dynamic Access Lists

Dynamic access lists, also known as lock-and-key, create specific, temporary openings in response to user authentication. It is highly recommended to use a TACACS+ server for the authentication of the user. TACACS+ provides authentication, authorization, and accounting services. In the example illustrated in Figure 7.2, no TACACS+ server has been included for authentication for the sake of simplicity. Figure 7.2 shows a user connected to the Internet. The user is trying to connect to a device in the internal network.

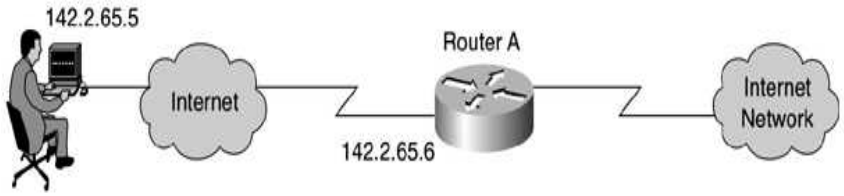


Figure 7.2: Dynamic Access List

To be able to connect to the device, the user needs a dynamic access list on Router A and a username for local authentication. Configure a username so that the user can access the device by using following command:

```
Tokyo(config)#username user password te5t
```

Because you should not count on the user to issue the access-enable command correctly, you need the line that follows under vty 0 4. The access-enable command is used to create a temporary access list entry in a dynamic access list.

The autocommand used in this example is executed immediately when a user logs in via Telnet access.

You can define an extended access list that is applied when any user logs in to the router and the access-enable command is issued. The maximum absolute time for this hole in the filter is set to 15 minutes. After 15 minutes, the hole closes whether or not anyone is using it. The name dyntest is needed but is not significant.

```
Tokyo(config)#access-list 101 dynamic dyntest timeout 15 permit ip any any
```

After that, define the access list needed to block everything except the ability to use Telnet to access the router. Users must telnet into this router to authenticate themselves as valid users. Therefore, the following line is needed for users to be able to telnet into this router:

```
Tokyo(config)#access-list 101 permit tcp any host 142.2.65.6 eq telnet
```

Now you only have to apply the access list to the interface on which users are coming.

```
Tokyo(config)#interface FastEthernet0/0
Tokyo(config-if)#ip access-group 101 in
```

When using the show access-lists command, the access list looks like this before any user has used Telnet to reach the router:

7.2.2.2. Time-Based Access Lists

In a time-based access list, the hole is created for a certain amount of time.

This example allows users coming in on Ethernet 0/0 to have web access from 8:00 to 18:00 during all weekdays. Instead of weekdays, you can use several other keywords, such as the following:

Friday Friday
Monday Monday
Saturday Saturday
Sunday Sunday
Thursday Thursday
Tuesday Tuesday
Wednesday Wednesday
daily Every day of the week
weekdays Monday thru Friday
weekend Saturday and Sunday

7.2.2.3.Reflexive Access Lists

With reflexive access lists, you have the ability to filter network traffic at a router, based on IP upper-layer protocol session information. Reflexive access lists can be defined by extended named IP access lists only. You cannot define reflexive access lists with numbered or standard named access lists. Reflexive access lists have significant differences from other types of access lists. They contain only temporary entries. These entries are automatically created when a new IP session begins and are removed when the session ends. Reflexive access lists are not applied directly to the interface, but are nested within an extended named IP access list that is applied to that interface. The syntax to define a reflexive access list is as follows:

```
ip access-list extended name
permit protocol any any reflect reflection-name [timeout seconds]
```

Define the reflexive access list using the permit entry and the reflect option. Then you can apply the extended access list to an interface. After you define a reflexive access list in one IP extended access list, you must nest the reflexive access list within a different extended named IP access list with the evaluate command. Example 8-9 should make that procedure clear.

7.3. CBAC

The Cisco IOS Firewall CBAC engine provides secure, per-application access control across network perimeters. CBAC allows administrators to implement firewall intelligence as part of an integrated, single-box solution.

CBAC works to provide network protection on multiple levels using the following functions:

- Traffic filtering CBAC intelligently filters TCP and UDP packets based on information of the application-layer protocol session. Using CBAC, Java blocking can be configured to filter HTTP traffic based on server address or to completely deny access to Java applets.
- Traffic inspection CBAC inspects traffic that travels through the firewall to discover and manage state information for the TCP and UDP sessions. This state information is used to create temporary openings in the firewall's access lists to allow return traffic and additional data connections for permissible sessions. Inspecting packets at the application layer and maintaining TCP and UDP session information provide CBAC with the ability to detect and prevent certain types of network attacks, such as SYN-flooding.
- Alerts and audit trials CBAC also generates real-time alerts and audit trails. Using CBAC inspection rules, you are able to configure alerts and audit trails on a per-application protocol basis.

CBAC does not provide intelligent filtering for all protocols. It works only for the specified protocols. If you do not specify a certain protocol for CBAC, the existing access lists determine how that protocol is filtered. No temporary openings are created for protocols not specified for CBAC inspection.

To configure CBAC, the following tasks are required:

- Pick an interface internal or external.
- Configure an IP access list on that interface.

- Configure global timeouts and thresholds.
- Define an inspection rule.
- Apply the inspection rule to an interface.
- Configure logging and audit trail.

Picking an interface means that you will have to decide whether you configure CBAC on the internal or external interface of your firewall. Internal refers to the side where sessions must originate. External is the side where sessions cannot originate. Sessions originating from the external side are blocked. If you want to configure CBAC in two directions, you have to configure it in one direction first. When you configure it in the other direction, the interface designations are swapped. In Figure 7.3, you can see a simple topology in which CBAC is configured on the external interface. In Figure 7.4, CBAC is configured for the internal interface.

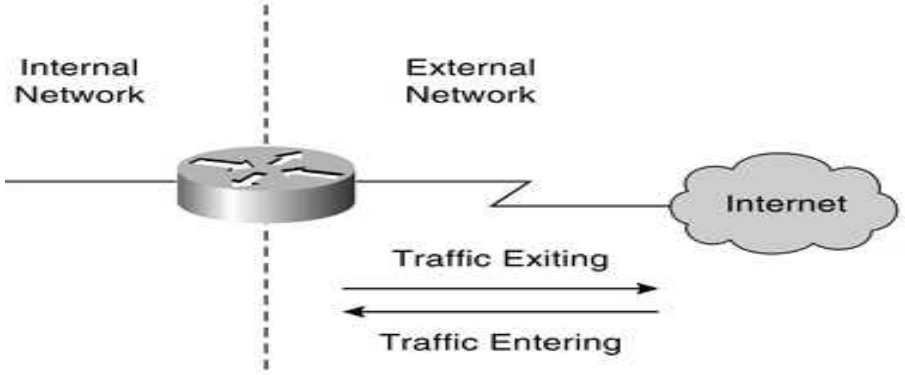


Figure 7.3: CBAC at the External Interface

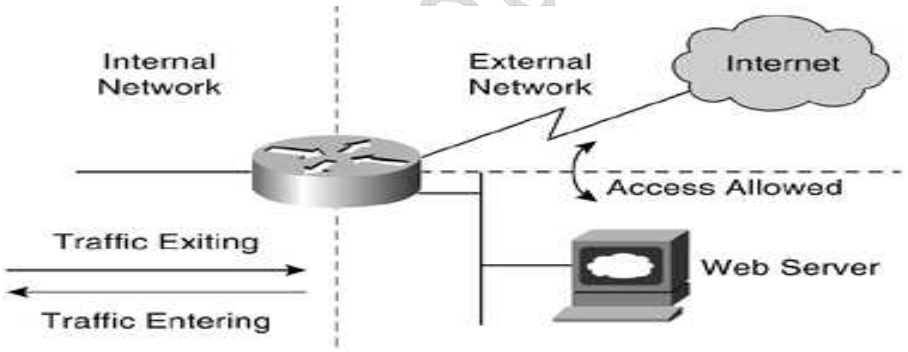


Figure 7.4: CBAC at the Internal Interface

CBAC uses timeouts and thresholds to determine how long to manage state information for a session and when to drop sessions that do not become fully established. These timeouts and thresholds apply globally to all sessions.

Table 7.5 describes the different inspect commands that are available on a Cisco router.

Table 7.5: inspect Command

Command	Description
ip inspect tcp synwait-time seconds	The length of time the software waits for a TCP session to reach the established state before dropping the session (default 30).
ip inspect tcp finwait-time seconds	The length of time a TCP session is still managed after the firewall detects a FIN-Exchange (default 5).
ip inspect tcp idle-time seconds	The length of time a TCP session is still managed after no activity occurs (default 3600).
ip inspect udp idle-time seconds	The length of time a UDP session is still managed after no activity occurs (default 30).
ip inspect dns-timeout seconds	The length of time a DNS name lookup session is still managed after no activity occurs (default 5).
ip inspect max-incomplete high number	The number of existing half-open sessions that cause the software to start deleting half-open sessions (default 500 existing half-open sessions).
ip inspect max-incomplete low number	The number of existing half-open sessions that cause the software to stop deleting half-open sessions (default 400 existing half-open sessions).
ip inspect one-minute high number	The rate of new sessions that causes the software to start deleting half-open sessions (default 500 half-open sessions per minute).
ip inspect one-minute low number	The rate of new sessions that causes the software to stop deleting half-open sessions (default 400 half-open sessions per minute).
ip inspect tcp max-incomplete host number block-time minutes	The number of existing half-open sessions with the same destination host address that cause the software to start dropping half-open sessions to the same destination host address (default 50 existing half-open TCP sessions).

7.4. Conclusion

As you can understand from reading this chapter, there are several ways to protect a router from being accessed by unauthorized persons. There are also many solutions for protecting the network behind a router. The method you use depends on the level of protection needed.

7.5. Important Questions

1: Give two commands to configure an enable password on a router.
2: Name three services that are running on a router that should be turned off if they are not used.
3: Name the different types of access lists that can be used.
4: What are dynamic access lists?
5: What is CBAC used for when it is configured on a router?
6: List five tasks to configure CBAC.
7: What does the ip inspect max-incomplete high command do?
8: Give three different types of enhanced access lists.
9: What can be filtered with reflexive access lists?
10: How can reflexive access lists can be defined?

Chapter 08
Firewall

8.1. Introduction

Protecting the confidentiality of information, preventing unauthorized access, and defending against external and internal attacks remain primary concerns of all network managers today. IT departments must defend against these threats. All network architectures should be based on sound security policies designed to address all the weaknesses and threats that can occur in today's large IP-based networks. Because of the ever-changing nature of remote connectivity especially with the increased use of virtual private networks (VPNs)and the requirement for instant access to core network resources, networks have policies that allow access to the Internet, where the amount of busy or noisy traffic from non-legitimate devices is vast. Firewalls play important roles in defending against these threats.

Every network should be based on a sound security policy. The security policy should describe firewalls in detail and, more specifically, the location, placement, and configuration of firewalls in the network, as well as whether the firewall is hardware based, software based, or even PC based.

Network vulnerabilities must be constantly monitored, found, and addressed because they define points in the network that are potential security weak points (or loopholes) that can be exploited by intruders or hackers. All networks are possible targets because an intruder's motivation can be based on a number of factors cash profit; revenge; vandalism; cyber terrorism; the excitement of a challenge; the search for prestige, notoriety, or experience; curiosity; or the desire to learn the tools of trade, just to name a few.

Sometimes the biggest security threat comes from within an organization, in particular from displeased employees who gain access to internal systems by abusing usernames and passwords. Identification of the weak points of the network and, therefore, the placement and configuration of the firewall are extremely important.

Now that you are aware of some of the reasons a network must have a sound security policy and why intruders (hackers) want to exploit a poorly designed network, let's discuss some of the firewall features and definitions before moving on to some of the available firewalls in today's marketplace.

8.2. Firewall Basics

A firewall is defined as a gateway or access server (hardware- or software-based) or several gateways or access servers that are designated as buffers between any connected public network and a private network. A firewall is a device that separates a trusted network from an untrusted network. It may be a router, a PC running specialized software, or a combination of devices. A Cisco firewall router primarily uses access lists to ensure the security of the private network.

Figure 8.1 displays a network in which firewalls are typically located between the trusted networks and untrusted networks.

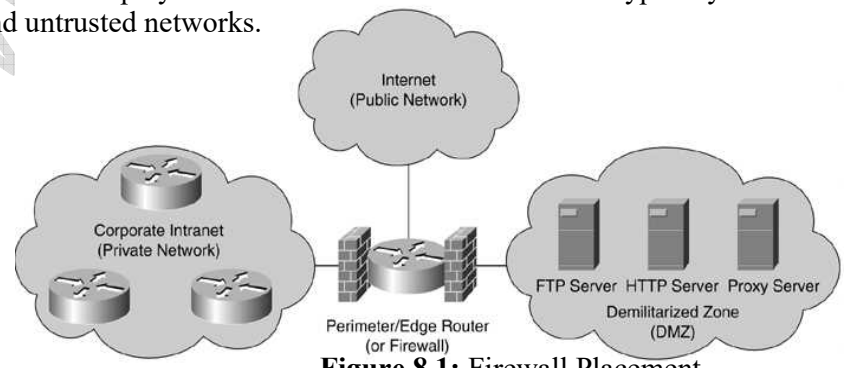


Figure 8.1: Firewall Placement

Data-driven, application-layer attacks have proliferated in recent years, with a dramatic rise in the late 1990s and the 21st century. With this increase, it has become clear that the existing solution set that was based on access lists is not adequate to counter these threats in a cost-efficient manner. Standalone devices are becoming an integral part of implementing effective security. Firewalls are primarily designed to address the countless threats posed to an organization's network by permitting access only to valid traffic.

8.2.1. Common Attacks:

Identifying valid traffic is a difficult task, and therefore security personnel should be well aware of existing intrusion techniques and attacks. Just as a reference, the following list presents a brief overview of common attack types.

- **TCP SYN flood attacks**: This form of denial-of-service (DoS) attack randomly opens up a number of TCP ports to make network devices use CPU cycles for bogus requests. By tying up valuable resources on the remote host (both CPU cycles and memory), the CPU is busy with bogus requests. In turn, legitimate users are affected by denial of access or poor network response. This type of attack renders the host unusable.
- **E-mail attacks:** This form of DoS attack sends a random number of e-mails to a host. E-mail attacks are designed to fill inboxes with thousands of bogus e-mails (also called e-mail bombs), thereby ensuring that the end user cannot send or receive legitimate mail.
- **CPU-intensive attacks**: This form of DoS attack ties up system resources by using programs such as Trojan horses (programs designed to capture usernames and passwords from a network) or enabling viruses to disable remote systems.
- **Teardrop:** A teardrop attack exploits an overlapping IP fragment implementation bug in various operating systems. The bug causes the TCP/IP fragmentation reassembly code to improperly handle overlapping IP fragments, causing the host to hang or crash.
- **DNS poisoning:** In this attack, the attacker exploits the DNS server, causing the server to return false IP addresses to a domain name query.
- **UDP bomb:** A UDP bomb causes the kernel of the host operating system to panic and crash by sending a field of illegal length in the packet header.
- **Distributed denial-of-service (DDoS):** This attack uses DoS attacks run by multiple hosts. The attacker first compromises vulnerable hosts using various tools and techniques. Then the actual DDoS attack on a target is run from the pool of all these compromised hosts.
- **Chargen attack:** This type of attack causes congestion on a network (high bandwidth utilization) by producing a high-character input after establishing a User Datagram Protocol (UDP) service or, more specifically, the chargen service.
- **Out-of-band attacks**: Applications or even operating systems such as Windows 95 have built-in vulnerabilities on data port 139 (known as WinNuke) if the intruders can ascertain the IP address.
- **Land.C attack**: This attack uses a program designed to send TCP SYN packets (TCP SYN is used in the TCP connection phase) that specify the target's host address as both source and destination. This program can use TCP port 113 or 139 (source/destination), which can also cause a system to stop functioning.
- **Spoof attack:** In a spoof attack, the attacker creates IP packets with an address found (or spoofed) from a legitimate source. This type of attack can be powerful when a router is connected to the Internet with one or more internal addresses.
- **Smurf attack:** The Smurf attack, named after the exploitive Smurf software program, is one of the many network-level attacks against hosts. In this attack, an intruder sends a large amount of Internet Control Message Protocol (ICMP) echo (ping) traffic to IP broadcast addresses, all of it having the spoofed source address of a victim.

- Smurf attacks include a primary and a secondary victim and are extremely potent and damaging to any IP network.

- **Man-in-the-middle attack:** With a man-in-the-middle attack, an intruder intercepts traffic that is in transit. The intruder can then either rewrite the traffic or alter the packets before the packets reach the original destination.

8.3. Different Types of Firewalls

Companies such as Cisco and other major vendors have introduced a multitude of firewall products that are capable of monitoring traffic using different techniques. Some of today's firewalls can inspect data packets up to Layer 4 (TCP layer). Others can inspect all layers (including the higher layers) and are referred to as deep packet firewalls. This section defines and explains these firewalls. The three types of inspection methodologies are as follows:

- Packet filtering and stateless filtering
- Stateful filtering
- Deep packet layer inspection

8.3.1. Packet Filtering

Packet filters (basic access-list filters on routers) are now easy to break, hence the introduction of proxy servers that limit attacks to a single device. A proxy server is a server that sits between a client application, such as a web browser, and a real server. It intercepts all requests to the real server to see if it can fulfill the requests itself. If not, it forwards the request to the real server. A proxy requests a connection to the Internet based on requests from internal or hidden resources. Proxy servers are application based, slow, and difficult to manage in large IP networks. The next generation of packet filters is stateless firewalls. Basically, a stateless firewall permits only the receipt of information packets that are based on the source's address and port from networks that are trusted.

8.3.2. Stateless Filtering

A stateless firewall was introduced to add more flexibility and scalability to network configuration. A stateless firewall inspects network information based on source and destination address. Figure 8.2 illustrates the inspection depth of a packet filter or stateless firewall. Packets are inspected up to Layer 3 of the OSI model, which is the network layer. Therefore, stateless firewalls are able to inspect source and destination IP addresses and protocol source and destination ports.

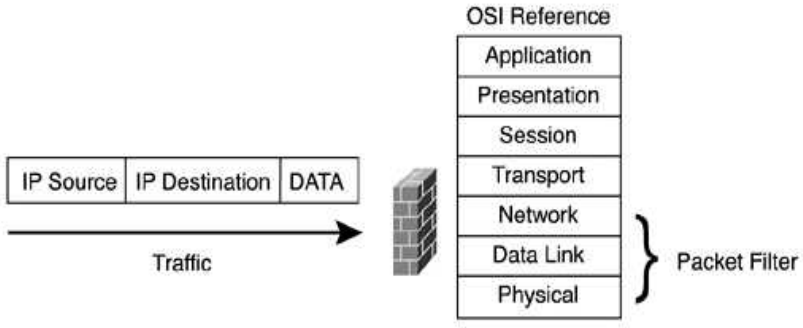


Figure 8.2: Stateless Firewall

8.3.3. Stateful Filtering

A stateful firewall limits network information from a source to a destination based on the destination IP address, source IP address, source TCP/UDP port, and destination TCP/UDP port. Stateful firewalls can also inspect data content and check for protocol anomalies. For example, a stateful firewall is much better equipped than a proxy filter or packet filter to detect and stop a denial-of-service attack. A proxy filter or packet filter is ill-equipped and incapable of detecting such

an attack. Because the source and destination address are valid, the data is permitted through whether it is legitimate or an attempted hack into the network. Figure 8.3 illustrates the inspection depth of a stateful firewall. Packets are inspected up to Layer 4 of the OSI model, which is the transport layer. Therefore, stateful firewalls are able to inspect protocol anomalies.

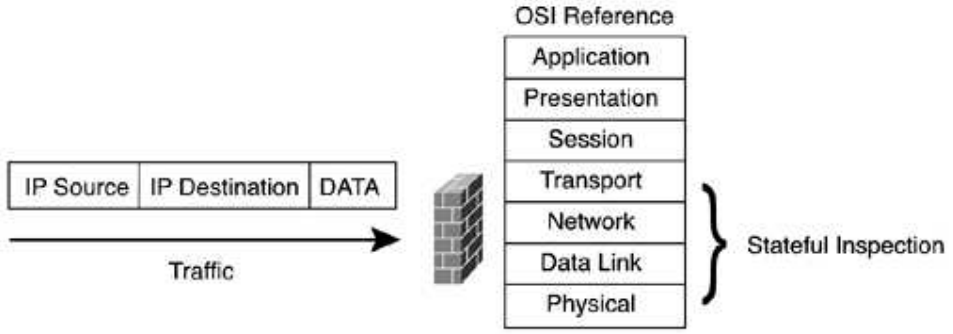


Figure 8.3: Stateful Firewall

8.3.4. Deep packet layer inspection:

With deep packet layer inspection, the firewall inspects network information from a source to a destination based on the destination IP address, source IP address, source TCP/UDP port, and destination TCP/UDP port. It also inspects protocol conformance, checks for application-based attacks, and ensures integrity of the data flow between any TCP/IP devices. The Cisco Intrusion Detection System (IDS), which is discussed in Chapter 10, "Intrusion Detection System Concepts," and NetScreen firewall products support deep packet layer inspection. The Cisco PIX Firewall supports stateless and stateful operation, depending on your product. Please refer to the Cisco website for the specific support for your product. Figure 8.4 displays how a device inspects packets with deep packet layer inspection.

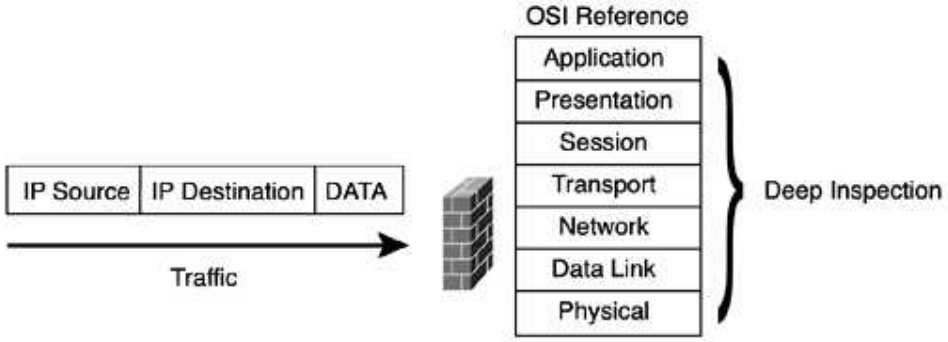


Figure 8.4: Deep Packet Layer Firewall

Figure 8.4 displays how a deep packet layer device inspects packets to

- Ensure that the packets conform to the protocol
- Ensure that the packets conform to specifications
- Ensure that the packets are not application attacks
- Police integrity check failures

Typically, these functions are performed in hardware or are ASIC based and are extremely fast. Any data that matches criteria such as that defined for DoS is dropped immediately and can be logged to an internal buffer, e-mailed to the security engineers, or can send traps to an external Network Management Server (NMS).

8.3.5. Hardware Firewalls: PIX and NetScreen

This section covers two of the most common hardware-based firewalls in the marketplace today, namely the CiscoSecure Private Internet Exchange (PIX) Firewall and the NetScreen firewall.

8.3.5.1.PIX

The PIX is a dedicated hardware-based networking device that is designed to ensure that only traffic that matches a set of criteria is permitted to access resources from networks defined with a secure rating. The PIX Firewall was an acquisition by Cisco Systems in the 1990s. The command-line interface (CLI) is vastly different from Cisco IOS, although recent software developments have made the CLI closer to the traditional Cisco IOS syntax that most readers are familiar with.

The Cisco PIX and Cisco IOS feature sets are designed to further enhance a network's security level. The PIX Firewall prevents unauthorized connections between two or more networks. The latest released versions of Cisco code for the PIX Firewall also perform many advanced security functions such as authentication, authorization, and accounting (AAA) services, access lists, VPN configuration (IPSec), FTP logging, and Cisco IOS-like interface commands. All these features are discussed in the remaining chapters of this book. In addition, the PIX Firewall can support multiple outside or perimeter networks in the demilitarized zones (DMZs).

It is mnemonically convenient to make E0 the "0"utside interface and E1 the "1"nside. On a PIX with additional interfaces, the interfaces are usually separate service subnets or additional inside networks. Other vendors follow the same methodology, although they rename their interfaces to names that are configurable, such as the "Internet" interface.

Typically, the Internet connection is given the lowest level of security, and a PIX ensures that only traffic from internal networks is trusted to send data. By default, no data is permitted at all. Therefore, the biggest problem or issue with a PIX Firewall is misconfiguration, which most crackers use to compromise network functionality. Figure 8.5 illustrates the different PIX interfaces and connections.

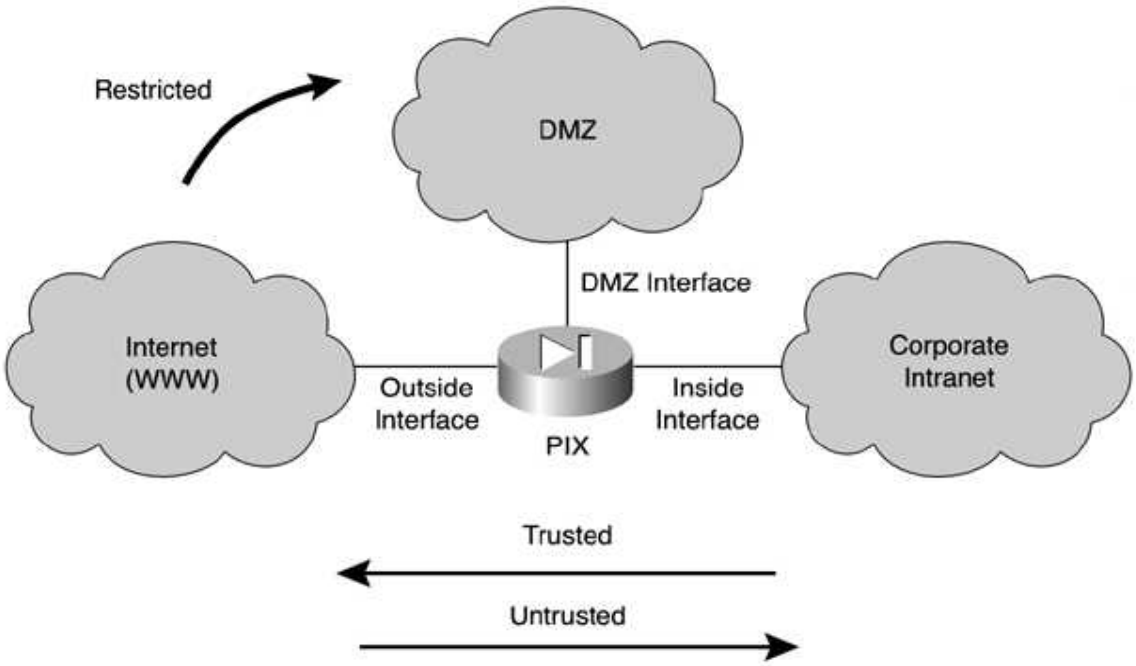


Figure 8.5: PIX Interfaces

A PIX Firewall permits a connection-based security policy. For instance, you might allow Telnet sessions to be initiated from within your network but not allow them to be initiated into the network from outside the network.

The PIX Firewall's popularity stems from the fact that it is solely dedicated to security. A router is still required to connect to wide area networks (WANs), such as the Internet, and to perform additional routing tasks and processes (recent versions of PIX OS do support some routing protocols). Some companies also use the PIX Firewalls for internal use to protect sensitive networks such as those of payroll or human resources departments.

`

As previously mentioned, the Cisco PIX Firewall is a stateful inspection device and bases all its decisions on a Cisco propriety algorithm, namely the Adaptive Security Algorithm (ASA).

8.3.5.1.1. ASA

The ASA is based on static and dynamic translation slots (or TCP/UDP-IP stateful inspection flow) configured in the PIX.

All IP packets incoming on any of the interfaces are checked against the ASA and against connection state information in memory.

