CompTIA

SY0-401 Study Guide

CompTIA Security+

Certifications:

- COMPTIA Security+ SY0-401

The CompTIA Security+ Certification is a vendor neutral credential. The CompTIA Security+ exam is an internationally recognized validation of foundation-level security skills and knowledge, and is used by organizations and security professionals around the globe.

The CompTIA Security+ exam will certify that the successful candidate has the knowledge and skills required to identify risk and participate in risk mitigation activities, provide infrastructure, application, operational and information security, apply security controls to maintain confidentiality, integrity and availability, identify appropriate technologies and products, and operate with an awareness of applicable policies, laws and regulations.

About This Study Guide

The CompTIA Security+ Certification is aimed at an IT security professional who has:

- A minimum of 2 years' experience in IT administration with a focus on security
- Day to day technical information security experience
- Broad knowledge of security concerns and implementation including the topics outlined in Certification Exam Objectives: SY0-401

This study guide provides information formatted to help you study the exam objective wise as described in the SY0-401 objectives. The topics are discussed in detail, which helps you to undertake the exam and be successful in it. It does not represent a complete reference work but is organized around the specific skills tested in the exam. It is focused on discussing the topics covered in SY0-401 rather than COMPTIA Security+.

Topics covered in this study guide are: Network security, compliance and operational security, Threats and vulnerabilities, Application, Data and Host Security, Access Control and Identity Management, and cryptography.

Good Luck!

Contents

Topics 1.0 Network Security 5
Section 1.1- Implement security configuration parameters on network devices and other technologies. 5
Section 1.2 Given a scenario, use secures network administration principles. 22
Section 1.3 - Explain network design elements and components. 29
Section 1.4 -Given a scenario, implement common protocols and services 36
Section 1.5 -Given a scenario, troubleshoot security issues related to wireless networking 59
Topic 2 - Compliance and Operational Security 67
Section 2.1-Explain the importance of risk related concepts 68
Section 2.2 - Summarize the security implications of integrating systems and data with third parties 74
Section 2.3 Given a scenario, implement appropriate risk mitigation strategies. 77
Section 2.4-Given a scenario, implement basic forensic procedures 80
Section 2.5 Summarize common incident response procedures 81
Section 2.6 Explain the importance of security related awareness and training 84
Section 2.7 Compare and contrast physical security and environmental controls 88
Section 2.8 Summarize risk management best practices. 97
Section 2.9 given a scenario; select the appropriate control to meet the goals of security. 101
Topic 3 - Threats and Vulnerabilities 105
Section 3.1 Explain types of malware. 106
Section 3.2- Summarize various types of attacks. 112
Section 3.3- Summarize social engineering attacks and the associated effectiveness with each attack. 119
Section 3.4-Explain types of wireless attacks. 122
Section 3.5- Explain types of application attacks 127
Section 3.6- Analyze a scenario and select the appropriate type of mitigation and deterrent techniques. 131
Section 3.7- Given a scenario, use appropriate tools and techniques to discover security threats and vulnerabilities. 140
Section 3.8 Explain the proper use of penetration testing versus vulnerability scanning. 146
Topic 4.0- Application, Data and Host Security 148
Section 4.1-Explain the importance of application security controls and techniques. 148

Section 4.2- Summarize mobile security concepts and technologies. 152
Section 4.3 Given a scenario, select the appropriate solution to establish host security. 155
Section 4.4- Implement the appropriate controls to ensure data security. 161
Section 4.5- Compare and contrast alternative methods to mitigate security risks in static environments. 165
Topic 5.0 Access Control and Identity Management 167
Section 5.1 Compare and contrast the function and purpose of authentication services. 167
Section 5.2 Given a scenario, select the appropriate authentication, authorization or access control 171
Section 5.3 Install and configure security controls when performing account management, based on best practices. 178
Topic 6.0 Cryptography 188
Section 6.1- Given a scenario, utilize general cryptography concepts. 188
Section 6.2- Given a scenario, use appropriate cryptographic methods. 194
Section 6.3- Given a scenario, use appropriate PKI, certificate management and associated components. 205

Topics 1.0 Network Security

Section 1.1- Implement security configuration parameters on network devices and other technologies.

It's very important to understand that in security, one simply cannot say ``what's the best firewall?" There are two extremes: absolute security and absolute access. The closest we can get to an absolutely secure machine is one unplugged from the network, power supply, locked in a safe, and thrown at the bottom of the ocean. Unfortunately, it isn't terribly useful in this state. A machine with absolute access is extremely convenient to use: it's simply there, and will do whatever you tell it, without questions, authorization, passwords, or any other mechanism. Unfortunately, this isn't terribly practical, either: the Internet is a bad neighborhood now, and it isn't long before some bonehead will tell the computer to do something like self-destruct, after which, it isn't useful to you.