The ASA follows a certain set of rules, including the following:

- By default, allow any TCP connections that originate from the higher-security network.
- By default, deny any TCP connections that originate from the lower-security network.
- Ensure that if an FTP data connection is initiated to a translation slot, there is already an FTP control connection between that translation slot and the remote host. If not, drop and log the attempt to initiate an FTP data connection. For valid connections, the firewall handles passive and normal FTP transparently without the need to configure your network differently.
- Drop and log attempts to initiate TCP connections to a translation slot from the outside.
- Drop and log source-routed IP packets sent to any translation slot on the PIX Firewall.
- Silently drop ping requests to dynamic translation slots.
- Answer (by the PIX Firewall) ping requests directed to static translation slots.

It is clear that devices using the ASA offer a more secure environment than devices implementing only the stateless and packet filtering technology. This explains the popularity of the PIX in the industry.

8.3.5.1.2. Data Flow for the PIX

The ASA uses the configured security levels at each interface to either permit or deny data flow from one interface to the other. The security levels are numeric values ranging from 0 to 100. Figure 8.6 shows the different security levels.

Figure 8.6: Security Levels

In Figure 8.6, the outside interface has security level 0 and is the least secure. The inside interface has security level 100 and is the most secure. The DMZ interface can be configured with varying security levels. This becomes complex for devices with multiple interfaces. By default, traffic can flow from high-security-level interfaces to low-security-level interfaces. All other traffic flows that are required must be configured. A distinction needs to be made between inbound and outbound traffic.

Imagine that an outbound packet (going from the inside network to the outside world) arrives at the PIX Firewall's inside interface. (PIX Firewalls name interfaces by default as inside and outside; another common interface name is DMZ.) The ASA verifies whether the traffic is permitted. The PIX Firewall checks to see if previous packets have come from the inside host. If not, the PIX Firewall creates a translation slot (also called an xlate) in its state table for the new connection. The translation slot includes the inside IP address and a globally unique IP address assigned by network address translation (NAT). A PIX can perform NAT and often does. However, it is also possible to

perform NAT on a different device, such as a packet filtering router placed between the PIX and the inside network (Belt and Braces Firewall architecture). It is also possible to use a registered address inside and not translate at all.

The PIX Firewall then changes the packet's source IP address to the globally unique address (unless your network is set up to use a fully public routable address space). The firewall then modifies the checksum and other fields as required and forwards the packet to the appropriate outside interface.

When an inbound packet arrives at the outside interface, it must first pass the PIX Firewall Adaptive Security criteria before any translation occurs. If the packet passes the security tests, the PIX Firewall removes the destination IP address, and the internal IP address is inserted in its place. The packet is forwarded to the inside interface. If there are no matching criteria found by the ASA, the packet is dropped and the threat is removed.

Figure 8.7 displays a typical network with PIX located between an internal and external network.

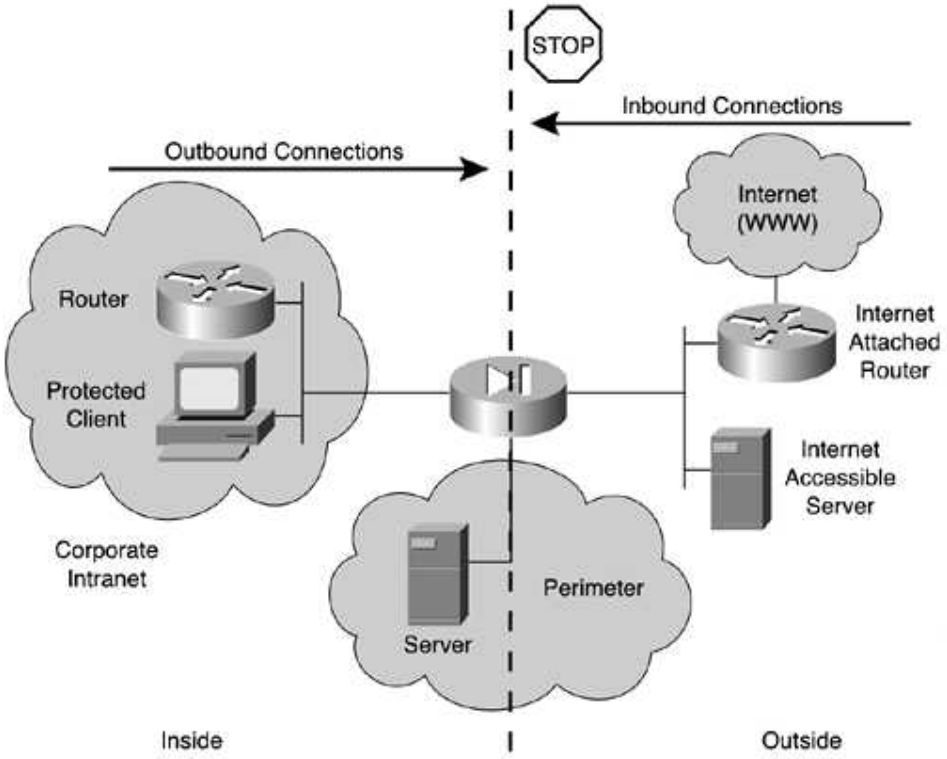


Figure 8.7: PIX Placement

Figure 8.7 shows a typical network design in which the internal network is protected from devices on the Internet, and only connections made from internal hosts are permitted to the outside (or to the Internet). You can, however, permit outside hosts to connect to resources internally by using access lists (in the older software versions of PIX, these were called conduits). A conduit or PIX access list is basically a rule that breaks the default behavior of the PIX (or the ASA) by permitting connections to internal devices located in the inside interface or the perimeter zone. Why would you permit outside untrusted devices access to sensitive hosts? The answer is that basically most companies, including Cisco, permit the following:

- FTP or HTTP to host devices so that orders can be placed
- Download of the latest technology white papers
- Download of the latest patches of Cisco IOS software

As long as you have a sound security policy in place, it provides the network administrator control of security vulnerabilities for hosts and servers with specific access from the outside world. Unfortunately, no one is immune to hackers trying to break into the network or trying to bring down your websites.

Although it is beyond the scope of the book to explore these in detail, the following list presents some additional features and functions of the PIX:

- Authentication based on AAA (RADIUS or TACACS+)
- Authorization based on AAA (RADIUS or TACACS+)
- Content filtering, URL filtering, Java filtering
- Dynamic Host Configuration Protocol (DHCP)
- Routing Information ProtocolRIPv2/Open Shortest Path First (OSPF)
- VPN capability
- Logging
- DC power (security in telephone environments)
- Failover

8.3.5.2.NetScreen Firewall

The NetScreen firewalls are deep inspection firewalls providing application-layer protection, whereas the PIX can be configured as stateful or stateless firewalls providing network- and transport-layer protection. Both NetScreen and PIX Firewalls are certified by the ICSA labs and have Common Criteria EAL 4 ratings.

NetScreen was founded on the vision of providing integrated security technologies that offer wire speed performance and are easy to deploy throughout an enterprise network. Juniper Networks acquired Netscreen in April 2004. Unlike Cisco, which is a networking company that provides hardware and software for nearly any network requirement, NetScreen provides network security products only.

NetScreen firewalls are bundled with Ethernet only. There is no support for Token Ring or high speed ISDN, for example; you need a routing device to perform these types of connections. There is, however, a gigabit-enabled firewall solution allowing, for example, a 1 Gb connection to a local-area network (LAN) infrastructure to enable fast processing per port. This operates much as a switch does for users on a large TCP/IP network.

The NetScreen firewall is a deep packet layer, stateful inspection device. It bases all its verification and decision making on a number of different parameters, including source address, destination address, source port, and destination port. The data is checked for protocol conformities.

NetScreen's Deep Inspection firewall is designed to provide application-layer protection for the most prevalent Internet-facing protocols such as HTTP, DNS, and FTP. The Deep Inspection firewall interprets application data streams in the form that a remote device would act upon. Deep Inspection firewalls defragment and reassemble packets and ensure that all data is reorganized into the original state.

Once the Deep Inspection firewall has reconstructed the network traffic, it employs protocol conformance verification and service-field attack pattern matching to protect against attacks within that traffic. These features are all controlled and acted upon by hardware-based ASIC chips to increase performance.

It is important to understand the dataflow for NetScreen firewalls. Except with low-end firewalls, by default, all NetScreen firewalls deny all traffic from any given interface. NetScreen's terminology for inside and external interfaces is user configurable. For example, the interfaces are called trusted interface and untrusted interface or the red zone and blue zone. A zone is merely a collection of physical or logical interfaces. Once the interfaces are placed in user-defined zones (UDZs), policies dictate what traffic is permitted or denied between the defined zones, as per Cisco access-list architecture. As soon as a policy match is made, the packet is sent to the appropriate queue. If no match is made, the packet is thrown into the bit bucket.

NetScreen devices maintain a session table that outlines, among other things, the source, the destination, the source port, and the destination port, and the number of active sessions. Figure 8.8 displays a typical session table entry on the NetScreen firewall and the detailed explanations of each field.

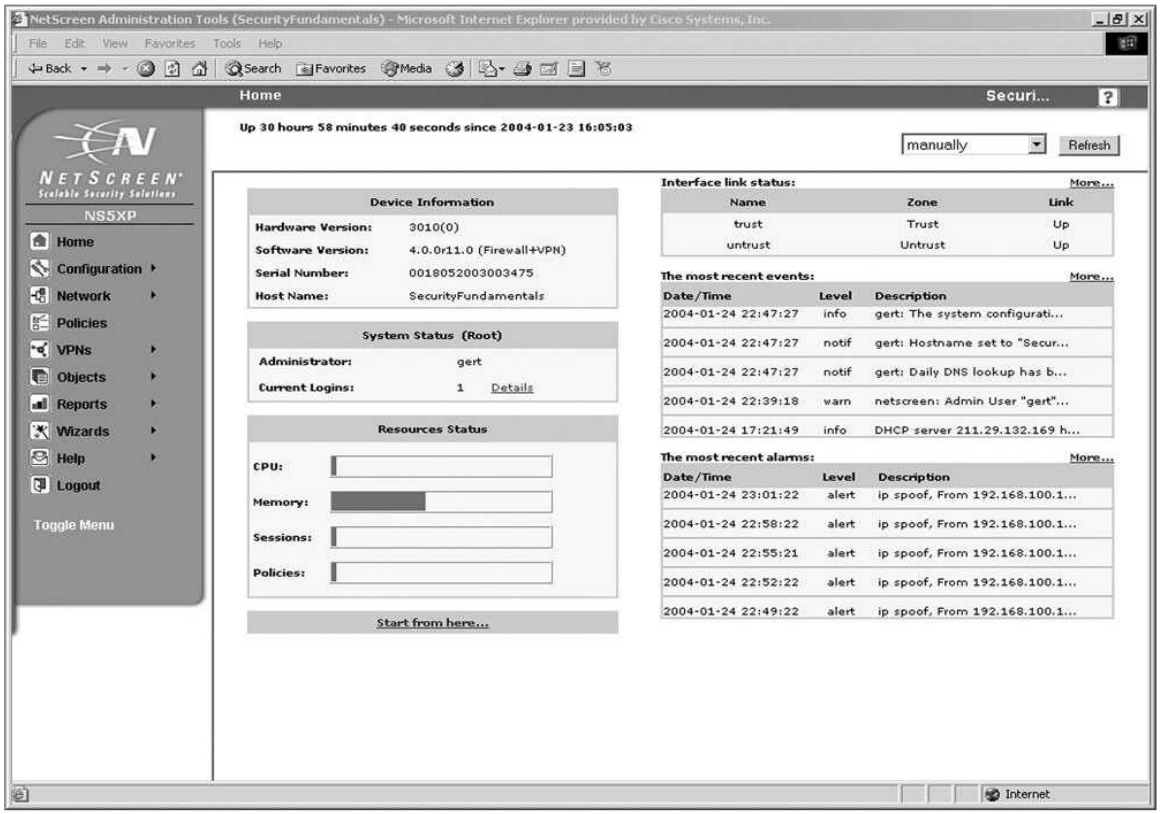


Figure 8.8: NetScreen Firewall Session Information

Additionally, a NetScreen firewall can operate at Layer 2 or Layer 3 mode. This allows a NetScreen firewall to be placed at the edge of the network with no IP address space required, except one address for management. This can be a significant advantage in large IP address networks when there may be a need to readdress IP address space when a firewall is strategically placed. Figure 8.9 illustrates this firewall placement.

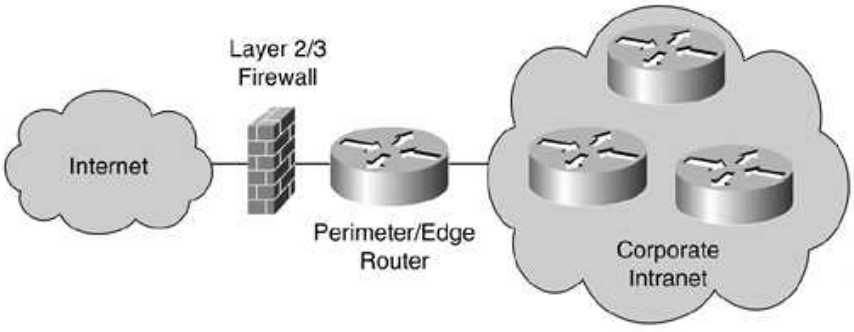


Figure 8.9: NetScreen Firewall Placement

Additionally, the NetScreen firewall can perform the following functions:

- Support for NAT and policy-based NAT
- Support for Port Address Translation (PAT)
- Ability to support inbound connections to hosts such as FTP servers
- Support for VPN

- DHCP
- URL filtering
- Management via a simple web HTTP interface
- Support for routing protocols such as BGP (only 8000 entries), OSPF, and RipV2

8.3.6. Check Point Software Firewalls

As most, hardware firewalls provide effective access control, many are not designed to detect and thwart attacks specifically targeted at the application level. Tackling these types of attacks is most effective with software firewalls.

Check Point is a major vendor in the software firewall marketplace today. Software firewalls allow networks and, more specifically, network applications to be protected from untrusted sources such as the Internet. The fact that millions, if not billions, of devices such as PCs, PDAs, and IP phones have instant access to the entire Internet means that commercial enterprises and networks based on country controls are vulnerable to attacks. The relative openness of the web has made it possible for anyone to potentially access a private network. Securing the network perimeter is the core foundation of the Check Point solution.

The Check Point Enterprise suite is an integrated product line that ties together network security, quality of service, and network management for large IP networks.

In short, Check Point can provide the following services:

- Firewall services
- VPN
- Account management
- Real-time monitoring
- Secure updates over the Internet
- User-friendly management interface

A Check Point firewall is a software solution and is hardware independent. The firewall software can be installed on a variety of different platforms, including the following:

- Windows 2000
- Solaris based on UNIX
- Red Hat Linux

Windows XP has a very basic firewall built into the client adapters that restricts ICMP traffic. ZoneAlarm and Sygate personal firewalls allow the PC user to permit or deny IP-based traffic to and from the client device, such as a PC. For example, a HTTP session initiated to the Internet triggers the personal firewall to prompt the user on whether to forever allow, deny, or block the request. Of course, it still requires an intelligent user and hence is not as popular as the hardware-based solution this chapter has introduced. These software applications basically allow users to be prompted or notified by alarm when remote devices initiate connections that are supposed to be blocked.

8.4. Enhancements for Firewalls

Of the many enhancements to firewalls, this section concentrates on four of the most important feature enhancements present in today's firewalls, namely:

- NAT
- Proxy services
- Content filtering
- Antivirus software

8.4.1. NAT

NAT is a router or firewall function whose main objective is to translate the addresses of hosts behind a firewall or router. NAT can also be used to overcome the IP address shortage that users currently experience with IPv4.

NAT is typically used for internal IP networks that have unregistered (not globally unique) IP

addresses. NAT translates these unregistered addresses into the legal addresses of the outside (public) network. This allows unregistered IP address space connectivity to the web and also provides added security.

Cisco IOS 12.0 and higher support full NAT functionality in all images. Cisco IOS 11.2 and higher need the "PLUS" image set for NAT feature support. (Cisco extended NAT with port address capabilities to increase the utility of each outside address. This is called Port Address Translation [PAT] in the Cisco terminology.)

PAT provides additional address expansion but is less flexible than NAT. With PAT, one IP address can be used for up to 64,000 hosts by mapping several IP port numbers to one IP address. PAT is secure because the source IP address of the inside hosts is hidden from the outside world. The perimeter router typically provides the function of NAT or PAT.

Figure 8.10 displays a typical scenario in which a private address space is deployed that requires Internet access. The private subnetted Class A 10.10.10.0/24 is not routable in the Internet.

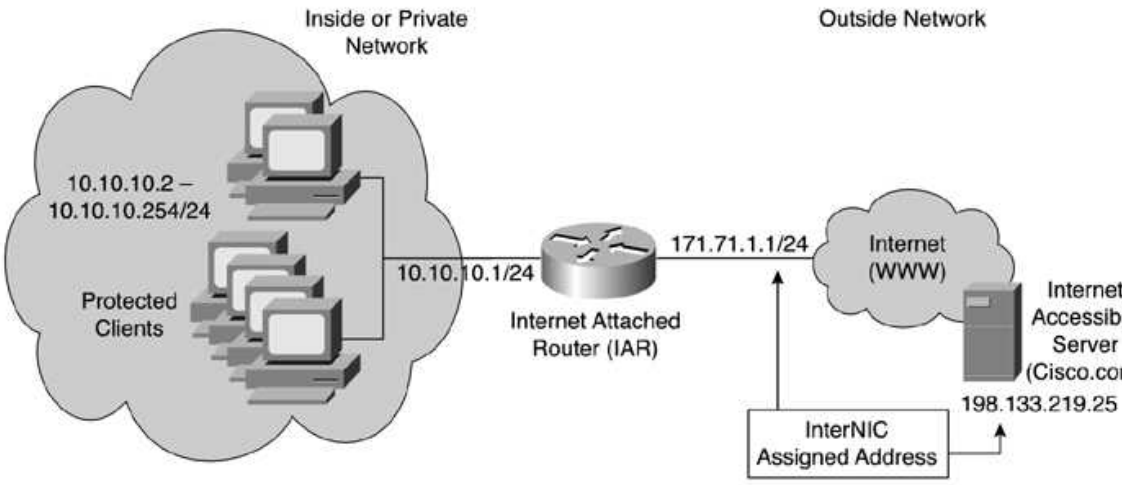


Figure 8.10: Typical PAT Scenario

The users in Figure 8.10 are configured with an inside local address ranging from 10.10.10.2/24 to 10.10.10.254/24. To allow Internet access, NAT is configured on Router IAR to permit the inside local addresses access to the Internet. (In this case, only PAT is configured because only one IP address was allocated by InterNIC, namely 171.71.1.1.) The advantages of using NAT include

- Hiding the Class A address space 10.10.10.0/24
- Internet access provided to all protected users without IP address changes

To view the NAT translation table on a Cisco router, apply the exec command show ip nat translations on the CLI interface. Example 9-1 illustrates the show ip nat translation configuration command on the Internet Accessible Router (IAR).

Before examining a demonstration of the configuration on the router and PIX Firewall, you need to become familiar with the NAT environment terminology set out in Table 8.1.

Term Meaning

Inside local ddress An IP address that is assigned to a host on the internal network, which is the logical address that is not being advertised to the Internet. This is an address that is generally assigned by a local administrator. This address is not a legitimate Internet address.

Inside global address A legitimate registered IP address as assigned by the InterNIC.

Outside local address The IP address of an outside host of the network that is being translated as it appears to the inside network.

Outside global address The IP address assigned to a host on the outside of the network that is being translated by the host's owner.

The disadvantages of NAT/PAT include the following:

- They are CPU processing power intensive.
- The Layer 3 header and source address changes.
- Voice over IP is not yet supported.

Some multimedia-intensive applications do not support NAT, especially when the data stream inbound is different from the outbound path, for example, in multicast environments.

8.4.2. Proxy Services

The use of proxy services in the network has multiple goals. Proxy services can be used to hide the real IP address of users. This means that when crackers or intruders try to spoof IP addresses, for example, they have no idea about the hidden addresses and in fact attack a proxy server designed to drop the packets and alert network administrators of the event.

There are even websites dedicated to home users and corporate users that offer proxy-like services.

Today's firewalls can act as proxy servers on behalf of clients such as UNIX hosts, Windows users, or HTTP servers.

Proxy servers can also cache information that is frequently used by end users and thus can act as an intermediate device between a web client and a web server. This allows other web clients to access web content much faster by downloading web content from a local device rather than from the web (proxies protect clients and reverse proxies protect servers).

8.4.3. Content Filters

With content filtering (also known as URL filtering), an organization designs a policy defining which websites are permitted to be accessed by local resources and which are not. Content filters can monitor, manage, and provide restricted access to the Internet. This means that employees do not tie up valuable and expensive WAN connections to the Internet for nonbusiness matters. You might, for example, allow access to www.cisco.com but deny employees access to music websites that permit large downloads of sheet music or MP3 files.

Cisco provides a number of content-filtering engines that can perform the following functions:

- Deny access to URLs specified in a list
- Permit access only to URLs specified in a list
- Use an authentication server in conjunction with a URL filtering scheme

The scenario illustrated in Figure 8.11 briefly touches on this concept. User1 with the IP address 10.10.10.1 is granted full access to all Internet resources, whereas User2, who is a temporarily employee with the IP address 10.10.10.2, has access only to the Cisco website and the Cisco Press website.

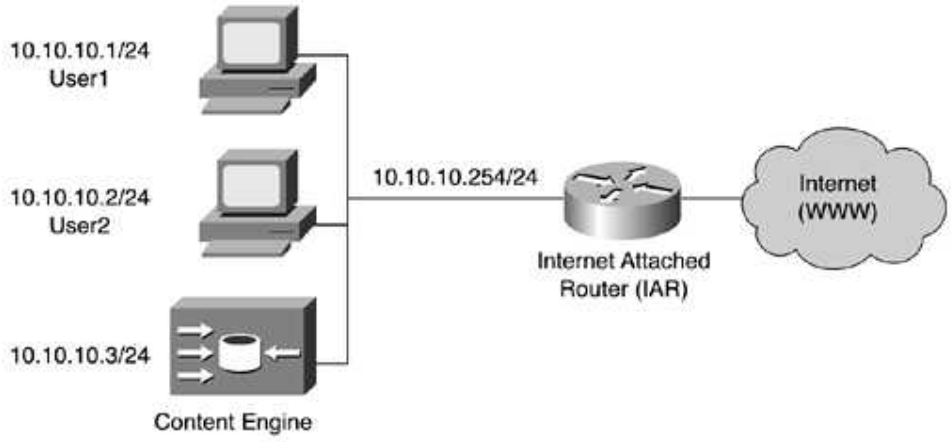


Figure 8.11. Typical Content Filtering Scenario

8.4.4. Antivirus Software

A computer virus can best be described as a small program or piece of code that penetrates into the operating system, causing an unexpected and usually negative event. Antivirus software applications scan the memory and hard disks of hosts for known viruses. If the application finds a virus (using a reference database with virus definitions), it informs the user. The user can decide what needs to happen next. These types of applications are becoming integrated features of newer software firewalls.

8.5. Summary

This chapter detailed a complex example and should provide the reader with the right tools and knowledge to tackle any PIX scenario. As you have seen, the PIX command set is rather easy, and once you understand what a command does and how to use it, the configuration of the firewall is not a difficult task at all.

Chapter 09
Intrusion Detection System

9.1. Introduction to Intrusion Detection

It is becoming increasingly important for network security personnel to defend company resources, not only passively by using firewalls, virtual private networks (VPNs), encryption techniques, and whatever other tricks they have up their sleeves, but also by deploying proactive tools and devices throughout the network. This is where IDSs come in.

In general, intrusion is when someone tries to break into, misuse, or exploit your system. More specifically, your organization's security policy defines what constitutes attempts to break into, abuse, or exploit your system. The security policy also defines the perpetrator of those attempts or actions.

There are two types of potential intruders exist:

- Outside intruders
- Inside intruders

Although the majority of intrusion attempts actually occur from within the organization or by inside intruders, the most common security measures that are put in place protect the inside network from the outside world. Outside intruders are often referred to as crackers.

It's clear that a mechanism is desirable and required to detect both types of intrusions continuously. IDSs are effective solutions for both types of attacks. These systems run constantly in your network, notifying network security personnel when they detect an attempt they consider suspicious. IDSs have two main components, namely, IDS sensors and IDS management.

IDS sensors can be software and hardware based used to collect and analyze the network traffic. These sensors are available in two varieties, network IDS and host IDS.

A host IDS is a server-specific agent running on a server with a minimum of overhead to monitor the operating system.

A network IDS can be embedded in a networking device, a standalone appliance, or a module monitoring the network traffic.

IDS management, on the other hand, acts as the collection point for alerts and performs configuration and deployment services for the IDS sensors in the network.

9.2. IDS Fundamentals

A solid understanding of the fundamentals and different IDS technologies is required before the actual analysis and deployment discussions can start.

9.2.1. Notification Alarms

The overall purpose of IDSs is to trigger alarms when a given packet or sequence of packets seems to represent suspicious activity that violates the defined network security policy. Although alarms are essential, it is critical for network security personnel to configure the IDS to minimize the occurrence of false negative and false positive alarms.

A **false positive** is a condition in which valid traffic or a benign action causes the signature to fire. A **false negative** is a condition in which a signature is not fired when offending traffic is transmitted. False negative alarms occur when the IDS sensor does not detect and report a malicious activity, and the system allows it to pass as nonintrusive behavior. This can be catastrophic for network operation. Therefore, minimizing false negatives has the highest priority. In general, there are two main reasons for a false negative to occur:

- The first results from the sensor lacking the latest signatures.
- The second can occur because of a software defect in the sensor.

The IDS configuration should be continuously updated with new exploits and hacking techniques upon their discovery.

False positive alarms occur when the IDS sensor classifies an action or transaction as anomalous (a possible intrusion) although it is actually legitimate traffic. A false alarm requires an

unnecessary intervention to analyze and diagnose the event. Clearly, network administrators try to avoid this type of situation because a large number of false positives can significantly drain resources, and the specialized skills required for analysis are scarce and costly.

As a central warehouse of security knowledge, Cisco has developed an encyclopedia to provide security professionals with an interactive database of security vulnerability information.

As stated previously, the process of updating the IDS configuration is a continuous activity because it is virtually impossible to completely eliminate false positives and false negatives. For instance, if new applications are deployed throughout your organization, retuning the sensors might be required to minimize false positives. Most sensors provide flexible tuning capability during steady state operations, so there is no need to take them off-line at any point.

9.2.2. Signature-Based IDS

The signature-based IDS monitors the network traffic or observes the system and sends an alarm if a known malicious event is happening. It does so by comparing the data flow against a database of known attack patterns. These signatures explicitly define what traffic or activity should be considered as malicious. Various types of signature-based IDSs exist, including the following:

- Simple and stateful pattern matching
- Protocol decode-based analysis
- Heuristic-based analysis

The pattern-matching systems look for a fixed sequence of bytes in a single packet, which has three advantages: It is simple, it generates reliable alerts, and it is applicable to all protocols. The weakness of pattern-matching systems is that any slightly modified attack leads to false negatives. Multiple signatures may be required to deal with a single vulnerability in stateful pattern-matching systems because matches are made in context within the state of the stream.

Protocol decode-based systems decode very specific protocol elements, such as header and payload size and field content and size, and analyze for Request for Comment (RFC) violations. These systems have the advantage of being highly specific and, as a result, minimize the chance for false positives.

Table 9.1 gives a general overview of the pros and cons of signature-based IDSs.

Table 9.1: Pros and cons of signature based IDS

Pros	Cons
Low false positive rate (reliable alerts)	Single vulnerability may require multiple signatures
Simple to customize	Continuous updates required
Applicable for all protocols	Modifications lead to misses (false negatives)
	Cannot detect unknown attacks
	Susceptible to evasion

The following example is an attack against a web server of Company X, in which the attacker is trying to find the passwords of known users in a file containing encrypted passwords for the systemthe /etc/shadow file. Commonly, web server attacks are specially crafted URLs that start with an HTTP request from the attacker. To detect these types of attacks, the IDS looks for the signature in the beginning of the dataflow when parsing all the incoming bytes. Figure 9.1 illustrates this attack, which can be prevented using a signature-based host IDS.

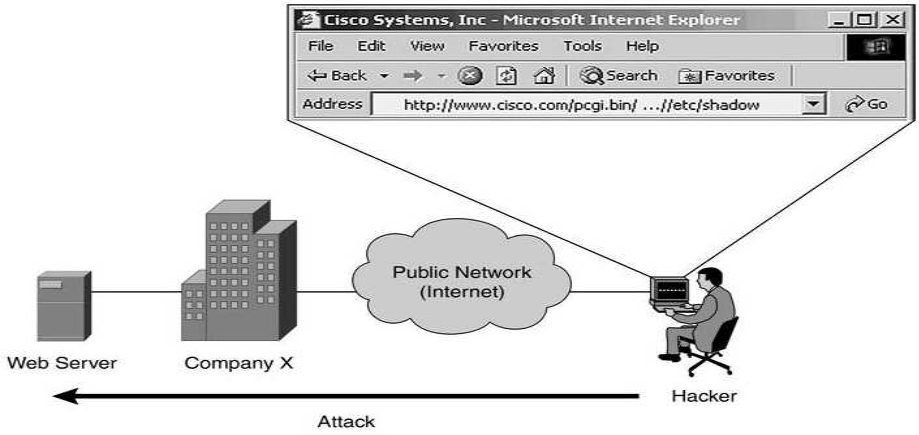


Figure 9.1: Attack That Can Be Prevented Using Signature-Based IDS

The Cisco Network Intrusion Detection Sensors keep complete collections of known malicious events in a database called the Network Security Database (NSDB).

The NSDB is an HTML-based encyclopedia of network vulnerability information. Figure 9.2 displays the Network Security Vulnerability Index. Figure 9.3 is a typical example of an exploit signature and how it is formatted in the database.

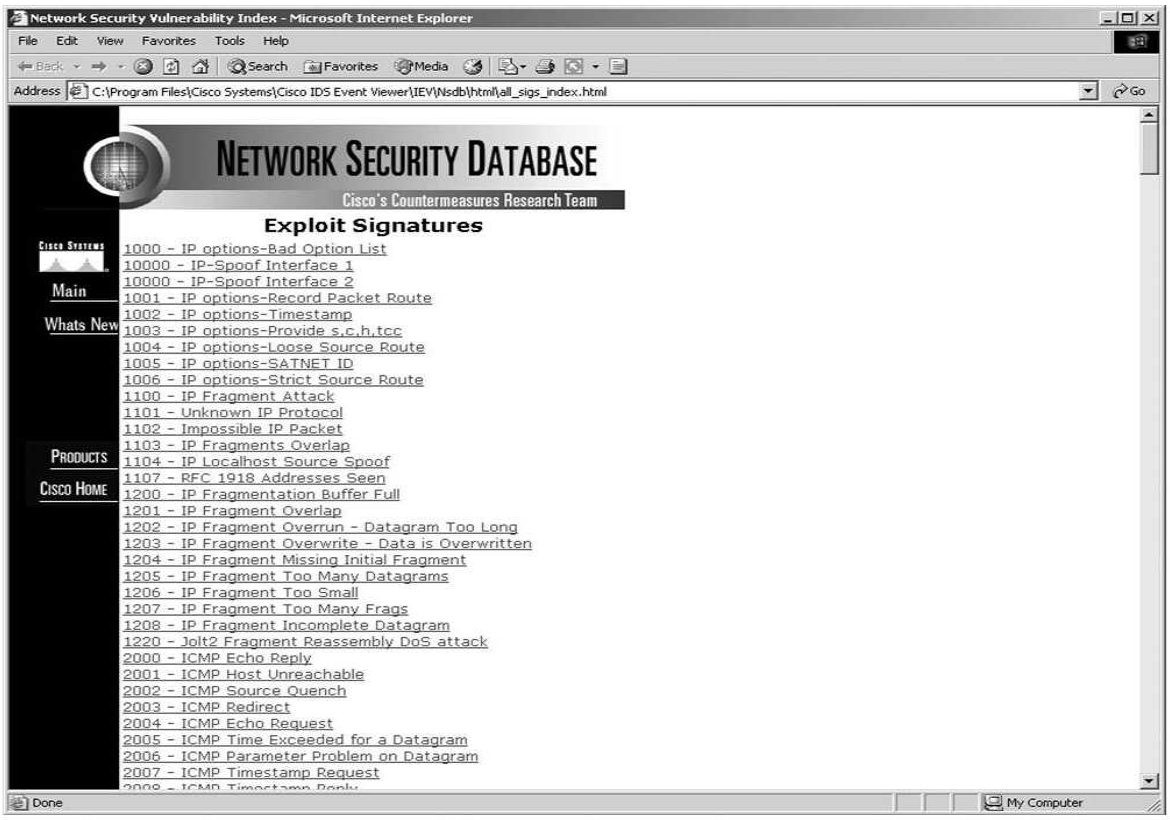


Figure 9.2: Network Security Database

A Smurf attack, which is named after the program used to perform the attack, is a denial-of-service (DoS) attack. It is a method by which an attacker can send a moderate amount of traffic and cause a virtual explosion of traffic at the intended target.

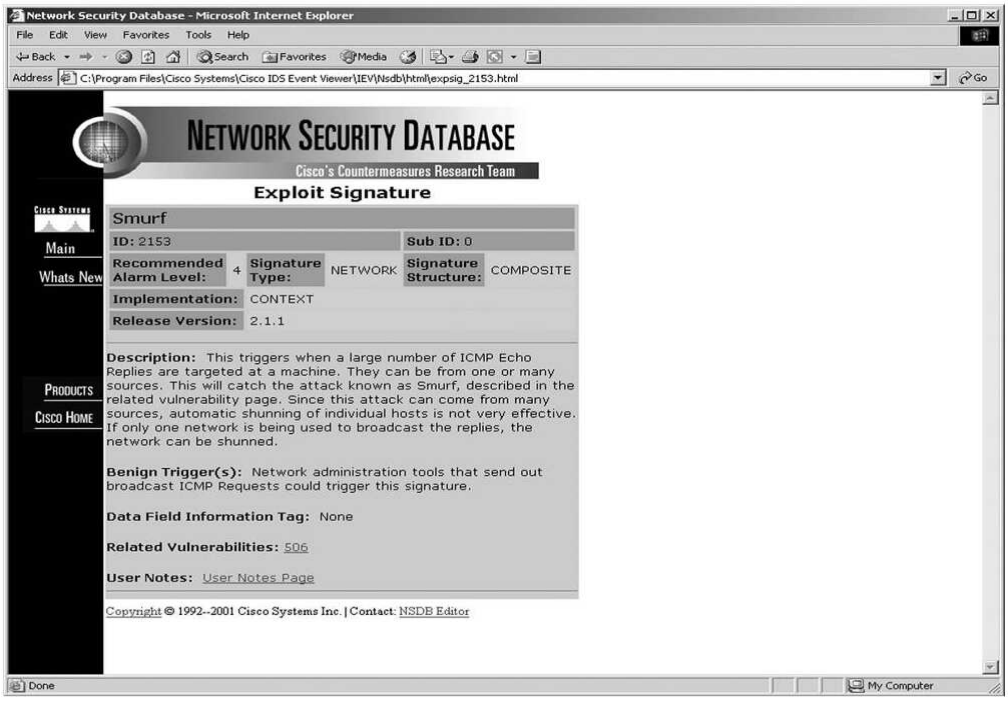


Figure 9.3: A Smurf Attack Signature (Name, Signature ID, and Description)

9.2.3. Policy-Based IDS

The policy-based IDSs (mainly host IDSs) trigger an alarm whenever a violation occurs against the configured policy. This configured policy is or should be a representation of the security policies. For instance, a network access policy defined in terms of access permissions is easy to implement. The marketing department on network x is allowed to browse only engineering websites and has no access to FTP software directories on segment y. This is a fairly simple example of network policy; other policies are much harder to implement. If, for instance, a company's management team does not allow the browsing of game sites, the IDS must be able to communicate with a database of blacklisted sites to check whether a policy violation has occurred.

Figure 9.4 illustrates this violation, which can be prevented by using a policy-based IDS. Employees from the engineering department should not be able to access either the marketing department VLAN or its servers.

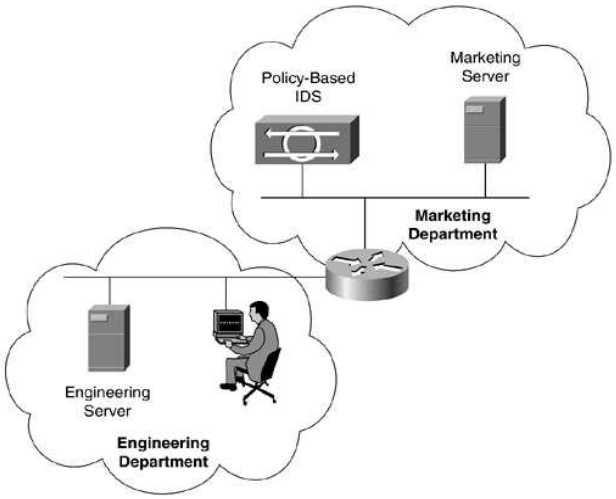


Figure 9.4: Attack That Can Be Prevented Using Policy-Based IDS

Table 9.2 gives a general overview of the pros and cons of policy-based IDS.

Table 9.2: Overview of Policy-Based IDS Pros Cons

Pros	Cons
Low false positive rate (reliable alerts)	Network administrator must design a set of policy rules from scratch
Simple to customize	Long deployment time

This type of IDS is flexible and can be customized to a company's network requirements because it knows exactly what is permitted and what is not. On the other hand, the signature-based systems rely on vendor specifics and default settings.

9.2.4. Anomaly-Based IDS

The anomaly-based IDS looks for traffic that deviates from the normal, but the definition of what is a normal network traffic pattern is the tricky part. Once the definition is in place, the anomaly-based IDS can monitor the system or network and trigger an alarm if an event outside known normal behavior is detected. An example of abnormal behavior is the detection of specific data packets (routing updates) that originate from a user device rather than from a network router. This technique is known in the world of crackers as spoofing. Table 9.3 gives a general overview of the pros and cons of anomaly-based IDS.

Table 9.3: Overview of Anomaly-Based IDS Pros Cons

Pros	Cons
Unknown attack detection	High false positive rate
Easy deployment for networks with well-defined traffic patterns	Interpretation of generated alarms is difficult

Easy deployment for networks with well-defined traffic patterns Interpretation of generated alarms is difficult

Two types of anomaly-based IDS exist: statistical and non-statistical anomaly detection. Statistical anomaly detection learns the traffic patterns interactively over a period of time. In the non-statistical approach, the IDS has a predefined configuration of the supposedly acceptable and valid traffic patterns.

9.2.5. Network IDS versus Host IDS

The previous sections outlined different analysis technologies. A good IDS has to be built around a solid implementation of these various technologies. Host IDSs and network IDSs are currently the most popular approaches to implement analysis technologies. A host IDS can be described as a distributed agent residing on each server of the network that needs protection. These distributed agents are tied very closely to the underlying operating.

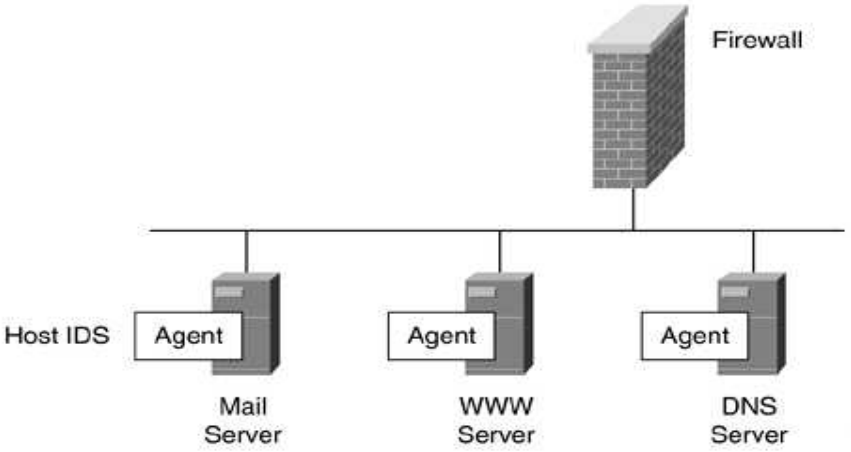


Figure 9.5: Host IDS

Network IDSs, on the other hand, can be described as intelligent sniffing devices. Data (raw packets) is captured from the network by a network IDS, whereas host IDSs capture the data from the host on which they are installed. This raw data can then be compared against well-known attacks and attack patterns that are used for packet and protocol validation. In addition to application validation, the network IDS is capable of keeping track of connection and flow status. Figure 9.6 illustrates the placement of a network IDS on a network segment.

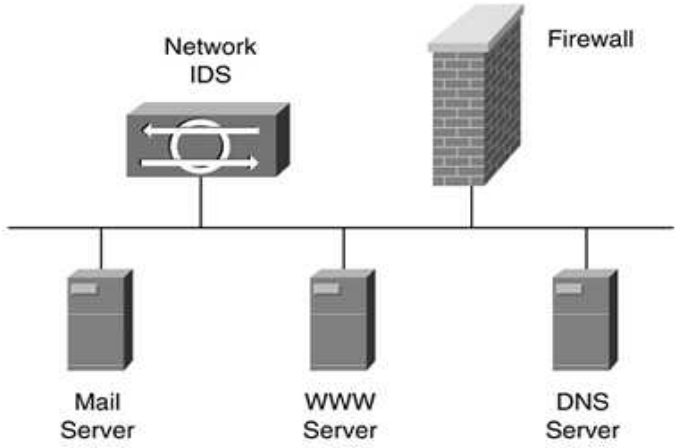


Figure 9.6: Network IDS

Host IDS and network IDS should be seen as complementary because the systems fill in each other's weaknesses. Table 9.4 lists the most important pros and cons of these systems.

Table 9.4: Comparison of Host IDS and Network IDS IDS Type

IDS Type	Pros	Cons
Host IDS	Verification of success or failure of an attack possible. Has a good knowledge of the host's context and, as a result, is more focused on a specific system. Not limited by bandwidth restrictions or data encryption.	Operating system/platform dependent. Not available for all operating systems. Impact on the available resources of the host system. Expensive to deploy one agent per host.
Network IDS	Protects all hosts on the monitored networkcost effective. Independent of the operating system and has no impact on the host (runs invisibly). Especially useful for low-level attacks (network probes and DoS attacks).	Deployment is very challenging in switched environment. Network traffic may overload the NIDS (CPU intensive). Not effective for single packet attacks, and hidden attacks in encrypted packets.

Generally speaking, the most efficient approach is to implement network-based IDS first. It is much easier to scale and provides a broad coverage of the network. Furthermore, less organizational coordination is required, with no or reduced host and network impact. If only a few servers need to be protected, a network administrator may want to start with host-based IDS.

9.2.6. Evasion and Antievasion Techniques

Network IDSs have a fundamental problem whereby a skilled attacker can evade the detection mechanism by exploiting ambiguities in the traffic patterns, network topology, and the IDS architecture. Network IDS evasion enables the attacker to use techniques that challenge the detection mechanisms and therefore allow certain attacks to pass unnoticed.

If the attacker suspects that a network IDS may be monitoring the network, he may start using alternative techniques to try and avoid detection. The attacker can try to evade the detection mechanism in the sensor. The attacker can try to convince the network IDS by masking the traffic as legitimate. The attacker can also try to generate lots of false positives to overwhelm the operator and the sensor hardware that is monitoring the logs and events. In this way, real threats to the network are not visible because the IDS is unable to capture and analyze all the traffic. Examples of these common evasion techniques are flooding, fragmentation, and obfuscation, as explained in Chapter 2.

As you can imagine, most vendors are aware of these evasion techniques and combat them by using antievasion countermeasures. Antievasion techniques can range from fragmentation alarms, packet loss alarms, and protocol decodes to tunable TCP stream reassembly options, alarm summarization, and others.

9.2.7. Organizational Issues and Complications

Intrusion detection spans many business functions within an organization. Organizational issues and complications are a direct result of the required interaction between the different groups.

Similar to designing a completely new network, the design, integration, and maintenance of IDSs in your network is an exercise in meeting strict requirements while simultaneously working with certain constraints. As discussed in Chapter 6, "Secure Design," these constraints can be markedly different in nature and can include technological constraints, social constraints, and political constraints.

9.2.7.1.Technological Constraints

The changing needs of consumers and society in general are obvious. All these developments cause Internet traffic to double every few months, whereas CPU processing speed is only doubling about every year to year-and-a-half. Because of the far more rapid increase of Internet traffic levels, computation is still a constraint for network designers, particularly in the case of routers and switches. Typically, the computation (processing) limitations that apply to network design are associated with the processing of the routing table calculations, encryption and decryption of secured packets, accounting, incoming and outgoing access lists, or even normal packet forwarding. The processing of network traffic from IDSs may overload the sensor or appliance (such processing is CPU intensive) because it sniffs all packets being sent on a specific segment.

Technological issues also include the bandwidth of the interfaces, tap placement, and switch configuration.

9.2.7.2.Social Constraints

Manpower or labor in general is clearly a concern in any network design. The more often a task must be executed, the more the design should focus on making that particular task simple and efficient to manage. Considering that 24 hours a day, 7 days a week, 365 days a year (24 x 7 x 365) monitoring and response capabilities are required for a proper IDS, a good IDS management design reduces labor costs. Network security personnel in charge of the IDSs require a cross-functional skill set, ranging from networking and security to operating systems. Staffing and personnel training should be considered as a top priority when designing an IDS for your network.

Some larger enterprises can consider outsourcing their IDS management so that internal resources can be employed elsewhere. But when you consider the complexity of tuning the IDS according to the security policy, service-level agreements are not easy to negotiate.

9.2.7.3. Political Constraints

A company should have an incident response policy and procedure in place that has been approved by the senior management team. This policy includes recovery procedures in case of a severe attack. In addition, the following should be absolutely clear to the network administrator: the circumstances that require senior management notification and the stage at which the company's legal department calls for law enforcement.

Organizational politics can become involved in the compulsory use of standards and legacy applications that are difficult to understand, implement, and use. Some companies have a single-vendor prearranged partnership agreement, whereas other leadership teams require a multivendor type of environment.

9.3. Host-Based IDSs

By now, all network administrators are aware that network security should be seen as a continuous process built around the security policy. This process is a four-step method, as described in Chapter 5: Secure the system, monitor the network, test the effectiveness of the solution, and improve the security implementation. Testing the effectiveness of the IDS host sensor is an integral part of the monitoring step.

A host IDS can be described as a distributed agent residing on each server of the network that monitors the network activity in real time. The host IDS detects the security violations and can be configured so that an automatic response prevents the attack from causing any damage before it hits the system. The section that follows focuses on the Cisco Secure Agent.

9.3.1. Host Sensor Components and Architecture

The Cisco Intrusion Detection Host sensor has two main components:

- Cisco Secure Agent
- Cisco Secure Agent Manager

9.3.1.1. Cisco Secure Agent

The Cisco Secure Agent is a software package that runs on each individual server or workstation to protect these hosts against attacks.

The Cisco IDS sensor (based on Entercept Security technology) provides real-time analysis and reaction to intrusion attempts. The host sensor processes and analyzes each and every request to the operating system and application programming interface (API) and proactively protects the host if necessary. The next generation Cisco Secure Agents (based on Okena's technology) extend these capabilities even further by automating the analysis function and creating protective policies for the operating system and applications. These agents control all events on files, network buffers, registry, and COM access. The architecture of the Cisco Secure Agent is the Security Agent's Intercept Correlate Rules Engine (INCORE) architecture.

Host IDSs are nowadays referred to as Host Intrusion Protection Systems (HIPS). Figure 9.7 illustrates the architecture of the Host Sensor Agent based on the Entercept technology.

The Host Sensor Agent is installed next to the operating system. The host sensor software has to run adjacent to the operating system to guarantee protection of the operating system itself. The agent protects the host against attacks launched via the network and also protects against attacks or malicious activity by a user who is logged in to the protected host. The rules engine consists of console, agent, general, operating system, web, and FTP rules. The database contains the security policy parameters, user-defined exceptions, and a list of shielded applications.

Let's assume that an attempt is made to compromise the Internet Information Services (IIS) on a web server. The agent core evaluates the incoming data using the FTP rules, which are stored in the rules engine, and applies the policy and exception parameters. If malicious activity is detected, the appropriate reaction is determined. These actions can range from logging to notifications to SNMP traps.

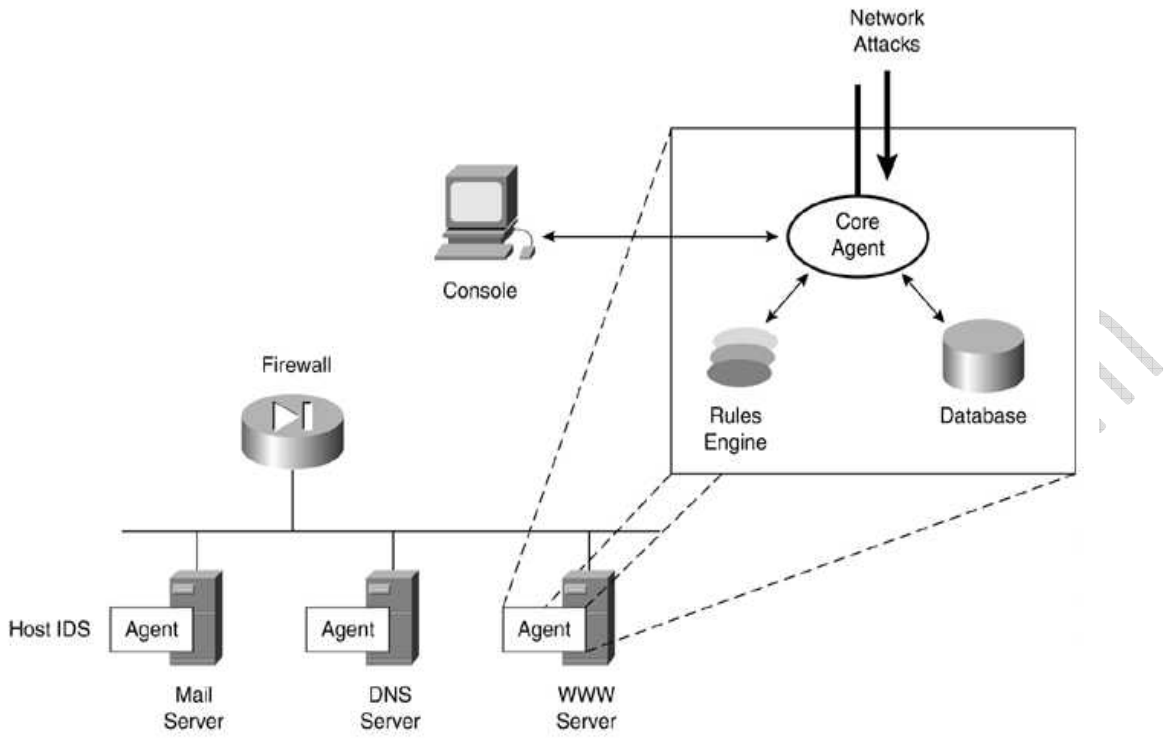


Figure 9.7: Architecture of the Host Sensor Agent

9.3.1.2.Cisco Secure Agent Manager

The Cisco Secure Agent Manager is responsible for managing the Cisco Secure Agent and communication with the agent. The Cisco Secure Agent Manager provides all management functions for all agents in a centralized manner. It also has components that notify security personnel in case of an attack and that generate reports. This management session should use data encryption technologies to be robust, private, and secure. The Cisco Secure Agent Manager has three main components: the graphical user interface (GUI), the server, and the notification handler. Both the GUI and the server are linked to a database where the configuration information is stored.

The agents are directly connected with the server. When an agent sends an alarm to the server, the server is responsible for instructing the notification handler to take care of all configured notification requests such as e-mail and pager notification.

Deploying Host-Based Intrusion Detection in the Network

The deployment of host-based IDSs throughout the organization's network requires a very well-thought-out design. A few design and deployment considerations are discussed in this section, but details on deploying host-based IDSs are far beyond the scope of this book.

Based on what is defined in the organization's security policy, the network designer is responsible for identifying and deciding which systems to protect. A clear objective during the design phase is defining the different system types: Are the servers UNIX or Windows platforms, do you need to protect only servers or should you worry about desktop computers as well as laptops, and so on.

The number of installed Cisco Secure agents is in direct correlation to the number of necessary Cisco Secure Agent Managers. The number of Agents and Agent Managers has a direct impact on personnel, as described in the section "Organizational Issues and Complications" earlier in this chapter.

Figure 9.8 illustrates the host IDS deployment for a company with remote users connecting over a public infrastructure to the corporate network.

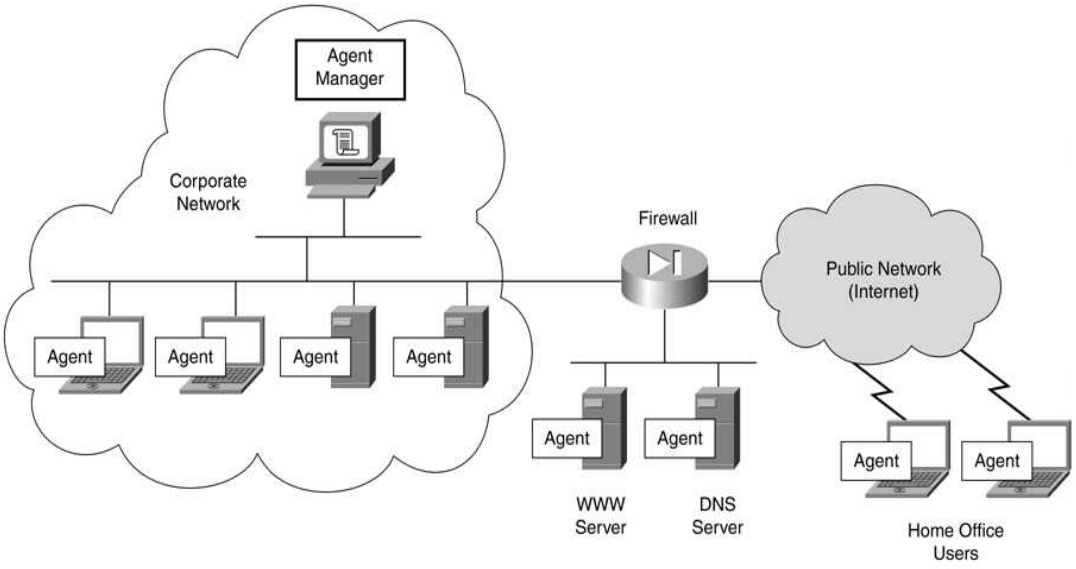


Figure 9.8: Host IDS Deployment

Probably one of the most important considerations in the design phase is the IDS management communication. The agents communicate with the Agent Manager on a specific TCP port. This becomes important when agents are residing on networks other than the Agent Manager network. This is especially true for agents running in a DMZ zone or in a branch or remote home office.

Common strategies for a company's infrastructure are the deployment of web servers, mail servers, Domain Name System (DNS), FTP, and other agents on the DMZ network. Traffic to and from the agents running on these servers to the Agent Manager should be allowed through the firewall.

For remote offices or home offices, VPN and IPSec should be considered when designing the management communication channel between the Agent and the Agent Manager. More details on management communication will follow later in this chapter.

A last criterion to consider when designing your IDS deployment plan is database management. Special attention should go to disk space, disk redundancy, backup scenarios, and so on.

9.4. Network-Based IDSs

Network-based IDSs are an integral part of the monitoring phase of the security policy.

Network-based intrusion detection is the deployment of real-time monitoring probes at vital locations in the network infrastructure. These probes, also called network sensors, analyze the traffic and detect unauthorized activity as well as malicious activity. Depending on the type of offensive strategy an organization has chosen, the probes take appropriate action, as discussed later in this section.

One of the main advantages of deploying network-based systems over host-based systems is the fact that network administrators are able to continually monitor their networks no matter how the networks grow. Adding hosts does not necessarily require the addition of extra network-based intrusion sensors.

9.4.1. Network Sensor Components and Architecture

The network IDS has two interfaces, which are typically connected to different segments of an organization's network. The first one, called the monitoring port, is responsible for capturing data

for analysis. The monitoring port should be connected to the network segment that has potential targets connected, such as mail servers, web servers, and so on. The second port, often referred to as the command and control port, is responsible for sending triggers (alarms) to the management platform. Similar to the host-based Cisco Secure Agent Manager, this platform is used for configuring the network sensors, logging and displaying the alarms, and generating reports on request. Figure 9.9 illustrates the configuration being attacked.

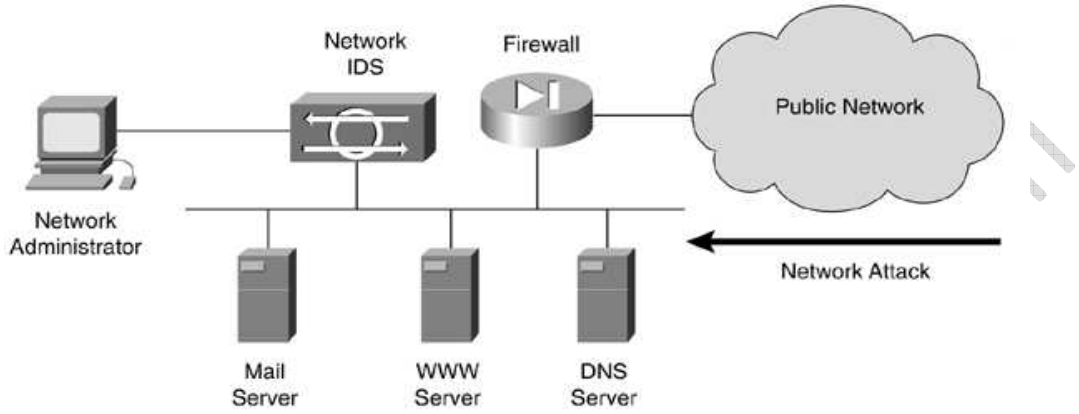


Figure 9.9: Network-Based IDS Overview

The following list outlines the steps involved in the attack and its rebuff:

1. An attack is launched on the mail server via the Internet (public network).
2. Packets travel over the network to the destination, which is the mail server in this case. The data port of the network sensor also captures all these packets.
3. For fragmented packets in different frames, packet reassembly is required. This happens at the packet's final destination (the mail server) and also at the network sensor.
4. The network sensor compares the data against the configured rules set.
5. For all detected attacks, the network sensor generates a log and notifies the network management station.
6. The network management station sends alarms, generates a log, and starts a response action to the attack.

From an architectural viewpoint, the network-based IDSs have three separate components: the network sensor, the director, and the communication mechanism between the previous two. This section focuses on the network sensor architecture. Figure 9.10 illustrates the basic architecture of the IDS sensor.

The network-based IDS sensor runs on Linux and has multiple components (software services), each interconnected and handling different processes. One of the main components is the cidWebServer. The web server uses different servlets to provide IDS services. The cidWebServer communicates with the event server, transaction server, and IP log server servlets using the Remote Data Exchange Protocol (RDEP). RDEP serves as the sensor's communication protocol.

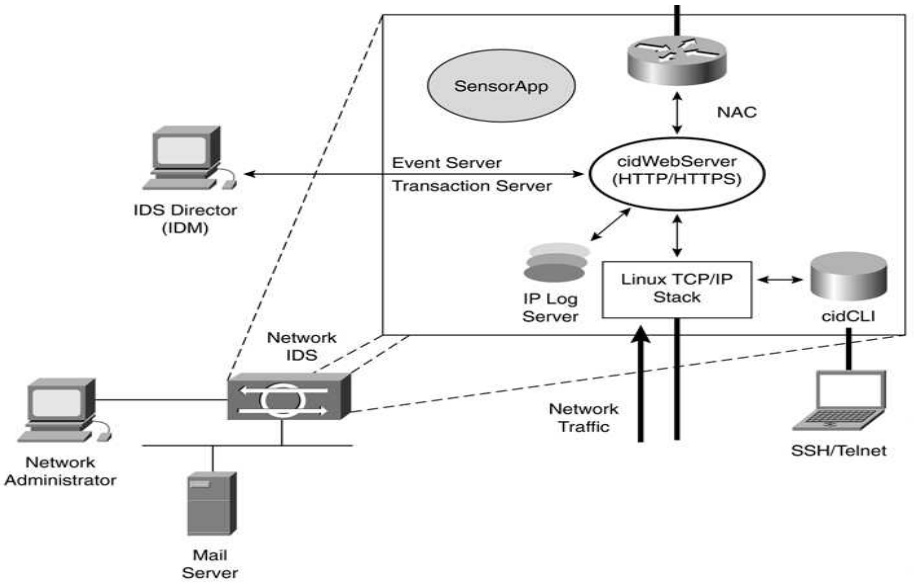


Figure 9.10: Network-Based IDS Architecture

9.4.2. Deploying Network-Based Intrusion Detection in the Network

Network IDSs are developed so that when deployment is carefully planned at designated network points, the network administrator or security personnel can monitor the data (network activity). When the monitoring takes place, the data is traveling only on the network. Therefore, the administrator has the opportunity to take proper action without needing to know what the exact target of the attack is because the IDS monitors the complete segment.

A number of steps or tasks need to be considered when deploying network sensors in your network. Installing the network sensors requires some planning before actually starting to connect the sensors to the network. It is the task of the security network administrators to determine what traffic needs to be monitored to protect all critical assets of the organization.

When planning for sensor placement, a network administrator must consider the size and complexity of the network, interconnectivity with other networks, and the amount and type of network traffic. After collecting this information and also knowing what information requires protection, the sensor location and sensor type (based on bandwidth) can be defined.

Sensors placed on the inside network have different duties than sensors placed on the outside network. Figure 9.11 illustrates the network sensor placement using a scenario that includes a number of attacks on a web server connected on a DMZ.

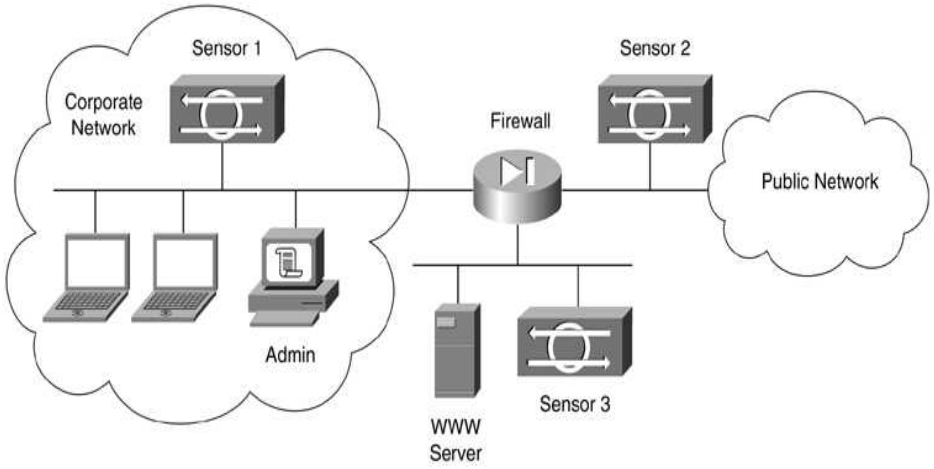


Figure 9.11: Network-Based IDS Sensor Placement

`

Sensor 1, connected on the inside network, sees only traffic that is permitted by the firewall or internal traffic that does not traverse the firewall. All intrusions reported by Sensor 1 require immediate attention and response from the network administrator. Protecting all internal connections on the firewall with a network sensor is the best practice. Sensor 2, connected on the outside network, sees all traffic targeted for the organization, including the traffic that is blocked by the firewall and all traffic leaving the organization's network. This sensor also monitors the DMZ traffic and inside traffic. Knowing what traffic is denied or permitted by the firewall, the network administrator must find out what reported intrusions reported by Sensor 2 are a danger for the network. This sensor also needs to protect the firewall itself against DoS attacks and tools generating noise on the network. Sensor 3 enables you to see which users are attempting to gain access to the protected network (DMZ). All three sensors provide visibility into which vulnerabilities are being exploited to attack servers, hosts, and so on.

Once you have decided which critical assets require network monitoring, the sensors can be connected, starting with the data capturing (sniffing) interface. It may sound ridiculous, but if the sensor cannot see the interested traffic, it does not function properly. It is straightforward to connect the sensor to a network segment by plugging the interface into an open port on a hub, but this becomes an issue in switched environments, where traffic is only aggregated on the backplanes of the devices. In these environments, you can solve the problem by using integrated switch sensors with traffic-capture functions. The SPAN feature or VACL feature can monitor traffic.

After connecting the network sensor interfaces, the sensor can be configured either locally via a console or remotely using a network management station.

Before starting to tune the sensor, which is the most important part of the network IDS deployment, it is recommended to use the sensor with the initial sensor configuration and analyze the alarms generated the first couple days. Analyzing the different alarms and tuning out the false positives produces a high-performing security system. Also keep in mind that not every sensor needs to trigger an alarm on every event. Here again, the importance of clearly defined network security policies is obvious. It is also clear that tuning the sensors is an iterative process. Traffic patterns can and do change over time, and sensor tuning is a must.

Once the initial tuning phase is finished, the network administrator can selectively implement response actions. Small organizations that are willing to investigate the deployment of IDSs can start deploying Cisco IOSbased IDSs on a router or PIX-based IDS, instead of buying standalone sensors.

9.4.3. Router IDS Features and Network Modules

The router IDS feature is a built-in functionality in Cisco IOS, enabling the router to be configured as network intrusion detection sensors. The sensors have only a limited number of signatures.

Because Cisco Secure Integrated Software is an in-line device, it inspects packets as they traverse the router's interfaces. This impacts network performance to a certain extent. When a packet, or a number of packets in a session, matches a signature, the router configured as network IDS can perform the following configurable actions:

- **Alarm** Sends an alarm to syslog server or management station
- **Drop** Drops the packet
- **Reset** Resets the TCP connection

The router IDS module is a hardware router module that can be installed in an empty slot in either a 2600, 3600a or 3700 router. Once the module is plugged into the router, it acts similar to a standalone IDS network sensor and can be configured and monitored via a remote management console.

9.4.4. PIX IDS

The PIX Firewall can also be configured as a network intrusion detection sensor in a manner similar to the router IDS.

The IDS integrated software for the PIX makes it possible, although in a very limited way, to customize the amount of traffic that needs to be audited and logged. Application-level signatures can be audited only for active sessions through the PIX. This audit needs to be applied to either the inbound or outbound interface of the PIX Firewall.

For auditing performed inbound, the PIX looks at the IP packets as they arrive at an input interface. For instance, if a packet triggers a signature and the configured action does not drop the packet, the same packet can trigger other signatures.

9.4.5. Response to Events and Alerts

IDSs can respond to attacks in a few different ways, including by passively creating IP session logs or by actively terminating the session or blocking the attacking host.

9.4.5.1.IP Session Logging

After a sensor detects an attack, an alarm is generated by the sensor and sent to the management station. The information is saved in a memory-mapped file on both the sensor and the management platform. This memory-mapped file is in binary format file. As discussed in the next section, the sensor uses RDEP to communicate with the external world; so does the IP logging feature. It is an HTTP communication that is client-server and two-way based, whereby the client (sensor) sends an RDEP request, which is answered by the management station with an RDEP response. All RDEP messages consist of two parts:

Header

Entity body

Figure 9.12 illustrates the IP logging capability of the network IDSs.

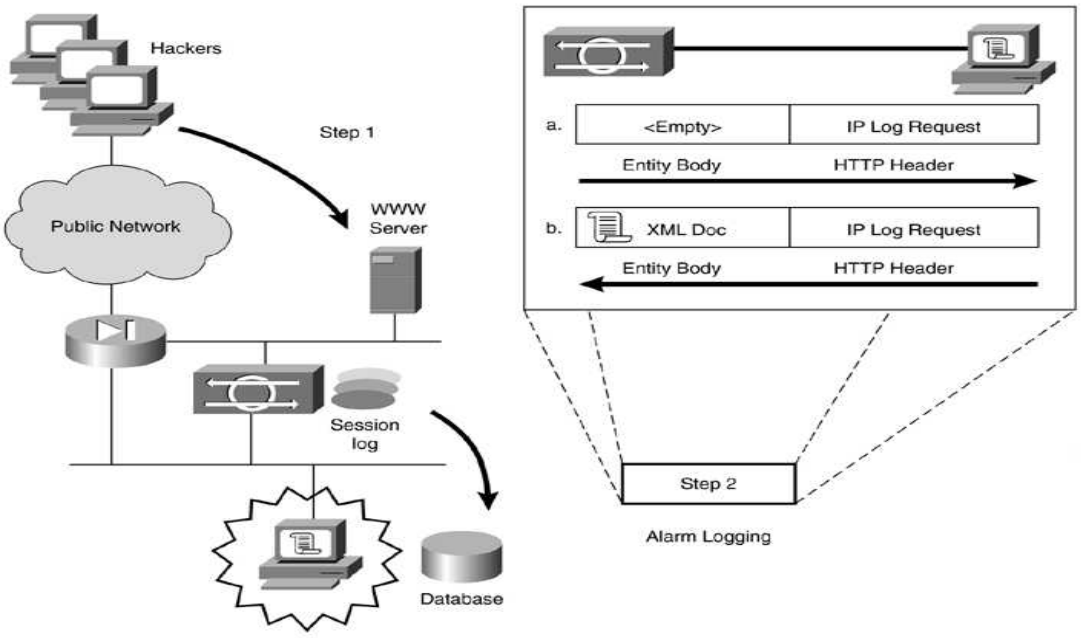


Figure 9.12: Network-Based IDS Logging

Step 1 illustrates the initial attack on the web server. The network IDS notices the attack and sends an alarm to the management server (step 2 in Figure 9.12). The communication between server and sensor is a two-way mechanism. The IP log feature captures the session in a pcap file. Once the event occurs, the IP log response that is sent from the server to the sensor is in HTML/XML format. This response contains an error status code and a description of the event. This response is sent from the server to the sensor.

The IP logging feature allows the network administrator to easily archive the data, write scripts for parsing the data, and monitor the attacks. The IP logging feature is helpful to analyze events, but it does impact sensor performance; therefore, disk utilization needs to be watched carefully.

9.4.5.2.Active ResponseTCP Resets

After a sensor detects an attack, an alarm is generated by the sensor and sent to the management station. The network IDS may terminate the Layer 4 session by sending a TCP RST packet to the attacked server and the host. Figure 9.13 illustrates the TCP reset capability of the network IDSs.

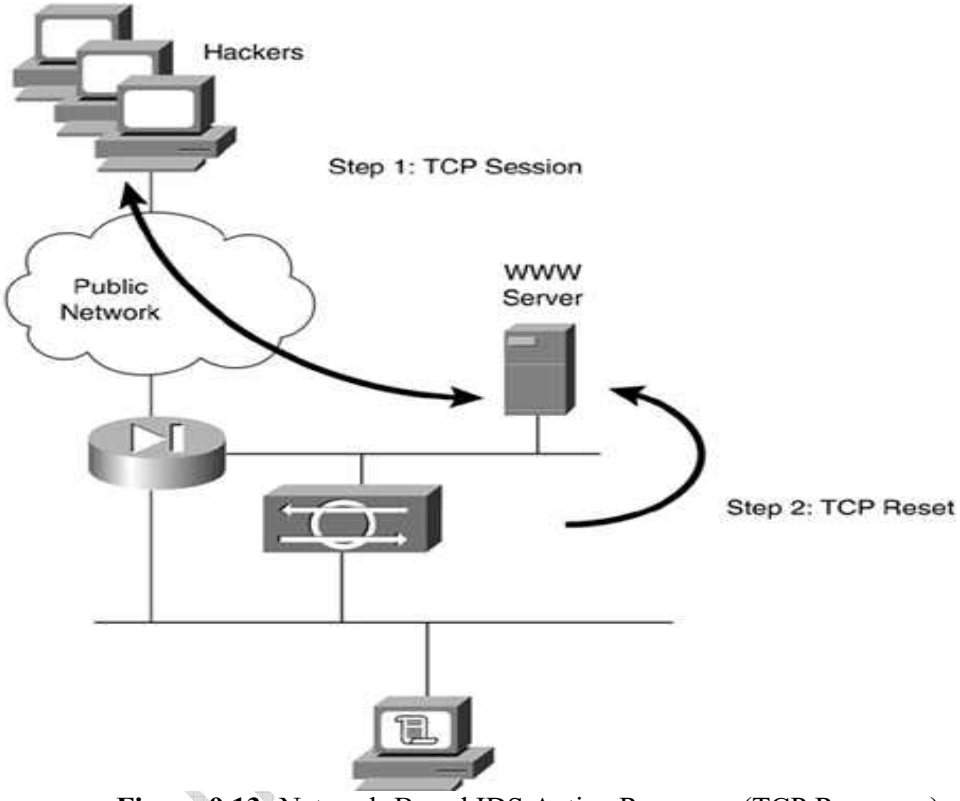


Figure 9.13: Network-Based IDS Active Response (TCP Response)

The TCP Reset is initiated from the data-capturing port to both the server and the cracker's host. The network administrator should be aware that certain applications automatically reconnect and resend data. A solution would be to implement a blocking mechanism.

9.4.5.3.Active Response Shunning or Blocking

After a sensor detects an attack, an alarm is generated by the sensor and sent to the management station. The network IDS can shut the attacker out of the network, usually by setting access control rules on a border device such as a router or firewall. Figure 9.14 illustrates the IP blocking capability of the network IDSs.

In Figure 9.14, the sensor connects to the router and configures an access list to block traffic originated for the offending host with IP address 10.0.0.1.

Special precautionary measures should be taken when implementing these active responses. The attacker (who is also aware of these features) can inappropriately deny service for authorized user traffic. General guidelines on responses to alerts and events are difficult to outline. But it is not recommended to use active responses during the tuning period. Shunning or blocking should be used only as the administrator gains experience with the traffic patterns in the network. Starting with TCP

resets is recommended instead. And last but not least, keep in mind that the initial trigger packet still makes it to the destination.

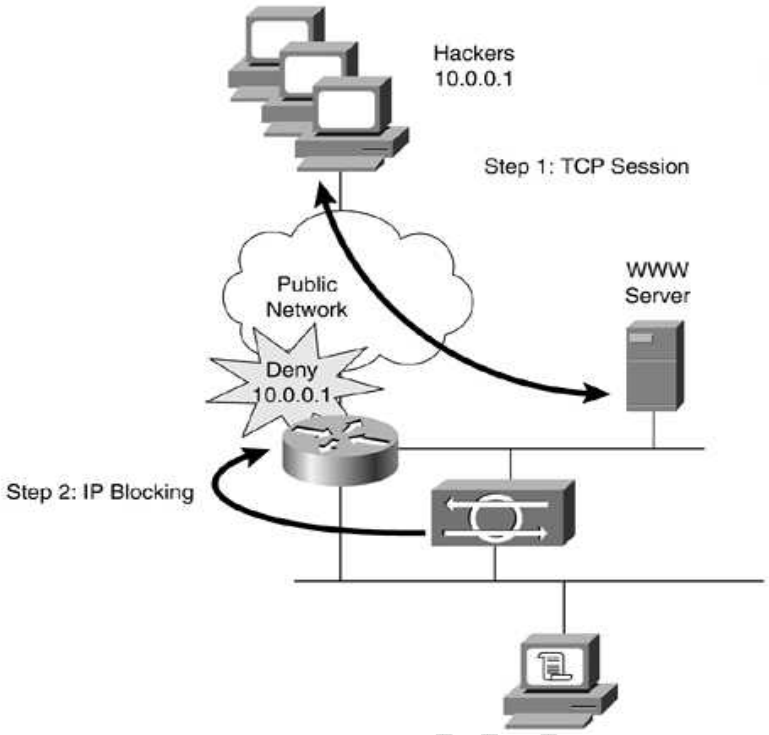


Figure 9.14: Network-Based IDS Active Response (Shunning or Blocking)

9.4.6. Notification and Reporting

The graphical user interface of the management station provides an excellent vehicle to view alarms generated by the various sensors throughout the network. Each alarm is displayed with a unique color based on the severity of the alarm. The administrator can quickly view all the intrusions occurring in the network at any time based on the generated alarms. This alarm information can also be saved in a text log file.

From a notification viewpoint, there are two options. The system can be configured to inform security personnel either by an e-mail message or by pager. Both mechanisms have their advantages and disadvantages, including notification time, ability to keep records for tracking, and so on.

The Cisco Secure Policy Manager and the Cisco VMS Management Center for IDS have a powerful alarm-reporting feature that provides the network security administrator with a tool to generate customized intrusion detection reports. These reports can be generated via HTTP, HTTPS, or on the network management console.

The following list gives an idea of some available reports:

- Intrusion detection summary
- Top sources of alarms
- Top destinations of alarms
- Alarms by day
- Alarms by sensor

9.5. IDS Management Communications Monitoring the Network

Network device management requires a communications channel to be available to the network devices. Devices may support out-of-band management, in-band management, or both. In-band management consumes bandwidth that could otherwise be used by network traffic. Out-of-band management increases bandwidth available for network traffic and typically improves the privacy and security of network management communications. The benefits are achieved in the

reduced cost of designing, provisioning, and managing the management network itself. In any case, the management channels should be robust, private, and secure.

9.5.1. Communication SyntaxRDEP

The data format used on the communication channel, which is set up between the network IDS sensor and the management station (often called the IDS director), is defined by the RDEP protocol. As of version 4.x of IDS sensor software, RDEP is used instead of PostOffice Protocol, which was used by earlier versions. The RDEP communication channel is critical to the success of an IDS and therefore must comply with some minimum requirements.

Figure 9.15 shows this communication channel, which is also referred to as the command and control network. The data link is referred to as the monitoring network.

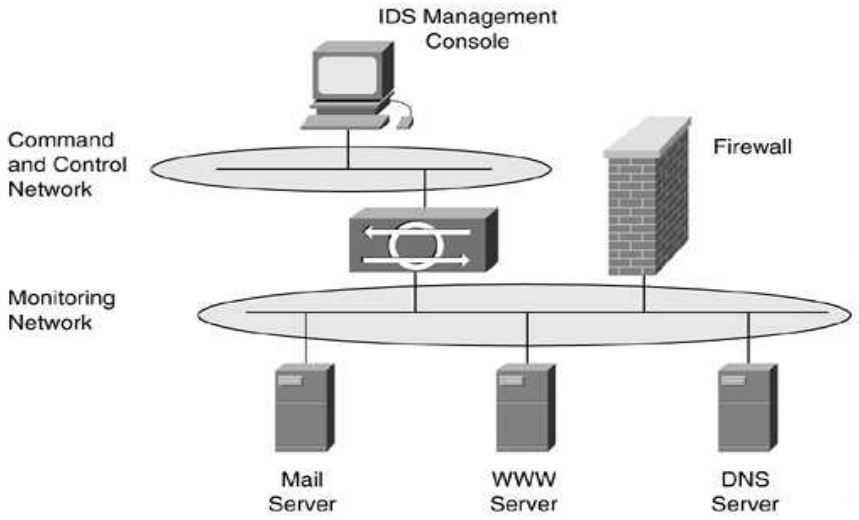


Figure 9.15: Example of IDS Installation with Device Management

External communication, or data exchange, between the sensor and the external systems uses XML data format. RDEP uses HTTP, or in some cases TLS/SSL, to pass these XML documents between the sensor and the director. The RDEP protocol communication consists of two message types, namely the RDEP request and the RDEP response message. These messages can be event messages or IP log messages, as you noticed in the previous section on IP logging.

The RDEP protocol is designed to be reliable, redundant, and fault tolerant. Guaranteed or reliable packet delivery is assured because all messages (alarms) sent by the sensor require an acknowledgement by the management station within a predefined period of time.

Figure 9.16 illustrates a fault-tolerant setup with the RDEP protocol.

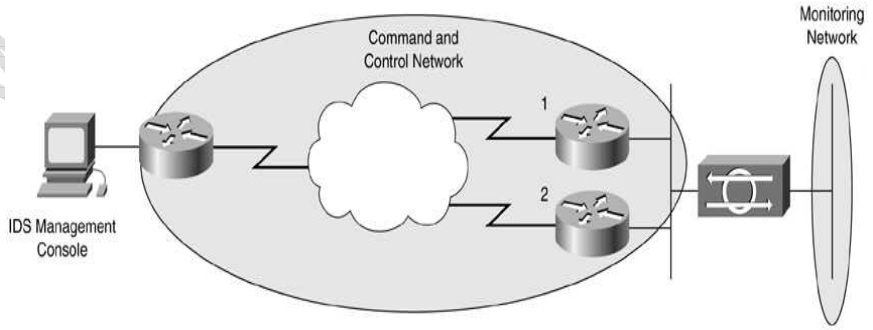


Figure 9.16: Fault Tolerant Setup with RDEP

9.5.2. Out-of-Band Management

Preparation for the worst-case network management scenario includes ensuring that there is a way to reach the devices when the usual access channel is unavailable. Out-of-band management using modem access through a management port is an attractive option when combined with

authentication and access controls. Direct connection to management ports using serial communication cables is a final, labor-intensive option.

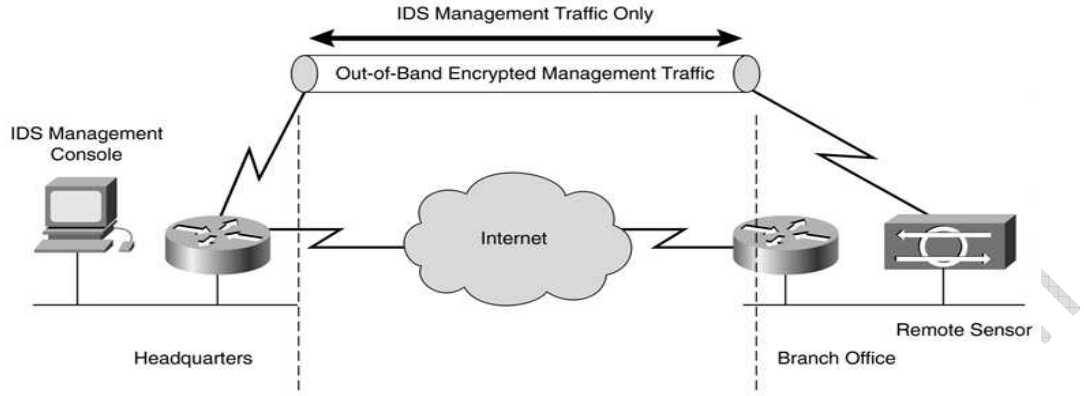


Figure 9.17: Remotely Installed Sensor as an Example of Out-of-Band IDS Management

Out-of-band management offers many significant advantages and becomes more desirable as the managed network grows. In this case, real-time monitoring and access can be performed over a protected channel, which does not impact transport bandwidth availability. In a large network, the costs of provisioning and maintaining the management network are less proportional than in a small network. Out-of-band management is a part of the Enterprise Composite Network Model and Security Architecture for Enterprises (SAFE) as applied to large enterprises.

9.5.3. In-Band Management

In-band management is appropriate in smaller networks and in networks with sufficient link capacities to support both application traffic and management activity. Securing access to the devices and management applications is an important consideration. When supported, secured VPN access in-band may provide access if a management network is lost. Mechanisms to secure the management command and data stream include IPSec tunnels, secure shell (SSH), and secure sockets layer (SSL). In-band communication channels are often the only option for managing remotely installed network sensors, such as securing management traffic for branch offices if the IDS directors are installed at the company headquarters. Figure 9.18 illustrates this scenario.

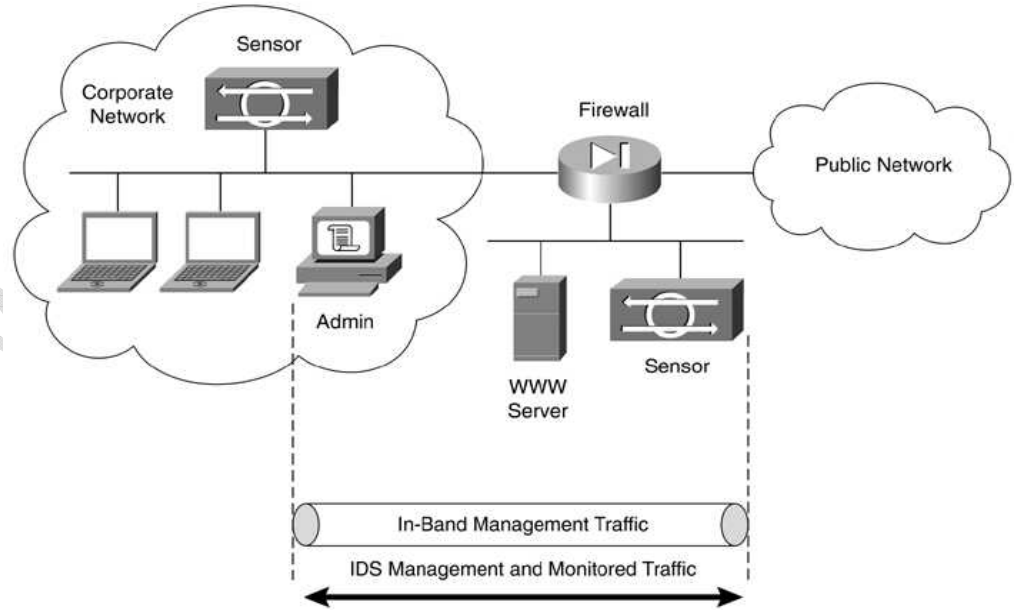


Figure 9.18: Example of In-Band IDS Management

9.6. Sensor Maintenance

Most IDSs are signature-based systems and require a level of maintenance. In particular, to detect recent attacks accurately, the sensor needs to install new signatures as they become available.

Signature updates, which also contain network security database (NSDB) updates, occur every two months. Service packs are released as needed to address software bugs or improvements to the core IDS software components (analysis engine, web software, and so on).

There are two ways to automate this process:

Automatic updates (Auto Update Server): A configuration option for some IDS sensors, providing the functionality to have signature updates applied automatically to the sensor.

Active update notification: A service available at Cisco.com. Using this service, the subscriber receives updates on changes to IDS signatures as well as information on how to obtain changes.

9.7. Conclusion

It is hard to tell which IDS method is best. The choice depends on what you are trying to achieve as network administrator. The Cisco philosophy to date has been to combine the use of pattern matching, stateful-pattern matching, protocol decodes and heuristic-based signatures.

Cisco and other vendors continue to research and monitor developments in the IDS arena and incorporate new techniques as they become efficient, practical, and commercially feasible.

9.8. Important Questions:

1. List two weaknesses of the signature-based IDS.
2. Why does the deployment of policy-based IDS take a long time?
3. Which IDS is not limited by bandwidth restrictions or data encryption?
4. Which IDS is very challenging in a switched environment?
5. Name the two main components of a Cisco host IDS.
6. Name the two interfaces of a network IDS.
7. What are the three main components of a network IDS?
8. List three responses to events or alerts.
9. What two processes are in place to automate sensor maintenance?
10. The RDEP protocol communication consists of what two message types?

Chapter 10
Remote Access

10.1. Introduction:

The overall goal of remote access is to grant trusted access for telecommuters, salespeople, and road warriors to the corporate network over an untrusted network such as the Internet. The concluding case study is a practical example of how organizations can provide access to their networks in a secure manner, thereby enabling a worldwide workforce to use remote access technology.

10.2. AAA Model

Authentication, authorization, and accounting (AAA, pronounced "triple A") provide security to Cisco IOS routers and network devices.

AAA provides a method for identifying users who are logged in to a router and have access to servers or concentrators. AAA also identifies the level of access that has been granted to each user and monitors user activity to produce accounting information.

Network data can be accessed via a variety of methods, including the following:

- Dialup connections
- Integrated services digital networks (ISDNs)
- Broadband cable and asymmetric digital subscriber lines (ADSLs)
- Access through the Internet via virtual private networks (VPNs)

The AAA model was designed in such a way that all these access methods can benefit from the AAA security features.

The three phases (authentication, authorization, and accounting) ensure that only legitimate users are permitted access, as explained in the following list:

- Authentication Verification of who you are. Remote users must be authenticated before being permitted access to network resources by confirming their identities.
- Authorization Control of what you can do. Once the user is identified, the accessible resources are defined by the authorization mechanism.
- Accounting tracking what you have done. Timestamps, command history, and type of resources are just a few examples of information collected by the accounting mechanism.

Authentication allows the users to submit their usernames and passwords through a series of challenges and responses.

Authorization defines what services in the network the users are permitted to access. The operations permitted may include the Cisco Internet Operating System (IOS) privileged executive commands that are permitted. For example, a user may be allowed to type commands, but only the certain show and debug commands that are authorized. This is demonstrated later in the chapter through examples.

Accounting allows the network administrator to log and view what actions were performed, such as whether a Cisco router was reloaded or the configuration was changed. The accounting function ensures that an audit allows network administrators to view which actions were performed and at what time.

The AAA server handles all three functions: authentication, authorization, and accounting. Figure 10.1 displays a typical network setup with a AAA server securing the network.

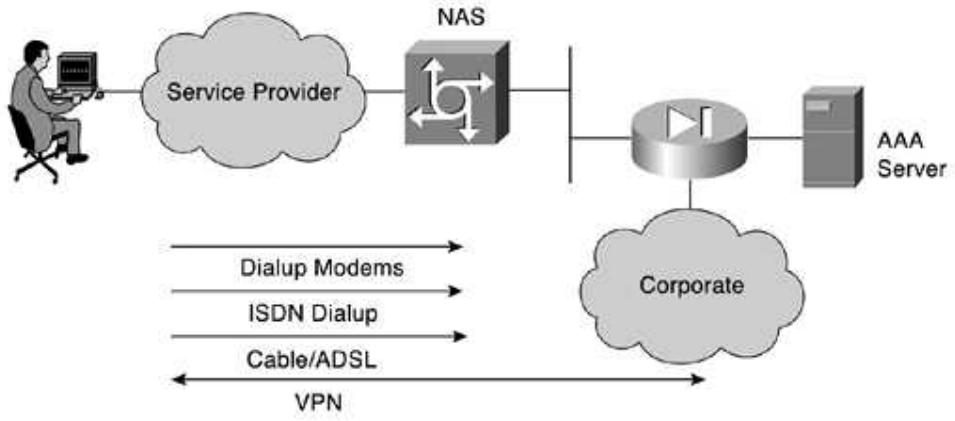


Figure 10.1: AAA Server Securing the Network

The remote users may be using dialup connections and running Async (PSTN) or using ISDN with Point-to-Point Protocol (PPP). Broadband access users could be using cable or ADSL connections. The Network Access Server (NAS) ensures that only authenticated users have access to the secure network. NAS also maintains resources and accounting information. The NAS depends on the AAA server to get the user-specific information.

Authorization controls which resources (FTP servers, web servers, and so on) are accessible. The NAS is configured with the AAA protocols and interacts with the AAA server to collect data on the network resources accessed.

10.2.1.Authentication

Authentication allows administrators to identify who can connect to a router by comparing the usernames and passwords of those seeking access with the usernames and passwords in an authorized list or database. Normally, when a user connects to a router remotely via Telnet, the user needs to supply only a password, and the administrator has no way of knowing the user's username. With AAA authentication, whenever a user logs on, the user must enter a username and a password, which have been assigned by the administrator.

Example 10.1 displays two types of remote access: a remote user accessing a router via Telnet without AAA and a remote user accessing a AAA-configured Cisco router.

Example 11-1. AAA vs. Router Configured Without AAA

```
Brussels#telnet nonAAA_router
User Access Verification
Password: xxxxxxxx
nonAAA_router>

Brussels#telnet AAA_router
Trying AAA_router (10.1.1.1)... Open User Access Verification
Username: Gert
Password: xxxxxxxx
AAA_router>
```

As you can see in Example 10.1, the user must enter a valid username and password to access a AAA-configured Cisco router. Both username and password are set to "Gert" in this case. Typically, a database contains the valid usernames that reside on a remote AAA server. Cisco IOS can also create a local database on the router, but this is not a scalable solution.

10.2.2.Authorization

Authorization is the second step in the AAA process. Authorization allows administrators to control the level of access users have after they have successfully gained access to a device. For the sake of simplicity, this section focuses on accessing a router. Cisco IOS allows certain access levels

(also called privilege levels) that control which Cisco IOS commands the user can issue. These levels range from 0 to 15. For example, a user with a privilege level of 0 cannot issue any Cisco IOS commands. A user with a privilege level of 15 can perform all valid Cisco IOS commands. The local database or remote security server (AAA server) can grant the required privilege levels.

Remote security servers, such as RADIUS and TACACS+ (which are discussed later in the chapter), authorize users for specific rights by associating attribute-value (AV) pairs, which define those rights, with the appropriate user. AAA authorization works by assembling a set of attributes that describe the tasks the user is authorized to perform. These attributes are compared with the information contained in a database for a given user, and the result is returned to the AAA software to determine the user's actual capabilities and restrictions.

You can display your privileged level on a Cisco router with the show privilege command. Example 10.2 displays the privilege level when the user has already been authenticated for the AAA_router.

Example 11.2: show privilege Command Output

```
AAA_router#show privilege
Current privilege level is 15
```

The higher the privilege, the more capabilities a user has with the Cisco IOS command set.

10.2.3.Accounting

Accounting occurs after the authentication and authorization steps have been completed. Accounting allows administrators to collect information about users. More specifically, administrators can track which user logged in to which router, which CISCO IOS commands a user issued, and how many bytes were transferred during a user's session. Accounting information can be collected by a router or by a remote security server. For simplicity's sake, the output of the router command is displayed. The case study at the end of the chapter supplies more details on the AAA server output.

To display local account information on a Cisco router that is collecting accounting information, issue the show aaa user all CISCO IOS command.

The most important accounting function records are

- Network
- EXEC
- Connect
- Command

The Network accounting function monitors dialup and PPP authentication. The EXEC function helps to monitor login authentication. The Connect function monitors connection parameters, and the Command function is used for the show command and debug monitoring.

Rather than maintaining a separate database with usernames and passwords and privilege levels, you can use an external security server to run external security protocols namely Remote Authentication Dial-In User Service (RADIUS), Terminal Access Controller Access Control System plus (TACACS+), and Kerberos. These protocols provide a more scalable solution for network environments that deploy large networks and need granular control.

These security server protocols allow you to stop unauthorized access to your network. The upcoming sections review these three security protocols.

10.3. AAA Servers

In many circumstances, AAA uses security protocols to administer its security functions. If your router, concentrator, or even PIX is acting as an NAS, AAA is the means through which you establish communication between your NAS and your TACACS+, RADIUS, or Kerberos security server.

10.3.1.TACACS+ Overview

Cisco IOS supports three versions of TACACS: TACACS, extended TACACS, and TACACS+. All three methods authenticate users and deny access to users who do not have a valid username and password pairing. This section covers only TACACS+ (also referred to as "TACACS plus").

Figure 10.2 displays a typical TACACS+ connection request. (This example shows PPP authentication using TACACS+, whereas the previous example showed user exec authentication).

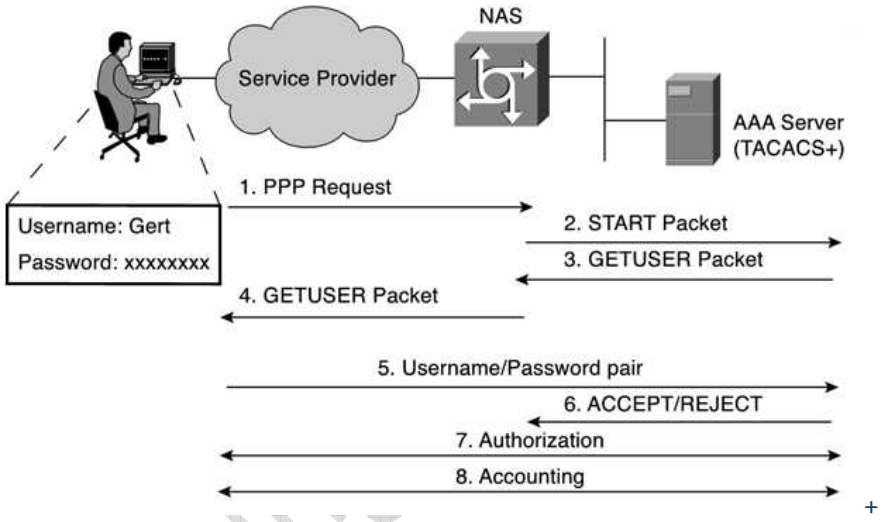


Figure 10.2: TACACS+ Authentication Example Sequence

When a TACACS+ server authenticates a remote user, the following events occur, which are illustrated in Figure 10.2:

Step 1. When the connection is established, the NAS contacts the TACACS+ service to obtain a username prompt, which is then displayed to the user. The user enters a username, and the NAS then contacts the TACACS+ service to obtain a password prompt. The NAS displays the password prompt to the user, the user enters a password, and the password is then sent to the TACACS+ service (steps 1 through 5 in Figure 10.2).

Step 2. The NAS eventually receives one of the following responses from the TACACS+ daemon:

- ACCEPT The user is authenticated and service may begin. If the NAS is configured to require authorization, authorization begins at this time (step 7 in Figure 10.2).
- REJECT The user fails to authenticate. The user may be denied further access

or will be prompted to retry the login sequence, depending on the TACACS+ daemon.
- ERROR An error occurs during authentication. This can be either at the daemon or in the network connection between the daemon and the NAS. If an ERROR response is received, the NAS typically tries to use an alternative method for authenticating the user.
- CONTINUE The user is prompted for additional authentication information.

Step 3. A Password Authentication Protocol (PAP) login is similar to an ASCII login, except that the username and password arrive at the NAS in a PAP packet instead of being typed in by the user. Therefore, the user is not prompted. Challenge Handshake Authentication Protocol (CHAP) logins are also similar in principle.

Step 4. Following authentication, the user is also required to undergo an additional authorization phase, if authorization has been enabled on the NAS. Users must first successfully complete TACACS+ authentication before proceeding to TACACS+ authorization.

Step 5. If TACACS+ authorization is required, the TACACS+ daemon is again contacted, and it returns an ACCEPT or REJECT authorization response. If an ACCEPT response is returned, the response contains data in the form of attributes that are used to direct the EXEC, NETWORK, COMMAND, or CONNECT session for that user, determining services that the user can access.

Services include the following:

- Telnet, rlogin, PPP, Serial Line Internet Protocol (SLIP), or EXEC services
- Connection parameters, including the host or client IP address, access list, and user timeouts

Table 10.1 shows the main TACACS+ characteristics.

Table 10.1: Summary of TACACS+ Protocol

Features	**Meaning**
TCP	Packets sent between client and server are TCP.
TCP destination PORT	Port 49.
Attributes	Packet types are defined in TACACS+ frame format as: Authentication 0x01 Authorization 0x02 Accounting 0x03
SEQ_NO	The sequence number of the current packet flow for the current session. The SEQ_NO starts with 1, and each subsequent packet increments by one. The client sends only odd numbers. The TACACS+ server sends only even numbers.
Encryption method	Entire packets are encrypted. Data is encrypted using MD5 and a secret key that matches on both the NAS (for example, a Cisco IOS router) and the TACACS+ server.

Figure 10.3 displays a screenshot of an ACS setup for TACACS+ authentication.

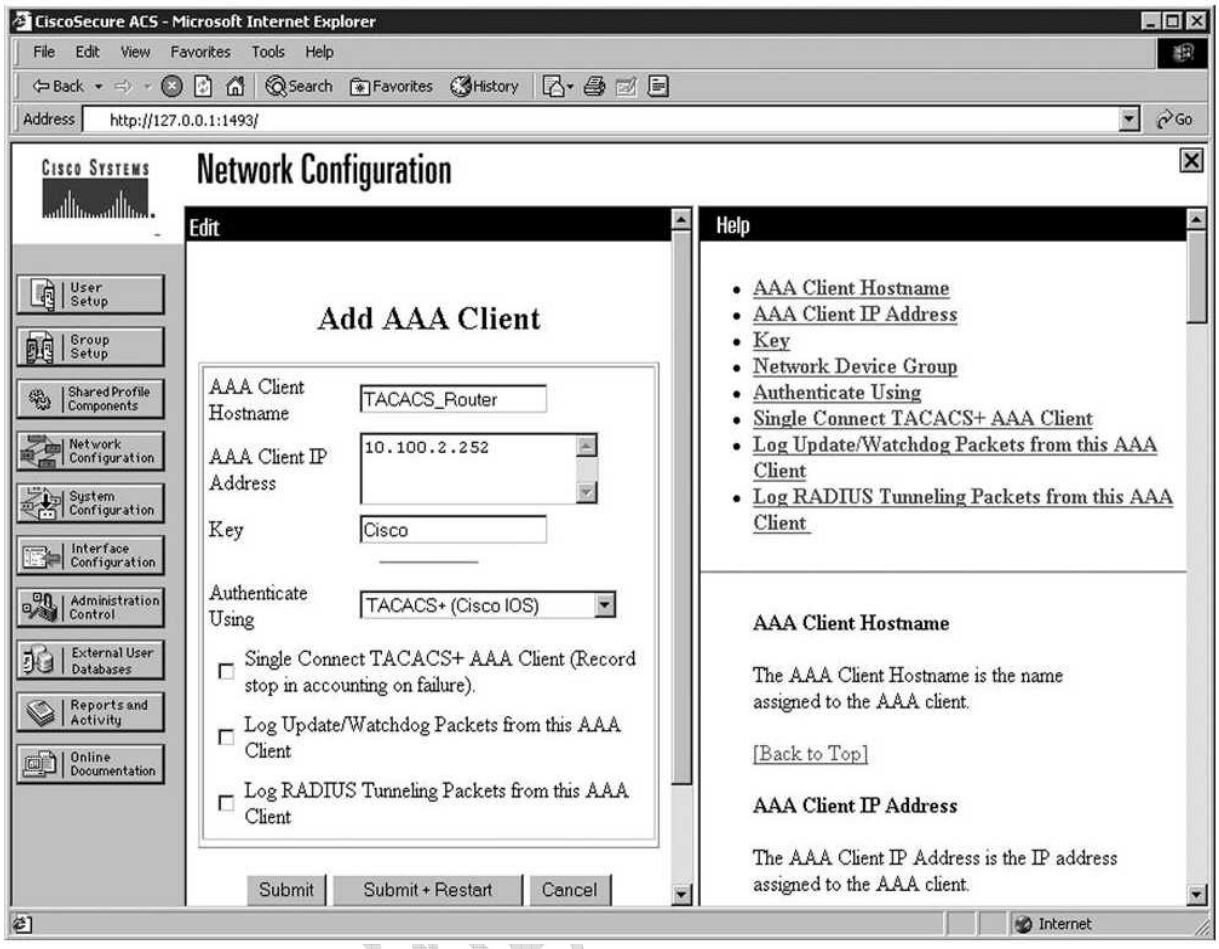


Figure 10.3.: ACS Setup for TACACS+ Authentication

TACACS+ accounting provides an audit record of what commands were completed. When NAS sends a record of commands, the TACACS+ server sends a response acknowledging the accounting record.

10.3.2.RADIUS Overview

RADIUS is a client-server based system that secures a network. RADIUS is a protocol that is implemented in all Cisco devices that send authentication requests to a RADIUS server. RADIUS is defined in RFC 2138/2139.

A RADIUS server is a device that has the RADIUS daemon or application installed. RADIUS must be used with AAA to enable the authentication, authorization, and accounting of remote users when using Cisco devices (routers, switches, firewalls, or concentrators).

Figure 10.4 displays a typical RADIUS connection request (authentication).

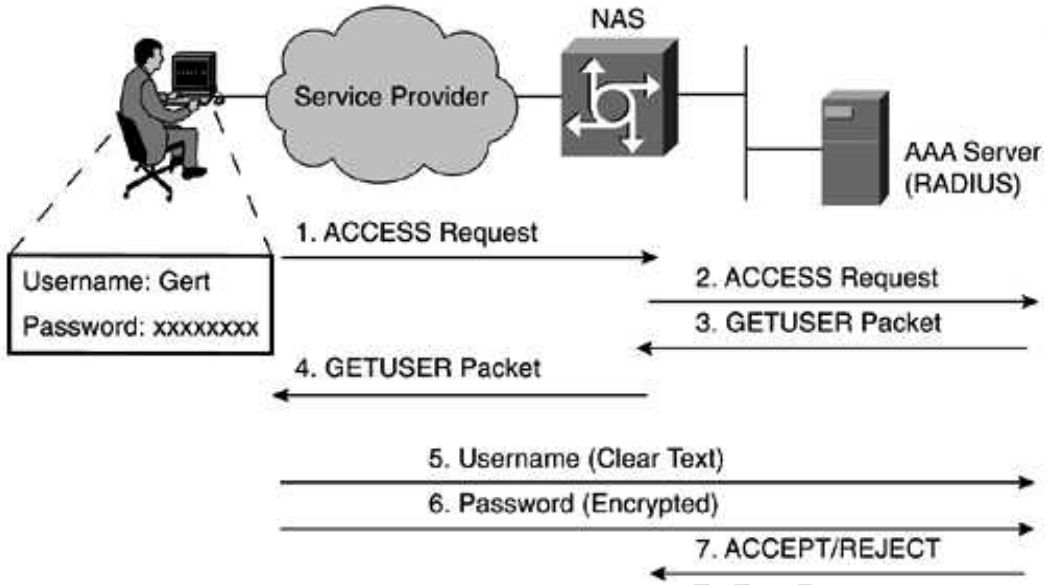


Figure 10.4: RADIUS Authentication Example Sequence

When a RADIUS server authenticates a remote user, the following events occur:

Step 1. When the connection is established, the NAS contacts the RADIUS daemon to obtain a username prompt, which is then displayed to the user. The user enters a username, and the NAS then contacts the RADIUS daemon to obtain a password prompt. The NAS displays the password prompt to the user, the user enters a password, and the password is then sent to the RADIUS daemon (steps 1 through 6 in Figure 10.4).

Step 2. When a RADIUS server authenticates a user, the following events occur:

a. The user is prompted for and enters a username and password.

b. The username and encrypted password are sent over the network to the RADIUS server.

c. The user receives one of the following responses from the RADIUS server:
- ACCESS-ACCEPT The user is authenticated.
- ACCESS-REJECT The user is not authenticated and is prompted to reenter the username and password, or access is denied. This response is sent from the RADIUS server when the user enters an invalid username/password pairing.
- CHALLENGE A challenge is issued by the RADIUS server. The challenge collects additional data from the user.
- CHANGE-PASSWORD A request is issued by the RADIUS server, asking the user to select a new password.

Services that are accessible for the user include Telnet, rlogin, or local-area transport (LAT) connections, and PPP, SLIP, or EXEC services.

The use of a shared secret authenticates transactions between the NAS and the RADIUS server. The username is sent as clear text. RADIUS supports both PAP and CHAP. It is important to realize that a RADIUS server never sends the user's password over the network. If the username and password pairing are entered incorrectly, the RADIUS server sends an ACCESS_REJECT response. The end user must reenter the pairings or the connection is rejected.

RADIUS supports a number of predefined attributes that may be exchanged between client and server, such as the client's IP address. RADIUS attributes carry specific details about

authentication. These attribute pairs are also referred to as AV pairs.

RFC 2138 defines a number of attributes. The following bulleted list provides details for the most common attributes:

- Attribute type 1Username Defines usernames such as numeric, simple ASCII characters, or a Simple Mail Transfer Protocol (SMTP) address.
- Attribute type 2User Password Defines the password, which is encrypted using MD5.
- Attribute type 3CHAP Password Only used in access-request packets.
- Attribute type 4NAS IP Address Defines the IP address of the NAS server; used only in access-request packets.
- Attribute type 5NAS Port Indicates the physical port number of the NAS; ranges from 0 to 65535.
- Attribute type 6Service Type Type of service requested; not supported for Cisco devices.
- Attribute type 7Protocol Defines required framing; for example, PPP is defined when this attribute is set to 1 and SLIP is set to 2.
- Attribute type 8IP Address Defines the IP address to be used by the remote user.
- Attribute type 9IP Subnet Mask Defines the subnet mask to be used by the remote user.
- Attribute type 10 Defines framed-routing to send and/or listen for routing packets.
- Attribute type 13 Defines utilization of framed compression.
- Attribute type 19 Defines the Callback ID used to authenticate.
- Attribute type 26Vendor specific Cisco (vendor-ID 9) uses one defined option: vendor type 1 named cisco-avpair; this attribute transmits TACACS+ A/V pairs.
- Attribute type 61NAS Port Type Defines the NAS port type (Async, ISDN Sync, ISDN Async, and Virtual.

Table 10.2 summarizes the main features of RADIUS.

Table 10.2: Summary of RADIUS Protocol

Features	**Meaning**
UDP	Packets sent between client and server use the User Datagram Protocol (UDP) primarily because the overhead of the Transmission Control Protocol (TCP) does not allow for significant advantages. Typically, the user can wait for a username and password prompt.
UDP destination PORT	RADIUS uses two sets of ports. The pre-RFC ports of 1645 and 1646 are widely used. Ports 1812 and 1813 are defined in RFC 2138.
Attributes	Attributes are used to exchange information between the NAS and the client.
Model	Client/server-based model, in which packets are exchanged in a unidirectional manner.
Encryption method	Password is encrypted using MD5; the username is not. RADIUS encrypts only the password in the access-request packet, from the client to the server. The remainder of the packet is transmitted in clear text. A third party can capture other information such as username, authorized services, and accounting.
Multiprotocol support	Does not support protocols such as AppleTalk, NetBIOS, or IPX. IP only is supported.

Figure 10.5 displays a screenshot of an ACS setup for RADIUS authentication

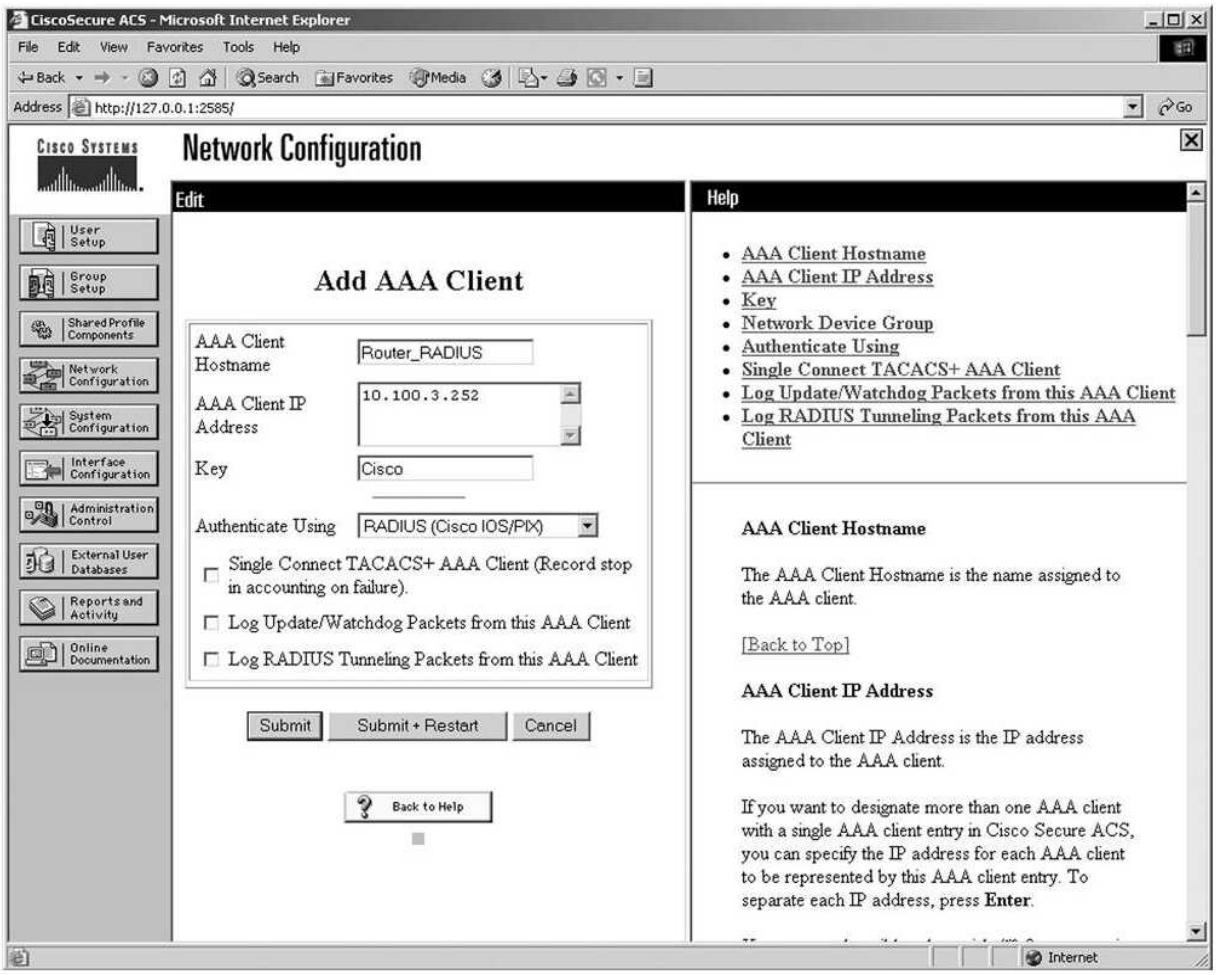


Figure 10.5: ACS Setup for RADIUS Authentication

A RADIUS server is usually software that runs on a variety of platforms, including Microsoft NT servers or a UNIX host. RADIUS can be used to authenticate router users, authenticate vendors, and even validate IP routes.

10.3.3.Kerberos

Kerberos is a trusted third-party authentication application layer service (Layer 7 of the OSI model), relying heavily on an authentication technique involving shared secrets. The basic concept is quite simple: If a secret is known by only two people, then either person can verify the identity of the other by confirming that the other person knows the secret.

Kerberos is a secret-key network authentication protocol, developed at the Massachusetts Institute of Technology (MIT), that uses the Data Encryption Standard (DES) cryptographic algorithm for encryption and authentication. In the Kerberos protocol, this trusted third party is called the key distribution center (KDC).

Figure 10.6 displays the authentication process Kerberos uses when a remote client initiates a remote Telnet session. (Kerberos supports Telnet, rlogin, remote shellrsh, and remote copyrcp.)

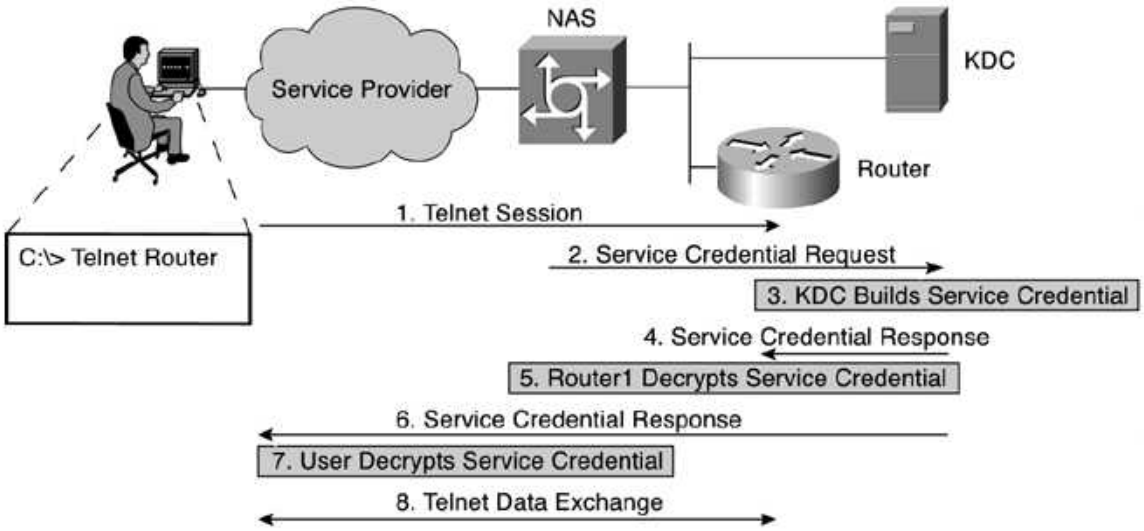


Figure 10.6: Kerberos Authentication Example Sequence

When Kerberos is used for authentication for a remote user, the following events occur, as shown in Figure 10.6:

Step 1.	User initiates a Telnet session to the router.
Step 2.	The NAS builds a credential request and sends it to the KDC.
Step 3.	The KDC decrypts the request and builds a service credential.
Step 4.	The KDC sends the service credential to the router.
Step 5.	The router decrypts the service credential.
Step 6.	The KDC sends the service credential to the user.
Step 7.	User decrypts the service credential.
Step 8.	An authenticated Telnet session to the router is established and data exchange can start.

The primary use of Kerberos is to verify that users and the network services they use are really who and what they claim to be. To accomplish this, a trusted Kerberos server issues tickets to users. These tickets, which have a limited lifespan, are stored in a user's credential cache and can be used in place of the standard username-and-password authentication mechanism.

The Kerberos credential scheme embodies a concept called "single logon." This process requires authenticating a user once and then allows secure authentication (without encrypting another password) wherever that user's credential is accepted.

Table 10.4 summarizes the key characteristics of Kerberos.

Table 10.4: Characteristics of the Kerberos Protocol

Attribute	Meaning
Packet delivery	A number of ports are defined: TCP/UDP ports 88, 543, 749, and TCP ports 754, 2105, 4444.
Packet encryption	This supports username/password encryption.
Telnet support	Telnet sessions can be encrypted.

10.4. Lock-and-Key Feature

The lock-and-key feature uses dynamic access lists to create specific temporary openings in the network in response to a user authentication success.

Lock-and-key is a traffic-filtering security feature that dynamically filters IP protocol traffic to grant access per user to a specific source/destination host. Lock-and-key is configured using IP dynamic extended access lists. It is the dynamic functionality that makes this feature so interesting. Access lists are typically created and maintained by manually defining the lists and then distributing or deploying them to all other devices in the network. This feature can be used in conjunction with other standard access lists and static extended access lists. It is recommended to use the lock-and-key feature in combination with a AAA server (either TACACS+ or RADIUS) to provide authentication, authorization, and accounting services. Although the lock-and-key is server independent, it is ideally designed for the TACACS+ server. TACACS+ has three components to provide authentication, authorization, and accounting services: protocol support within access servers and routers, protocol specification, and a centralized security database.

10.5. Two-Factor Identification

With increased focus on productivity, remote access for the workforce is a must. Network administrators are required to open more doors to more users, and an identity method that scales well and is cost effective is necessary. The more you know where your network is heading, the better you can plan your identification strategy.

Given the expense required to create an infrastructure for biometrics, a good compromise is two-factor identification: a combination of digital signatures and passwords. In general, two-factor identification consists of any two of the following: something you know, something you have, and something you are. Here are a few examples of two-factor identification. Organizations that adopt a PKI can do so with minimal expense and can protect their property much more effectively than they could with passwords alone. Everyone uses two-factor authentication technology on a daily basis. When retrieving money from an ATM account, for example, a customer needs both a PIN number and the magnetic-strip card. Even if someone attains the PIN number, the card is also needed for access. If the card is lost or stolen, it cannot be used without the PIN.

Other examples are a combination of two pieces of information to validate a person's identity: a password and a hardware or software token that supplies a unique, one-time-use, alphanumeric code. Aladdin eToken is a universal serial bus (USB) Smartcard key that provides two-factor authentication to networks and applications. eToken is used to store certificates during Phase 1 of the IP Security (IPSec) authentication, also referred to as Internet Key Exchange (IKE).

10.6. Summary

The overall goal of remote access is granting trusted access to the corporate network over an untrusted network such as the Internet. To secure these remote connections, the AAA model can be

used to secure the corporate network. The AAA model consists of authentication, authorization, and accounting functions.

10.7. Important Questions:

1. What does AAA stand for, and what is its function?
2. What is authentication used for?
3. What is authorization used for?
4. What is accounting used for?
5. What are the three types of authentication servers supported by Cisco IOS?
6. List three characteristics of the TACACS+ protocol.
7. List three characteristics of the RADIUS protocol.
8. What Cisco IOS command is used to enable AAA on a router?
9. What is the Cisco IOS lock-and-key feature?
10. Give an example of two-factor identification.

Chapter 11
Virtual Private Networks

11.1. Introduction:

A virtual private network (VPN) is a service that offers a secure, reliable connection over a shared public infrastructure such as the Internet. Cisco defines a VPN as an encrypted connection between private networks over a public network. To date, there are three types of VPNs:

1. Remote access
2. Site-to-site
3. Firewall-based

The remote access VPN solution is shown in Figure 11.1. Telecommuters and mobile phone users use remote access VPNs to work on the corporate network while out of the office.

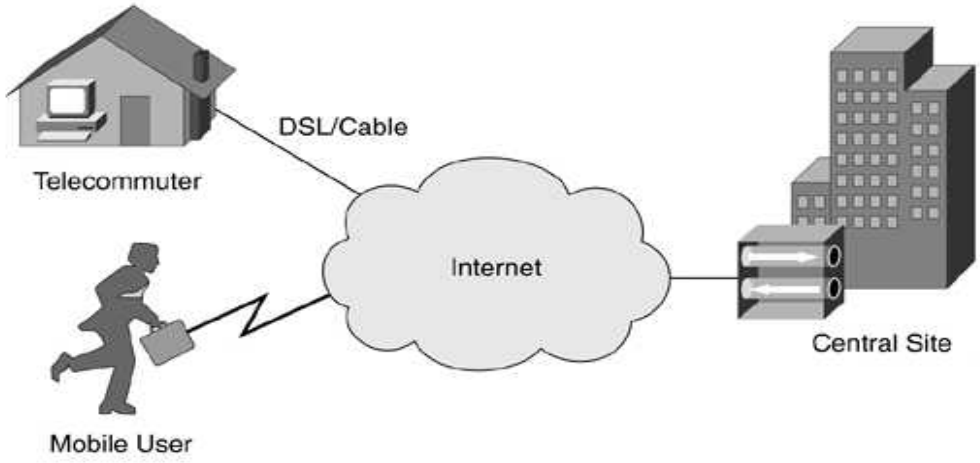


Figure 11.1: Remote Access VPN

In the past, telecommuters and mobile phone users used dial-in connections to access the corporate network, but corporations had to pay for phone lines and the speed was unsatisfactory. Now with the use of VPNs and broadband Internet access, a mobile user can access the corporate site from almost any location, and the speed has greatly improved.

Another VPN solution is site-to-site, as shown in Figure 11.2.

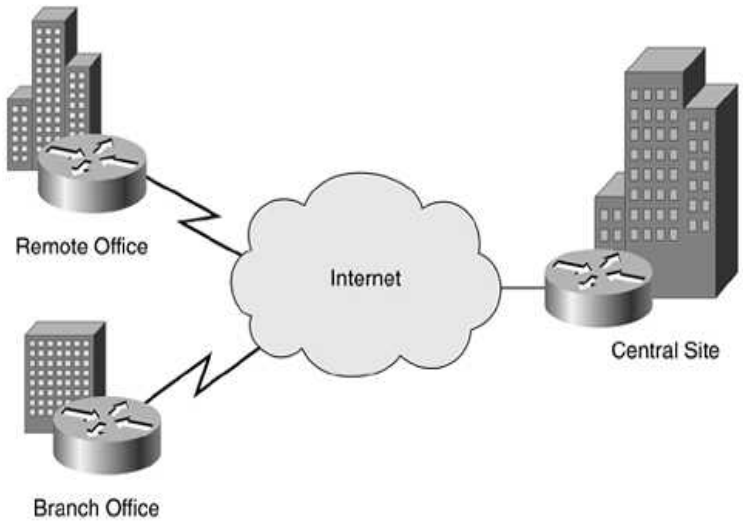


Figure 11.2: Site-to-Site VPN

In the past, leased lines and Frame Relay connections were used to connect different sites. Now, almost all companies have Internet access, so VPNs can be used to connect sites together.

The last available solution is the firewall-based solution. This is almost the same as a site-to-site setup, as you can see in Figure 11.3.

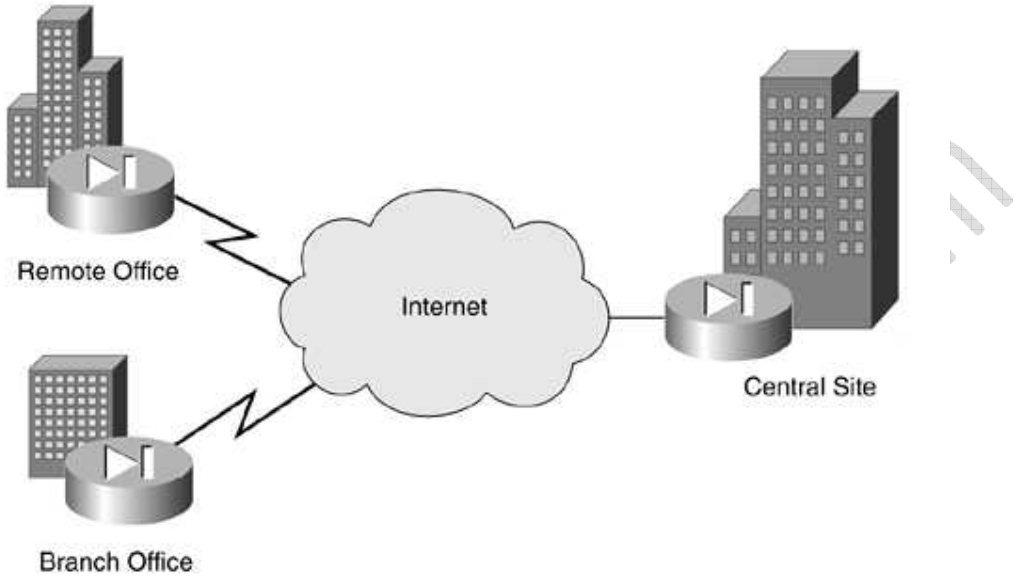


Figure 11.3: Firewall-Based VPN

In a site-to-site setup, the VPN originates on one router and ends on another, whereas in a firewall-based solution, the routers are replaced by firewalls. The difference between the two is not in setup but in security. Typically, this approach is used when corporate security manages the VPN connections because then corporate security is in control of everything.

11.2. Generic Routing Encapsulation Tunnels

Generic Routing Encapsulation (GRE) tunnels are the simplest form of VPNs, and they are very easy to configure. Figure 11.4 shows a GRE tunnel from Router A to Router D. When a packet is sent through the tunnel, it is encapsulated in a GRE packet, so Router B and Router C do not see the original packet.

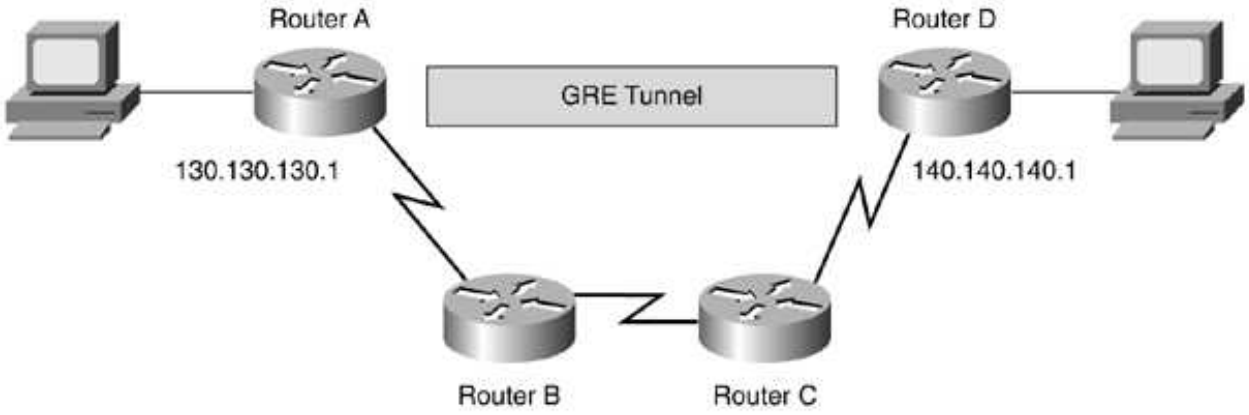


Figure 11.4: GRE Tunnels

11.3. IP Security

IPSec is a framework of open standards. It is not bound to any specific encryption or authentication algorithm keying technology. IPSec acts on the network layer, where it protects and authenticates IP packets between participating peers such as firewalls, routers, or concentrators. IPSec security provides four major functions:

Confidentiality: The sender can encrypt the packets before transmitting them across the network. If such a communication is intercepted, it cannot be read by anybody.

Data integrity: The receiver can verify whether the data was changed while traveling the Internet.

Origin authentication: The receiver can authenticate the source of the packet.

Antireplay protection: The receiver can verify that each packet is unique and is not duplicated.

11.3.1.Encryption

When packets are traveling on the Internet, they are vulnerable to eavesdropping. Clear-text messages can be intercepted and read by anybody. Therefore, to keep the data secure, it can be encrypted. For encryption to work, both the sender and the receiver need to know the rules that were used to encrypt the original message. There are two types of encryption:

- **Symmetric**
- **Asymmetric**

With symmetric key encryption, each peer uses the same key to encrypt and decrypt data. With asymmetric key encryption, each peer uses a different key to encrypt and decrypt the message. Both the Data Encryption Standard (DES) and Triple DES (3DES) require a symmetric shared secret key. The problem is then to give those keys to both users. The keys can be sent by mail, courier, or public key exchange. The easiest method to exchange the key is Diffie-Hellman public key exchange. This key exchange provides a way for the users to establish a shared secret key, which only they know, although they are sending it over an insecure channel.

Public key cryptosystems rely on a two-key system:

A public key, which is exchanged between the users

A private key, which is kept secret by the owners

The Diffie-Hellman public key algorithm states that if user A and user B exchange public keys and combine them with their private keys, the end result should be the same. This is shown in Figure 11.5.

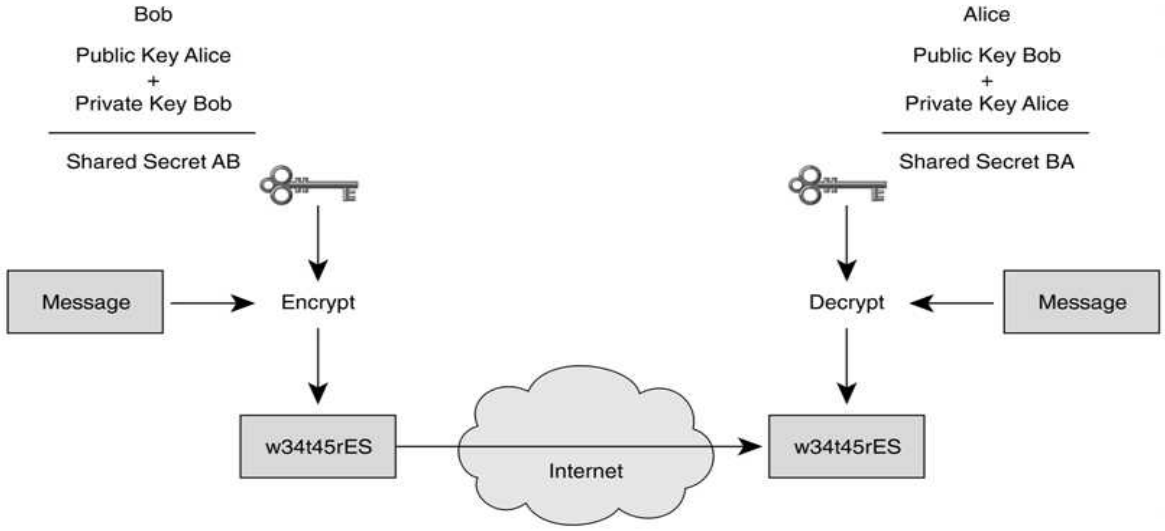


Figure 11.5: Diffie-Hellman Key Exchange

Figure 11.5 is greatly simplified to ensure that the concept of Diffie-Hellman key exchange is clear. There are different variations to this algorithm, known as DH groups 1 through 7. During tunnel setup, VPN peers negotiate which DH group to use.

Encryption can also be accomplished by using the Rivest, Shamir, and Adelman (RSA) algorithm. The RSA algorithm uses an asymmetric key for encryption and decryption. Each user generates two keys: a private key and a public key. The users keep the private key for themselves and exchange the public key. To send an encrypted message to the other end, the local end encrypts the message by using the remote end's public key and the RSA encryption algorithm. This message is then sent to the other end, where it is decrypted using that site's private key. With RSA encryption, the opposite can also be true. The remote end can encrypt a message using its own private key, and the receiver can decrypt the message using the sender's public key. This RSA encryption technique is used for digital signatures.

11.3.2.Data Integrity

Data integrity is also a critical function of VPN because data is sent over a public network and can be intercepted and modified. To guard against this interception, every message has an attached hash. This hash guarantees the integrity of the message. The receiver checks this by comparing the received hash with the hash it calculates from the message itself. If both values are equal, the message has not been tampered with. However, if there is no match, the receiver knows that the message was altered.

IPSec uses the Hashed Message Authentication Codes (HMAC) protocol to calculate the hash. At the sender's end, the message and the shared key are sent through a hash algorithm, which produces a hash value. Basically, this hash algorithm is a formula used to convert a variable-length message into a fixed-length hash. It is also important to understand that this is a one-way function. A message can produce a hash, but a hash cannot produce the original message. After the hash is calculated, it is sent over the network together with the message. At the other end, the receiver performs the same action. It sends the message and the shared key through the hash algorithm and then compares the two hashes to verify whether they match.

Two HMAC algorithms are commonly used:

HMAC-MD5 This protocol uses a 128-bit shared key. The key and the message are combined to a 128-bit hash.

HMAC-SHA-1 This protocol uses a 160-bit shared key. The length of the hash is 160 bits. This protocol is considered stronger because of the longer key.

11.3.3.Origin Authentication

Another important function is origin authentication. Before the electronic era, a seal or a signature on a letter guaranteed its origin. In the electronic era, a document is signed with the sender's private encryption key. This is also called a digital signature. This signature can be authenticated by decrypting it with the sender's public key. When doing business over a long distance, it is important to know who is at the other side of the phone, fax, and so on. The same is true for VPNs. The devices at the other end of the tunnel must be authenticated before the path is considered secure. There are three peer authentication methods:

Preshared keys: A secret key is entered into each peer manually.

RSA signatures: The exchange of digital certificates authenticates the peers.

RSA encryption nonces: Nonces (a random number generated by the peers) are encrypted and then exchanged between peers. The two nonces are used during the peer authentication process.

11.3.3.1. Preshared Keys

If preshared keys are used, the same key is configured on each IPSec peer. At each end, the preshared keys are combined with other information (device-specific information) to form the authentication key. They are both sent through a hash algorithm to form a hash. Then the hash is sent to the other site, as you can see in Figure 11.6.

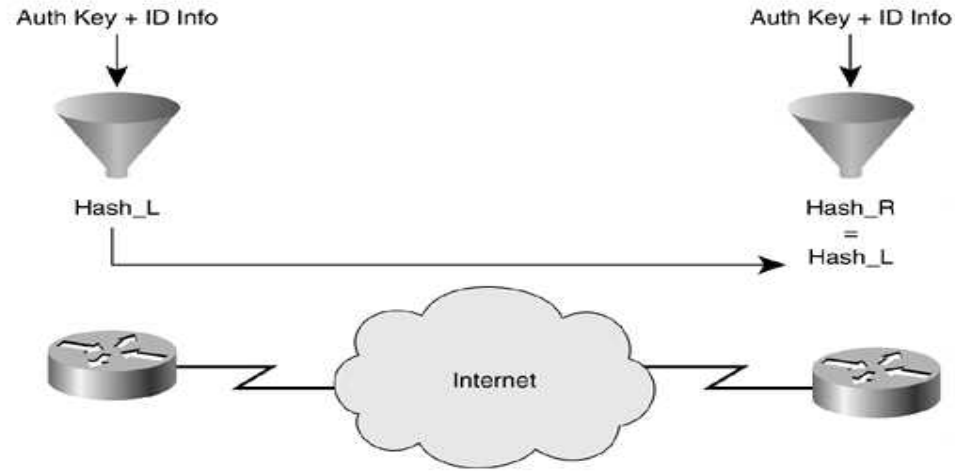


Figure 11.6: Preshared Keys

If the remote peer is able to independently create the same hash, the local peer is authenticated. After that, the authentication process continues in the opposite direction. The remote peer combines its specific information with the preshared key and sends the resulting hash to the local peer. If this peer can create the same hash from its stored information and the preshared key, the remote peer is authenticated. Each peer must authenticate its opposite peer before the tunnel is considered secure. This system with preshared keys is easy to configure manually but does not scale very well. Each IPSec peer must be configured with the preshared key of every other peer with which it wants to communicate.

11.3.3.2. RSA Signatures

With RSA signatures, both hashes are not only authenticated but also digitally signed. Digital certification is discussed in Chapter 13, "Public Key Infrastructure." At the local end, the authentication key and identity information are sent through the hash algorithm to form the hash, a process similar to that used with preshared keys. But with RSA signatures, the hash is then encrypted using the local peer's private key. The result of this procedure is a digital signature, as you can see in Figure 11.7. The digital signature and a digital certificate are both forwarded to the other site. The public encryption key that is also used to decrypt the signature is included in the digital certificate.

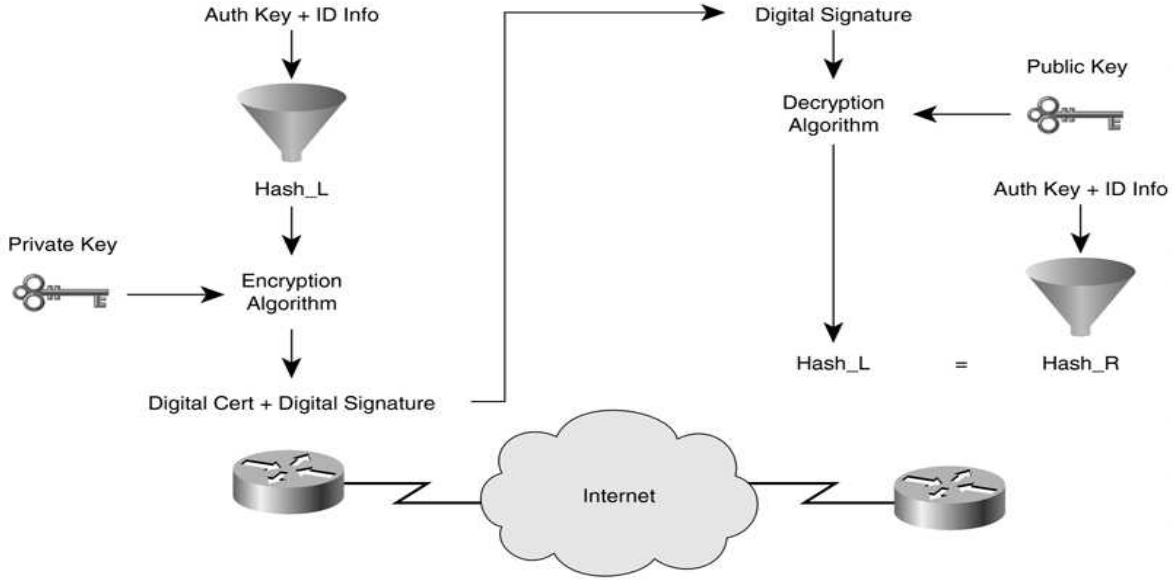


Figure 11.7: RSA Signatures

At the remote peer, the peer authentication is a two-step process. First, the remote site verifies the digital signature by decrypting it with the public key. The result should be the same hash that the local end made. Next, the remote peer independently creates a hash from its stored information and the authentication key, and this also results in a hash. If the hashes are equal, the local peer is authenticated.

After the local peer is authenticated, the process starts all over in the opposite direction. With this kind of authentication, both peers must authenticate their opposite peer before the tunnel is considered secure.

11.3.3.3. RSA-Encrypted Nonces

The word nonce comes from "number used once." RSA-encrypted nonces require that each site generate a nonce. As stated previously, a nonce is a pseudorandom number. The generated nonces are then encrypted and exchanged. When the other side receives the nonces, it makes an authentication key from both nonces and some other information. That nonce-based key is then combined with device-specific information and run though the hash algorithm, as shown in Figure 11.8. After this, the process is similar to that used for RSA signatures.

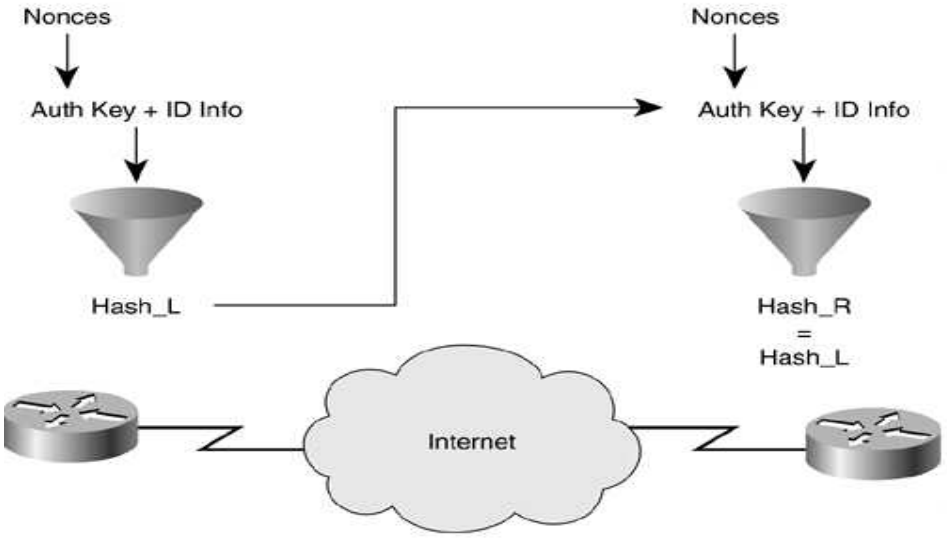


Figure 11.8: RSA-Encrypted Nonces

11.3.4.Antireplay Protection

Antireplay protection verifies that each packet is unique and not duplicated. IPSec packets are protected by comparing the sequence number of the received packets and a sliding window on the destination host. Packets in which the sequence number is before the sliding window are considered late, or duplicate. These packets are dropped.

11.3.5.Protocol Framework

The previous sections discussed encryption, integrity, and authentication. Now let's apply these three concepts to the IPSec protocol suite. IPSec is a framework of open standards. IPSec relies on existing technology, such as DES and 3DES, to secure the communication between two entities. There are two main IPSec framework protocols available:

- Authentication header (AH)
- Encapsulating security payload (ESP)

11.3.5.1. AH

AH is the protocol to use when confidentiality is not required. It provides data authentication and integrity for IP packets between two systems. It verifies that the origin of the packet is correct and that the packet is not modified during transport. It does not encrypt the data packet, so the text is transported in clear text.

Authentication is achieved by using a one-way hash function to create a message digest. The hash is then combined with the text and transmitted to the other site. When the packet reaches its destination, the receiver performs the same one-way hash function and compares the result with the message digest that the sender has supplied. Because the one-way hash uses a symmetric key between the two systems, the authenticity of the packet is guaranteed. The AH function is applied to the entire datagram, except for some header fields that change in transit, such as the Time-To-Live field. The workings of AH are shown in Figure 11.9 and are spelled out in the following steps:

Step 1. The IP header and data payload are hashed.

Step 2. The hash is used to build the AH, which is inserted into the original packet.

Step 3. The modified packet is send to the peer router.

Step 4. The peer router hashes the IP header and data payload.

Step 5. The router extracts the transmitted hash from the AH.

Step 6. The peer router compares the two hashes. The hashes have to match exactly to prove that the packet was not modified during transport.

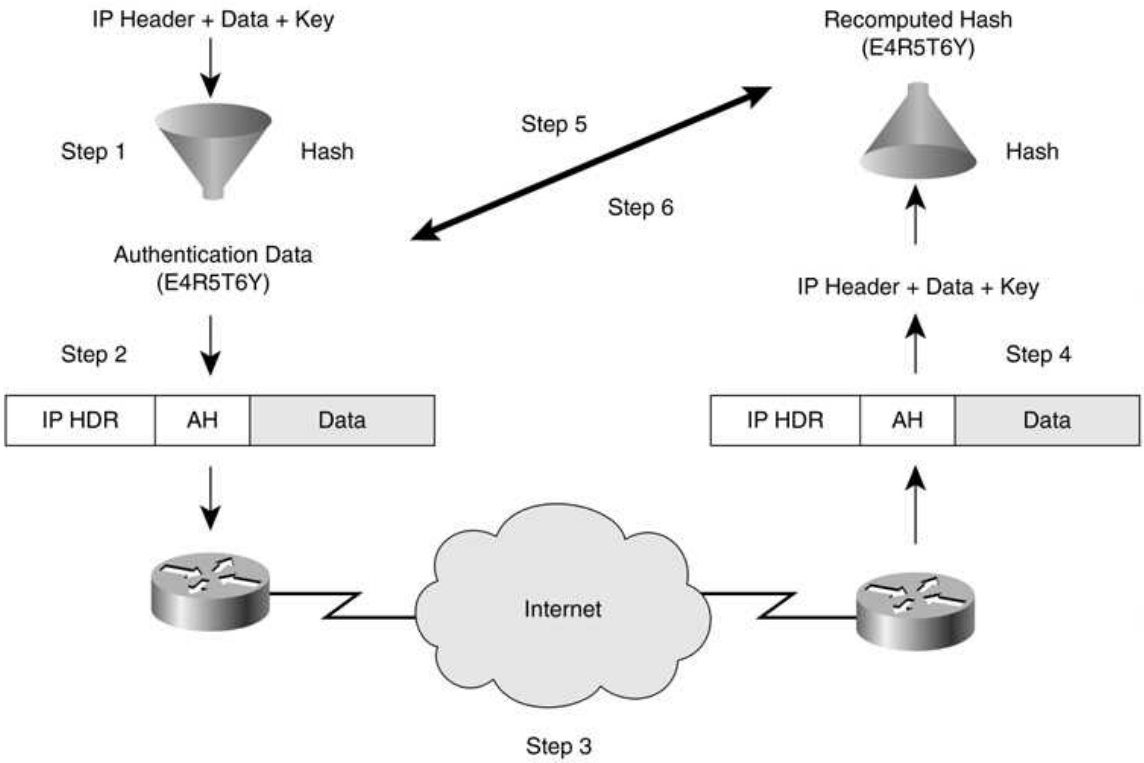


Figure 12-9: AH

11.3.5.2. ESP

ESP can be used to provide encryption and authentication. It provides confidentiality by performing encryption at the IP packet layer. ESP provides authentication for the IP packet payload and the ESP header. As with AH, ESP verifies the following: that the packet originated from where it declares it did, that it is what it declares it is, and that the packet was not modified during transport.

ESP provides confidentiality by encrypting the payload. It supports several symmetric encryption algorithms. The default for IPSec is 56-bit DES, but Cisco products also support 3DES and AES for stronger encryption. ESP can be used alone or in combination with AH. Between two security gateways, the original data is well protected because the entire IP packet is encrypted. An ESP header and trailer are added to the encrypted payload, as shown in Figure 11.10.

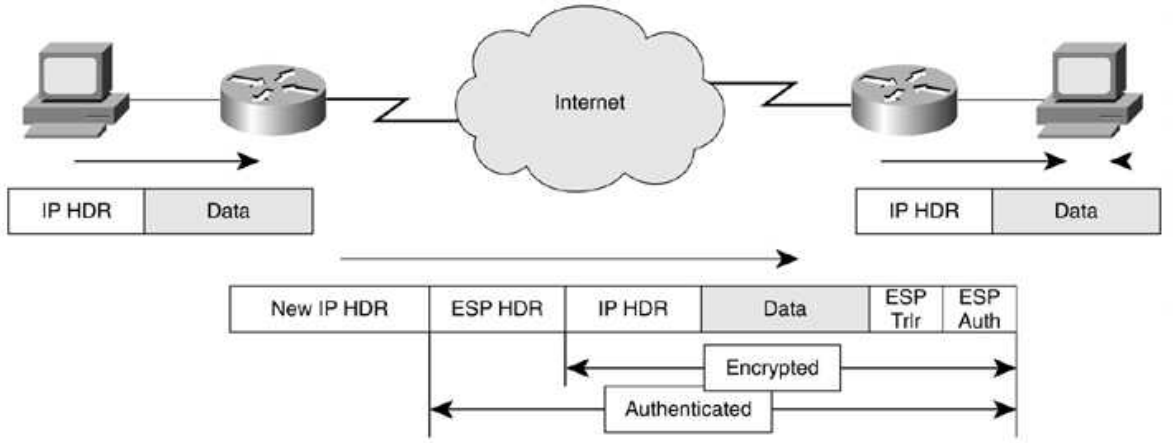


Figure 11.10: ESP

With authentication, the encrypted IP datagram and the ESP header and trailer are included in the hashing process. A new IP header is appended to the front of the packet. This new IP header is used to route the packet through the Internet. When both ESP authentication and encryption are selected, encryption is performed before authentication. One of the main reasons for this order of processing is that it facilitates rapid detection and rejection of incorrect packets at the receiving side. Before decrypting the packet, the receiver can check the authentication of the packets. This requires less processing time and can reduce the impact of denial-of-service (DoS) attacks.

11.3.5.3. Tunnel or Transport Mode

Both ESP and AH can be applied to IP packets in two different ways:

- **Transport mode**
- **Tunnel mode**

These two different modes provide a further level of authentication or encryption support to IPSec. The sections that follow discuss these two IPSec modes in more detail.

- **Transport Mode**

This mode is primarily used for end-to-end connections between hosts or devices acting as hosts. Transport mode protects the payload of the packet but leaves the original IP address readable. This address is used to route a packet through the Internet. Transport mode provides security to the higher layer protocols only. Figure 11.11 shows how transport mode affects AH IPSec connections.

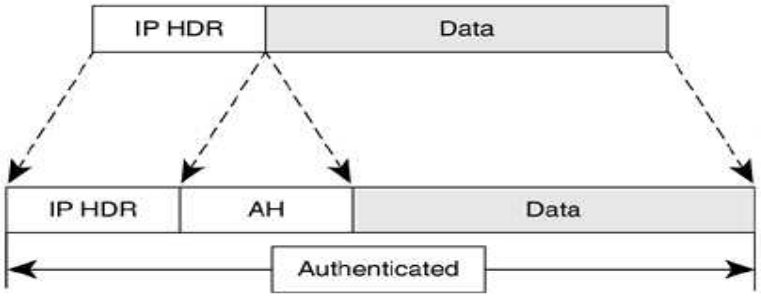


Figure 11.11: AH Transport Mode

The Layer 3 and Layer 4 headers are pried apart, and the AH is added between them.

Figure 11.12 shows ESP transport mode. Again, the IP header is shifted to the left, and the ESP header is inserted. The ESP trailer and ESP authentication are then appended to the end of the packet.

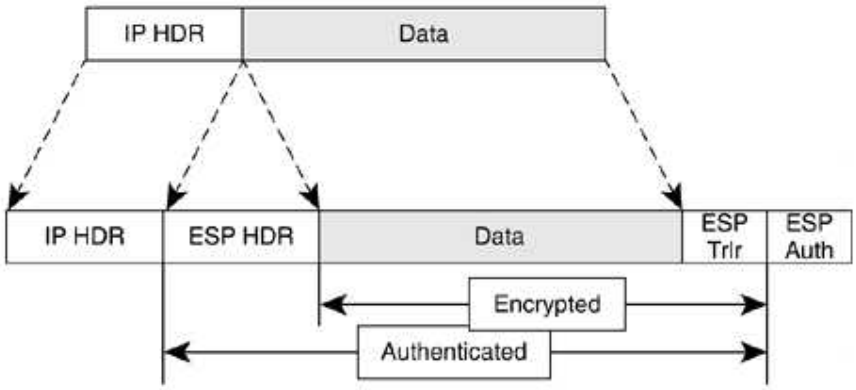


Figure 11.12: ESP Transport Mode

Although the original header remains intact in both situations, the AH transport does not support Network Address Translation (NAT) because changing the source address in the IP header would cause the authentication to fail. If NAT is needed with AH transport mode, make sure that NAT happens before IPSec. ESP transport mode does not have this problem. The IP header remains outside the authentication and encryption area.

- **Tunnel Mode**

IPSec tunnel mode is used between gateways such as routers, PIX firewalls, or VPN concentrators. Tunnel mode is used when the final destination is not a host but a VPN gateway. In this mode, instead of shifting the original IP header to the left and then inserting the IPSec header, the original header is copied and shifted to the left to form a new IP header. The IPSec header is then placed between the new and the original IP headers. The original datagram is left intact. Figure 11.13 shows AH tunnel mode.

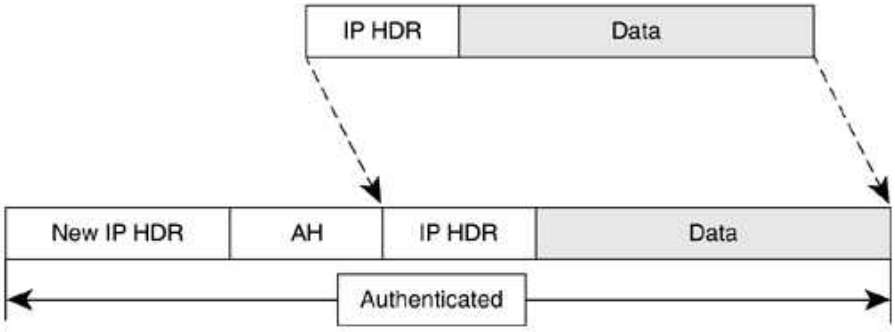


Figure 11.13: AH Tunnel Mode

Also in this mode, notice that the IP header is part of the authentication and that is does not support NAT. In Figure 11.14, you can see a depiction of the ESP tunnel mode. The entire original datagram can be encrypted and authenticated. When both are needed, encryption has to be performed first. This allows authentication to be done with assurance that the sender does not alter the datagram before transmission, and the receiver can authenticate the datagram before decrypting the packet. ESP supports NAT in either tunnel or transport mode, and only ESP supports encryption.

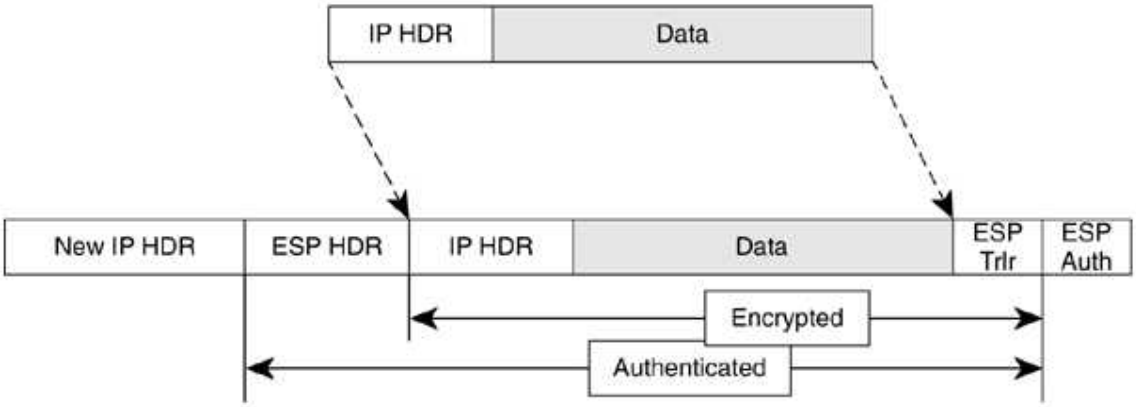


Figure 11.14: ESP Tunnel Mode

- **Transform Sets**

The protocol that brings all the previously mentioned protocols together is the Internet Key Exchange (IKE) protocol. IKE operates in two separate phases when establishing IPSec VPNs.
IKE Phase 1 is responsible for

i. Authenticating the IPSec peers
ii. Negotiating an IKE security association among the peers
iii. Initiating a secure tunnel for IPSec using the Internet Security Association and Key Management Protocol (ISAKMP)

IKE Phase 2 is responsible for

i. Negotiating the set of security parameters for the tunnel
ii. Creating the IPSec tunnel

Configuring IPSec on a Cisco router is fairly simple. You need to identify some parameters for IKE Phase 1, such as:

- Encryption algorithm 56-bit DES or the stronger 168-bit 3DES
- Hash algorithm MD5 or SHA-1
- Authentication method Preshared keys, RSA digital signatures, or RSA encrypted nonces
- Key exchange method 768-bit Diffie-Hellman group 1 or 1024-bit Diffie-Hellman group 2
- IKE SA lifetime 86,400 seconds or 1 day

These parameters need to be identical on both sides, or the connection will not be established. Once these are configured, the only other values you need to supply to establish the IPSec tunnel in IKE Phase 2 mode are as follows:

- IPSec protocol AH and/or ESP
- Hash algorithm MD5 or SHA-1
- Encryption algorithm for ESP DES or 3DES

To make the configuration process easier, the IPSec parameters are already grouped into some predefined configurations called transform sets. The transform sets identify the IPSec protocol, hash algorithm, and when needed, the encryption algorithm. The following transform sets are available, as shown in Table 11.1.

Table 11.1: Transform Sets

Type	Transform	Description
AH authentication	ah-md5-hmac	IPSec AH protocol using HMAC-MD5 for message integrity.
	ah-sha-hmac	IPSec AH protocol using HMAC-SHA-1 for message integrity.
	ah-rfc1828	IPSec AH protocol using MD5 for message integrity. This transform set is used to support older RFC 1828 IPSec implementations.
ESP encryption	esp-des	IPSec ESP protocol using DES encryption.
	esp-3des	IPSec ESP protocol using 3DES encryption.
	esp-null	IPSec ESP protocol with no encryption.
	esp-rfc1829	IPSec ESP protocol using DES-CBC encryption. For older RFC 1829 implementation.
ESP authentication	esp-md5-hmac	IPSec ESP protocol using HMAC-MD5 for message integrity.
	esp-sha-hmac	IPSec ESP protocol using HMAC-SHA-1 for message integrity.

11.4. VPNs with IPSec

As you noticed in the previous discussion, IPSec can use a robust set of protocols and processes. You can use them without knowing much about the protocols, but good practice dictates some preparation steps that need to be taken care of before you can effectively configure a device with IPSec. These steps can be organized as follows:

Step 1. Establish an IKE policy This policy must be identical on both sides of the VPN. The following elements go into an IKE policy:

- Key distribution method Manual or certificate authority. This is explained in more detail in Chapter 13.

- Authentication method This is mainly determined by the key distribution method you have selected. Manual distribution uses preshared keys, whereas certificate authority distribution uses RSA encrypted nonces or RSA signatures.

- IP address or hostnames of peers

Step 2. Establish an IPSec policy Only certain traffic has to go through the IPSec tunnel. Of course, you can decide to send all traffic between peers through that tunnel, but there is a significant performance penalty when using IPSec. It is better to be selective. As in step 1, both peers need to have the same IPSec policies. The following information is needed for an IPSec policy:

- IPSec protocol AH and/or ESP

- Authentication MD5 or SHA-1

- Encryption DES, 3DES, or AES

- Transform set One of the transform sets available in Table 11.1

- Identify traffic Identification of traffic to be sent through the tunnel; specify the protocol, source, destination, and port

- SA establishment

Step 3. Examine the configuration as it is at this stage Check your devices to avoid conflicts with existing settings on one of the devices.

Step 4. Test the network before IPSec Check whether you can ping the peers that are going to participate in IPSec. If you cannot ping them, you must fix this before you can configure IPSec.

Step 5. Permit IPSec ports and protocols If there are access lists enabled on the devices along the path of the VPN, make sure that those devices permit the IPSec traffic.

After completing these steps, you can begin the configuration process. You can think of configuring IPSec as the following five-step process:

Step 1. Interesting traffic initiates the setup of an IPSec tunnel.

Step 2. IKE Phase 1 authenticates peers and establishes a secure tunnel for IPSec negotiation.

Step 3. IKE Phase 2 completes the IPSec negotiation and establishes the tunnel.

Step 4. Secure VPN communication can occur.

Step 5. When there is no traffic to use IPSec, the tunnel is torn down, either explicitly or because the security association (SA) timed out.

11.5. Case Study: Remote Access VPN

This case study translates some of the material covered in this chapter into a real-life scenario. The same Company XYZ is used for this scenario as in previous chapters, and the topology of that company is shown in Figure 11.15.

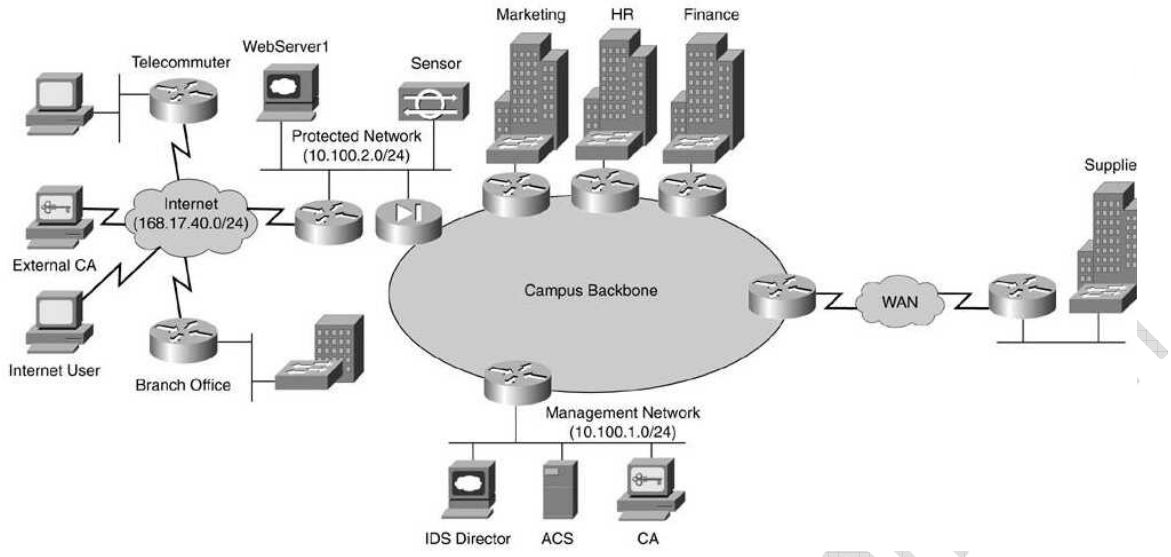


Figure 11.15: XYZ Topology

The whole topology from Figure 11.15 is not used in this scenarioonly a small part. The part that is useful for this case study is shown in Figure 11.16.

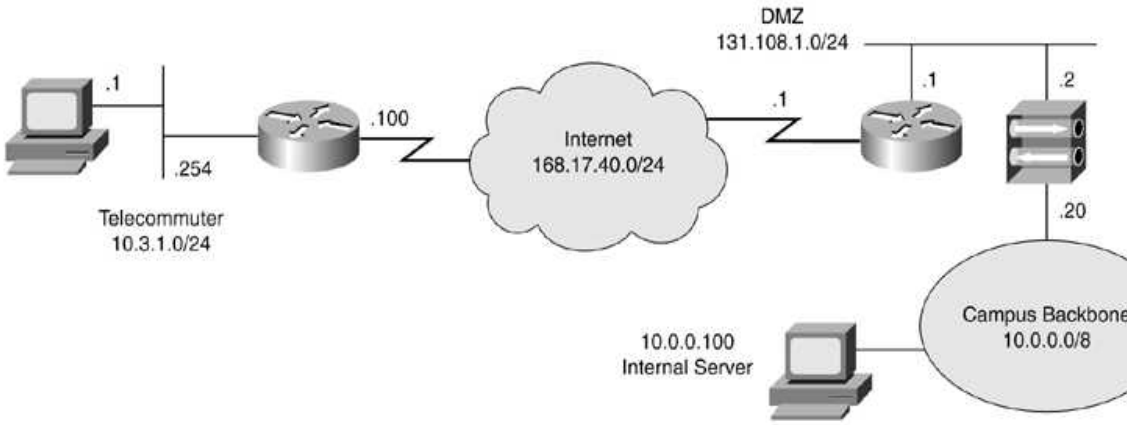


Figure 11.16. Remote Access VPNs

In Figure 11.16, you can see a telecommuter who is connecting to the corporate backbone via a VPN client on a PC. In this case study, the telecommuter is configured to use preshared keys. It is easier to configure the VPN 3000 Concentrator Series for remote access using preshared keys. The alternative method is to use a certificate authority (CA), which is explained in more detail in Chapter 13. Using preshared keys, the client needs to know only the address of the concentrator and the shared secret key. Although VPN configuration is relatively easy with preshared keys, this manual process does not scale well for large implementations. For now, try to configure the concentrator to use preshared keys.

For the initial part of the configuration, you need to attach a console cable to configure the private address of this device. Once the private interface is configured, you can access the concentrator from a workstation using a web browser. The concentrator enters into quick configuration mode the first time it is powered up. After the system has performed the boot functions, you should see the login prompt. When prompted, supply the default login name of admin and the default password, which is also admin. After you run through the menus and

you have configured the private interface (in this case, with address 10.0.0.20), you can access the concentrator from the server (10.0.0.100).

When the browser connects to the concentrator, you see the initial login screen, as shown in Figure 11.17.

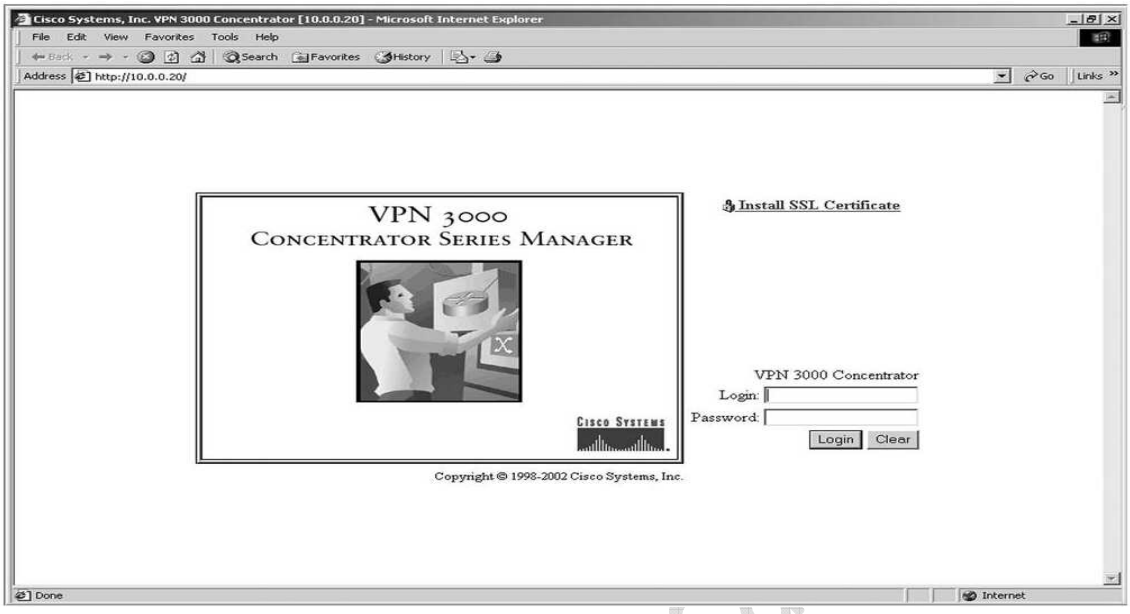


Figure 11.17: Concentrator Login Screen

To continue with the configuration that you started from the command-line interface (CLI), you have to log in with the same login and password you used before. After the VPN Concentrator has accepted your administration login, the screen shown in Figure 11.18 is displayed in your browser window.

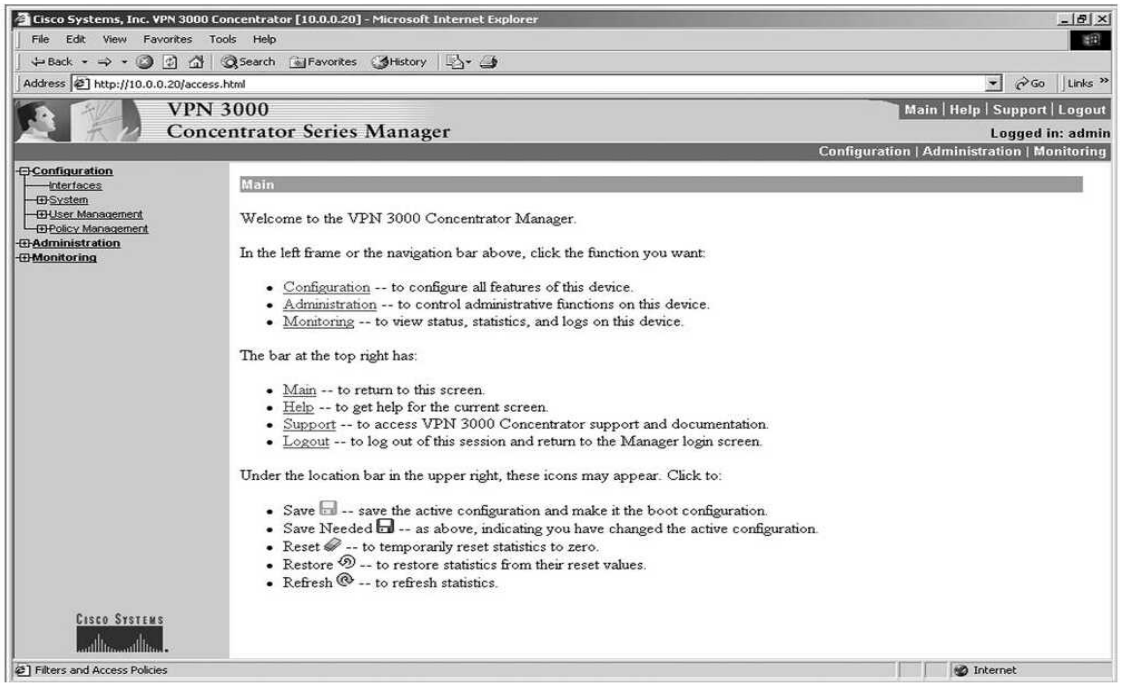


Figure 11.18: Concentrator Main Screen

Figure 11.18 shows Configuration, Administration, and Monitoring in the upper-left corner. These three keys are the primary navigation tools for the daily VPN manager

functions. To proceed with the case study, you have to click the word Interfaces that appears under Configuration. On the screen that displays, select Interface 2. This is the public interface, which brings you to the screen shown in Figure 11.19.

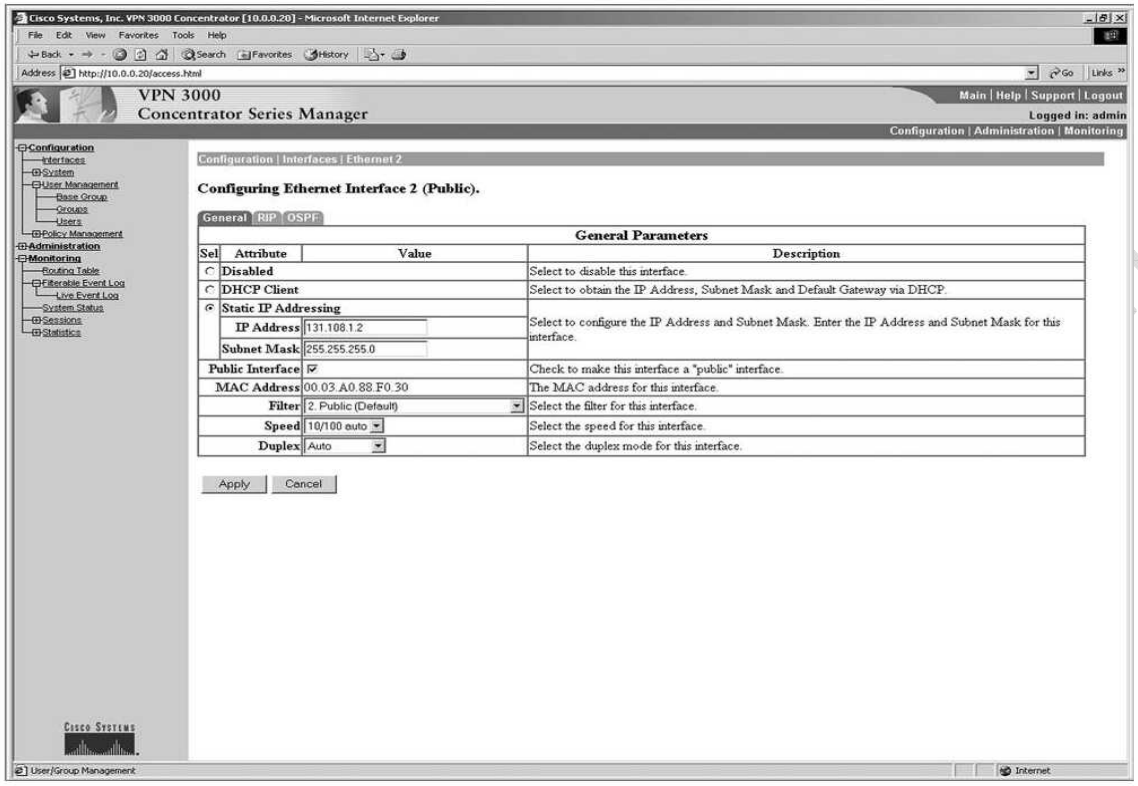


Figure 11.19: Concentrator Interface Screen

On this screen, you can disable the interface, make it a Dynamic Host Configuration Protocol (DHCP) client, or give it a static IP address. For this example, you are using a static IP address (131.108.1.2). You can also set the speed and the mode of the interface on that screen. They are left to default for this example. As a filter, select the default public filter, which is all you have to configure for the public interface. Now you have to perform the same steps for the private interface.

Once the interfaces are configured, you have to add a group and a user to the concentrator. To do this, click User Management under Configuration. Select Groups because you have to define a group before you can add users to that group. This is shown in Figure 11.20.

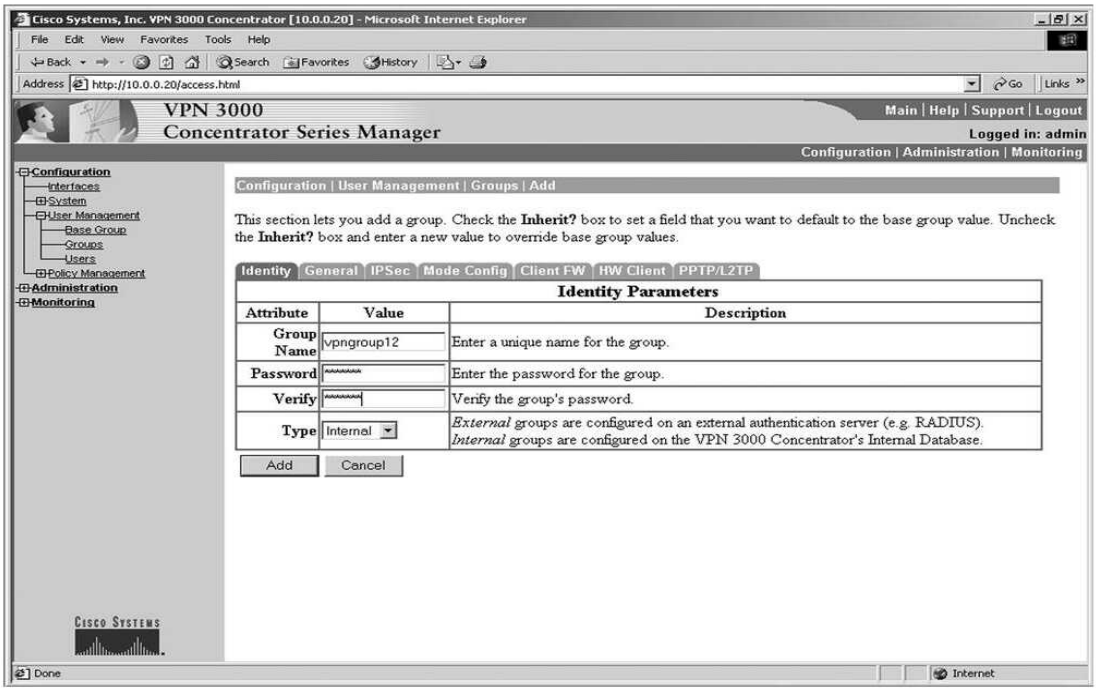


Figure 11.20. Concentrator Group Screen

As you can see, the Groups page has several tabs:

- Identity
- General
- IPSec
- Mode Config
- Client FW
- HW Client
- PPTP/L2TP

For this case study, you are concerned only with Identity, General, and IPSec. On the Identity screen, you have to enter a group name (in this case, the name is vpngroup12) and a password.

That password is also the shared key that the client uses to log in to the concentrator. You also have to define the type of authentication that is used for this group. Users can be authenticated via the following methods:

- RADIUS servers
- NT domain controllers
- Concentrator internal server

In this case study, you use the internal server, so the next step is adding a user to the concentrator internal server. This is done later in the case study. Now that you have defined a group, you can go to the next tab (General) that is shown in Figure 11.21.

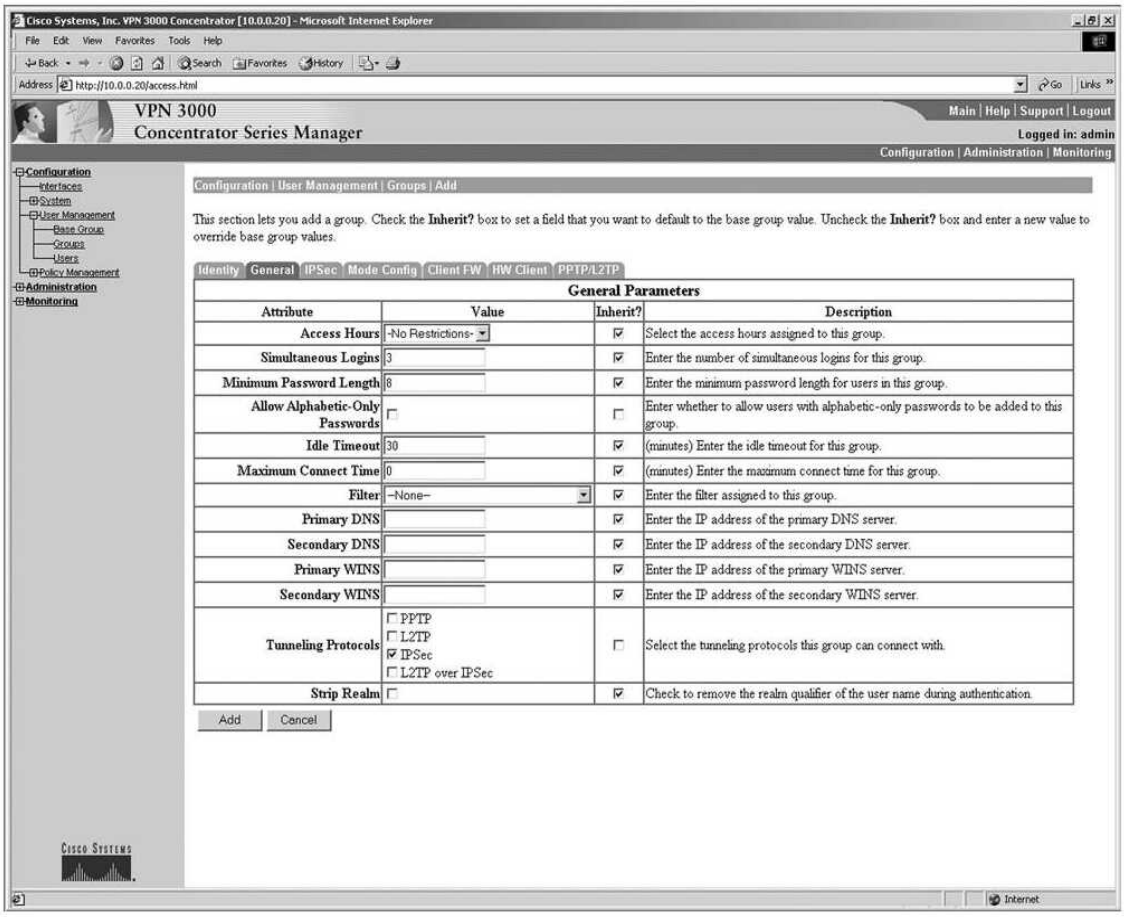


Figure 11.21: Group ScreenGeneral

On this screen, the following information is available:

- Access Hours Selected from the drop-down menu, this attribute determines when the concentrator is open for business for this group. It is currently set to No Restrictions, but you could also select Never, Business Hours (9 a.m. to 5 p.m., Monday through Friday), or a named access hour range that you created elsewhere in the VPN Manager.
- Simultaneous Logins The default is 3, and the minimum is 0. There is no upper limit, but security and prudence would suggest that you limit this value to 1.
- Minimum Password Length The allowable range is 1 to 32 characters. A value of 8 provides a good level of security for most applications.
- Allow Alphabetic-Only Passwords Notice that the Inherit? box has been unchecked. The default is to allow alphabetic-only passwords, which is a security risk. This value has been modified.
- Idle Timeout 30 minutes is a good value here. The minimum allowable value is 1, and the maximum is a value that equates to more than 4000 years. Zero disables idle timeout.
- Maximum Connect Time Zero disables maximum connect time. The range here is

again 1 minute to more than 4000 years.

- Filter Filters determine the "interesting traffic" that uses IPSec. There are three default filters: Public, Private, and External. You can select from those or from any that you may define in the drop-down box. The option None permits all traffic to be handled by IPSec.
- Primary/Secondary DNS/WINS These have been modified from the base groups default settings.
- SEP Card Assignment Some models of the VPN Concentrator can contain up to four Scalable Encryption Processing (SEP) modules that handle encryption functions. This attribute allows you to steer the IPSec traffic for this group to specific SEPs in order to perform your own load balancing. SEP Card Assignment is only visible when there is a SEP card in the concentrator.
- Tunneling Protocols IPSec has been selected, but you could allow the group to use PPTP, L2TP, and L2TP over IPSec as well.
- Strip Realm The default operation of the VPN Concentrator verifies users against the internal database using a combination of the username and realm qualifier, as in username@group. The @group portion is called the realm. You can have the VPN Concentrator use the name only by checking the value for this attribute.

When you have completed these steps, you can move on to the next screen, shown in Figure 11.22, where all IPSec parameters can be configured.

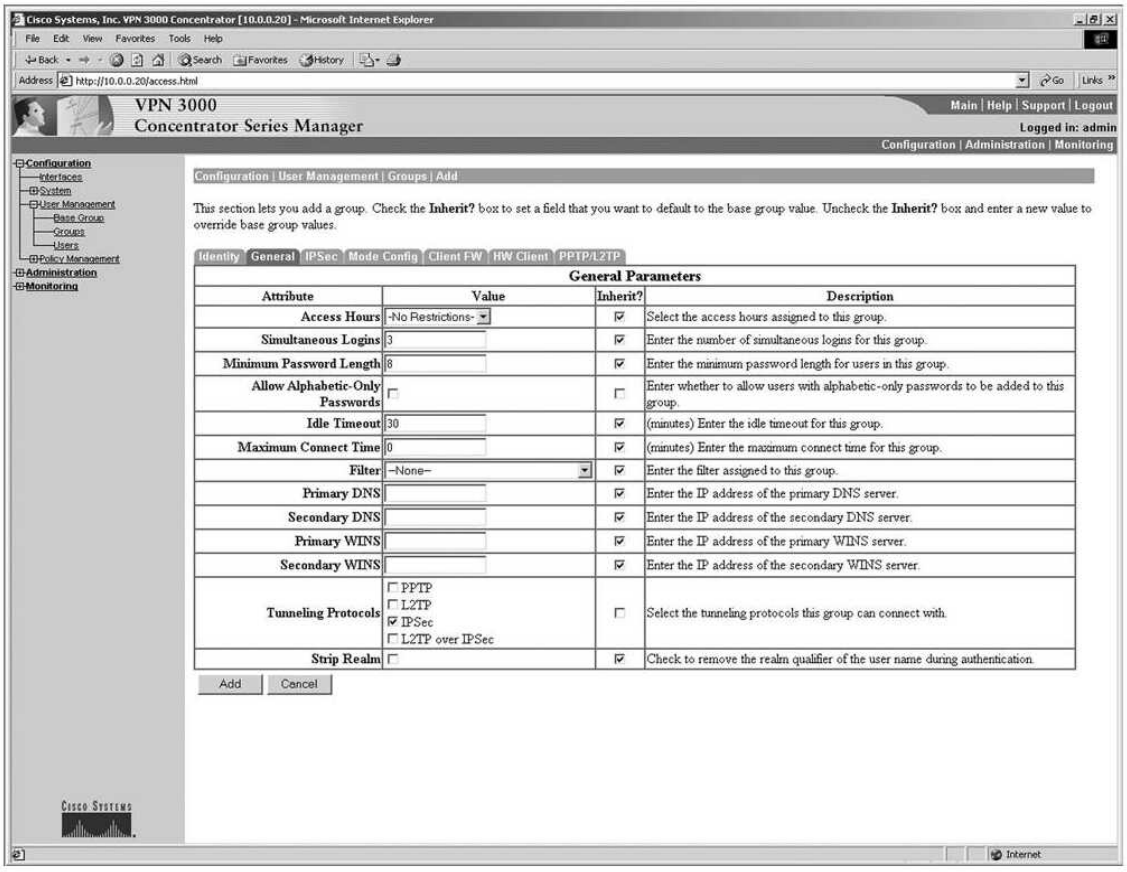


Figure 11.22: Group ScreenIPSec

On this screen, the following attributes can be configured:

- IPSec SA For remote access clients, you must select an IPSec Security Association (SA) from this list of available combinations. The client and server negotiate an SA that governs authentication, encryption, encapsulation, key management, and so on based on your selection here.

  These are the default selections supplied by the VPN Concentrator:

  - None No SA assigned.

  - ESP-DES-MD5 This SA uses DES 56-bit data encryption for both the IKE tunnel and IPSec traffic, ESP/MD5/HMAC-128 authentication for IPSec traffic, and MD5/HMAC-128 authentication for the IKE tunnel.

  - ESP-3DES-MD5 This SA uses 3DES 168-bit data encryption and ESP/MD5/HMAC-128 authentication for IPSec traffic, and DES-56 encryption and MD5/HMAC-128 authentication for the IKE tunnel.

  - ESP/IKE-3DES-MD5 This SA uses 3DES 168-bit data encryption for both the IKE tunnel and IPSec traffic, ESP/MD5/HMAC-128 authentication for IPSec traffic, and MD5/HMAC-128 authentication for the IKE tunnel.

  - ESP-3DES-NONE This SA uses 3DES 168-bit data encryption and no authentication for IPSec traffic, and DES-56 encryption and MD5/HMAC-128 authentication for the IKE tunnel.

  - ESP-L2TP-TRANSPORT This SA uses DES 56-bit data encryption and ESP/MD5/HMAC-128 authentication for IPSec traffic (with ESP applied only to the transport layer segment), and it uses 3DES 168-bit data encryption and MD5/HMAC-128 for the IKE tunnel. Use this SA with the L2TP over IPSec tunneling protocol.

  - ESP-3DES-MD5-DH7 This SA uses 3DES 168-bit data encryption and ESP/MD5/HMAC-128 authentication for both IPSec traffic and the IKE tunnel. It uses Diffie-Hellman Group 7 (ECC) to negotiate Perfect Forward Secrecy. This option is intended for use with the movianVPN client, but you can use it with other clients that support Diffie-Hellman Group 7 (ECC).

- IKE Peer Identity Validation This option applies only to VPN tunnel negotiation based on certificates. This field enables you to hold clients to tighter security requirements.
- IKE Keepalives This monitors the continued presence of a remote peer and notifies the remote peer that the concentrator is still active. If a peer no longer responds to the keepalives, the concentrator drops the connection, preventing hung connections that could clutter up the concentrator.
- Tunnel Type You can select either LAN-to-LAN or Remote Access as the tunnel type.

`

If you select LAN-to-LAN, you do not need to complete the remainder of this screen. For this case study, you need to select Remote Access.

- Group Lock Checking this field forces the user to be a member of this group when authenticating to the concentrator.
- Authentication This field selects the method of user authentication to use. The available options are as follows:

  - None No user authentication occurs. Use this with L2TP over IPSec.

  - RADIUS Uses an external RADIUS server for authentication. The server address is configured elsewhere.

  - RADIUS with Expiry Uses an external RADIUS server for authentication. If the user's password has expired, this method gives the user the opportunity to create a new password.

  - NT Domain Uses an external Windows NT Domain system for user authentication.

  - SDI Uses an external RSA Security Inc. SecurID system for user authentication.

  - Internal Uses the internal VPN Concentrator authentication server for user authentication.

- IPComp This option permits the use of the LZS compression algorithm for IP traffic. This could speed up connections for users connecting through low-speed dialup circuits.
- Reauthentication on Rekey During IKE Phase 1, the VPN Concentrator prompts the user to enter an ID and password. When you enable reauthentication, the concentrator prompts for user authentication whenever a rekey occurs, such as when the IKE SA lifetime expires. If the SA lifetime is set too short, this could be an annoyance to your users, but it does provide an additional layer of security.
- Mode Configuration During SA negotiations, this option permits the exchange of configuration parameters with the client. If you want to pass any configuration information to the client, such as Domain Name System (DNS) or Windows Internet Naming Service (WINS) addresses, you need to enable this option. If you check this box, you need to continue on to the Mode Config tab to complete the selection of attributes there.

If these settings are completed as shown in Figure 11.22, the only thing left is to add a user to the concentrator internal server user database. This can be done by clicking Users under User Management. This screen is shown in Figure 11.23.

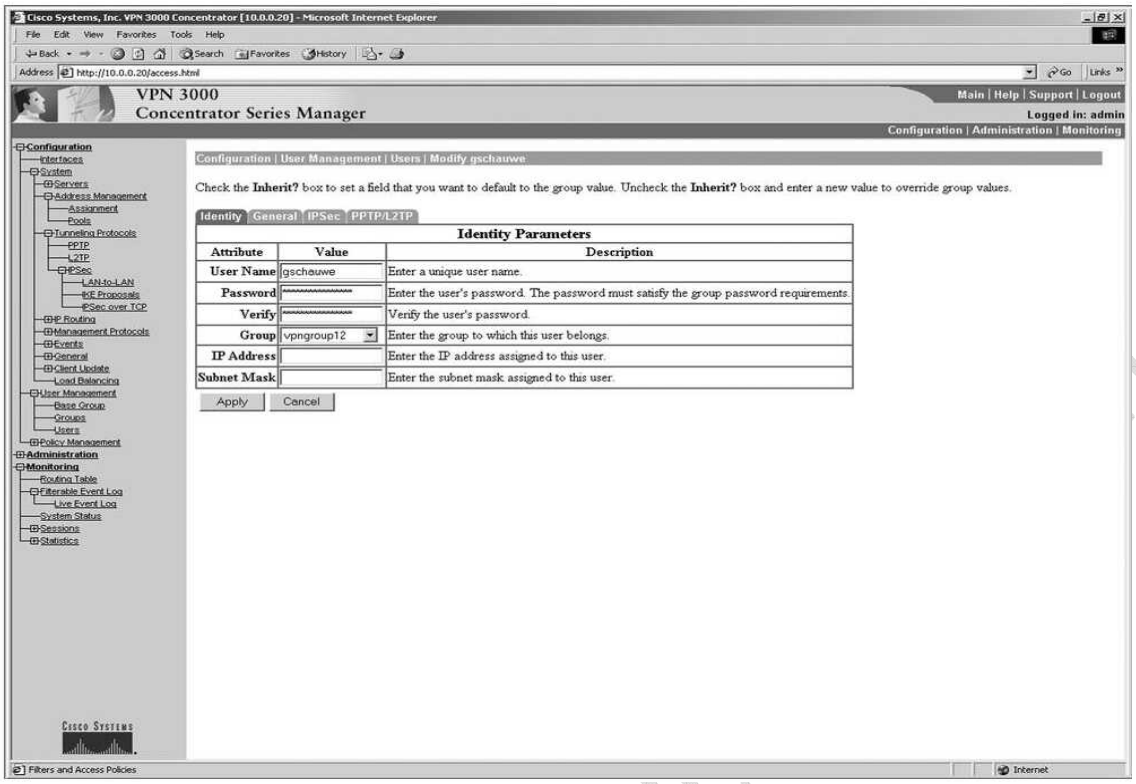


Figure 11.23: Concentrator User Screen

On this screen, add a user gschauwe and a password, and assign that user to the group you previously made. Then click Apply. At that point, the concentrator is ready for use.

The next step in this case study is setting up the VPN client on the telecommuter PC. To do this, start the VPN client by clicking Start > Programs > Cisco Systems VPN Client > VPN Dialer. This brings you to the screen shown in Figure 11.24.

Figure 11.24: VPN Client

On this screen, click New to add a new connection. On the first screen of the wizard, supply a name and a brief description. After you have entered a name and a description, click Next. Figure 11.25 displays the screen that you see.

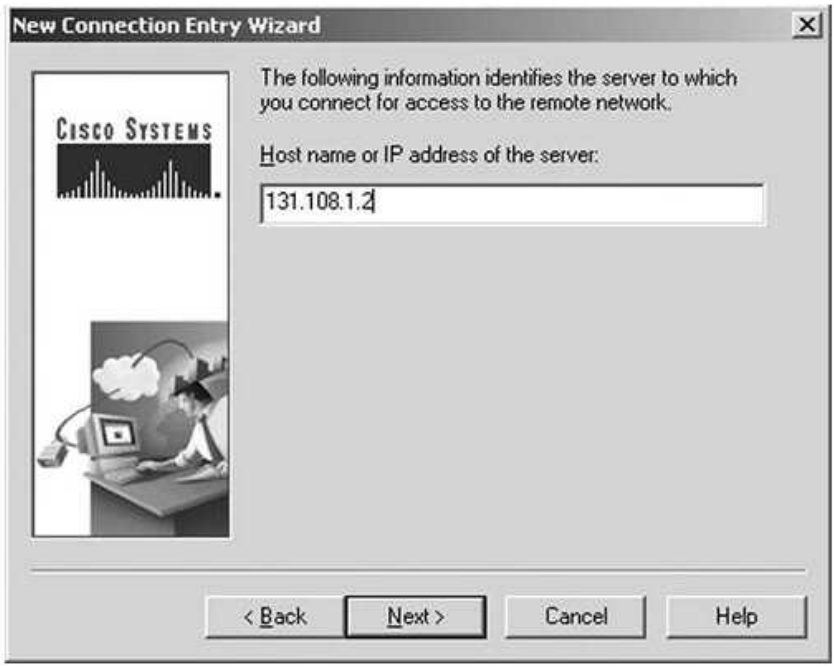


Figure 11.25: VPN ClientSetup Step 1

This screen asks you to identify the VPN server to which you will be connecting. The public address of the VPN concentrator is required, so enter 131.108.1.2 to reach the concentrator you configured earlier. Click Next after you have identified the host server. Figure 11.26 shows the next screen.

Figure 11.26: VPN ClientSetup Step 2

To configure the client to use preshared keys for the IPSec connection, enter the IPSec

group name and password in the appropriate fields of the Group Access Information section. The group name you established earlier was vpngroup12. Click Next and Finish to quit this wizard. Now you are able to connect to the concentrator by clicking Connect on the screen shown in Figure 11.24. This connects you to the VPN Concentrator. After you have established a connection, the concentrator asks you to log in to verify that the correct user is now using the VPN client. After you have entered your username and password, you can access to network behind the VPN concentrator.

11.6. Conclusion

This chapter showed you some methods for making a secure connection from one site to another or from a remote user to the corporate network.

11.7. Important Questions:

1. Name three types of VPN solutions.
2. What are the four major functions of IPSec?
3. Describe the two HMAC algorithms that are commonly used today to provide data integrity.
4. What are the three peer authentication methods used in IPSec?
5. There are two main IPSec framework protocols available. State their names and give a brief explanation of what they do.
6. Both ESP and AH can be applied to IP packets in two different ways. List those two modes and explain the difference between them.
7. List the functions for which IKE Phase 1 is responsible.
8. List the functions for which IKE Phase 2 is responsible.
9. What steps should be completed before configuring a device to use IPSec?
10. Describe briefly how the IPSec process works.