This is no different from our daily lives. We constantly make decisions about what risks we're willing to accept. When we get in a car and drive to work, there's a certain risk that we're taking. It's possible that something completely out of control will cause us to become part of an accident on the highway. When we get on an airplane, we're accepting the level of risk involved as the price of convenience. However, most people have a mental picture of what an acceptable risk is, and won't go beyond that in most circumstances. If I happen to be upstairs at home, and want to leave for work, I'm not going to jump out the window. Yes, it would be more convenient, but the risk of injury outweighs the advantage of convenience.

Every organization needs to decide for itself where between the two extremes of total security and total access they need to be. A policy needs to articulate this, and then define how that will be enforced with practices and such. Everything that is done in the name of security, then, must enforce that policy uniformly.

Firewall

A *firewall* can be hardware, software, or a combination whose purpose is to enforce a set of network security policies across network connections. It is much like a wall with a window: the wall serves to keep things out, except those permitted through the window. Network security policies act like the glass in the window; they permit some things to pass, such as light, while blocking others, such as air. The heart of a firewall is the set of security policies that it enforces. Management determines what is allowed in the form of network traffic between devices, and these policies are used to build rule sets for the firewall devices used to filter network traffic across the network.

Security policies are rules that define what traffic is permissible and what traffic is to be blocked or denied. These are not universal rules, and many different sets of rulesare created for a single company with multiple connections. A web server connected to the Internet may be configured to allow traffic only on port 80 for HTTP and have all other ports blocked, for

example. An e-mail server may have only necessary ports for e-mail open, with others blocked. The network firewall can be programmed to block all traffic to the web server except for port 80 traffic, and to block all traffic bound to the mail server except for port 25. In this fashion, the firewall acts as a security filter, enabling control over network traffic, by machine, by port, and in some cases based on application level detail. A key to setting security policies for firewalls is the same as has been seen for other security policies—the principle of least access. Allow only the necessary access for a function; block or deny all unneeded functionality. How a firm deploys its firewalls determines what is needed for security policies for each firewall.

How Do Firewalls Work?
Firewalls enforce the established security policies through a variety of mechanisms, including the following:

- Network Address Translation (NAT)
- Basic packet filtering
- State-ful packet filtering
- ACLs
- Application layer proxies

One of the most basic security functions provided by a firewall is NAT, which allows you to mask significant amounts of information from outside of the network. This allows an outside entity to communicate with an entity inside the firewall without truly knowing its address. NAT is a technique used in IPv4 to link private IP addresses to public ones. Private IP addresses are sets of IP addresses that can be used by anyone and by definition are not routable across the Internet. NAT can assist in security by preventing direct access to devices from outside the firm, without first having the address changed at a NAT device. The benefit is less public IP addresses are needed, and from a security point of view the internal address structure is not known to the outside world. If a hacker attacks the source address, he is simply attacking the NAT device, not the actual sender of the packet.

NAT was conceived to resolve an address shortage associated with IPv4 and is considered by many to be unnecessary for IPv6. The added security features of enforcing traffic translation and hiding internal network details from direct outside connections will give NAT life well into the IPv6 timeframe.

Basic packet filtering, the next most common firewall technique, involves looking at packets, their ports, protocols, source and destination addresses, and checking that information against the rules configured on the firewall. Telnet and FTP connections may be prohibited from being established to a mail or database server, but they may be allowed for the respective service servers. This is a fairly simple method of filtering based on information in each packet header, such as IP addresses andTCP/UDP ports. Packet filtering will not detect and catch all undesired packets, but it is fast and efficient.

Wireless
Wireless devices bring additional security concerns. There is, by definition, no physical connection to a wireless device; radio waves or infrared carry data, which allows anyone within range access to the data. This means that unless you take specific precautions, you have no control over who can see your data. Placing a wireless device behind a firewall does

not do any good, because the firewall stops only physically connected traffic from reaching the device. Outside traffic can come literally from the parking lot directly to the wireless device.

The point of entry from a wireless device to a wired network is performed at a device called a *wireless access point.* Wireless access points can support multiple concurrent devices accessing network resources through the network node they provide.

Several mechanisms can be used to add wireless functionality to a machine. For PCs, this can be done via an expansion card.

Modems

Modems were once a slow method of remote connection that was used to connect client workstations to remote services over standard telephone lines. *Modem* is a shortened form of *modulator/demodulator,* covering the functions actually performed by the device as it converts analog signals to digital and vice versa. To connect a digital computer signal to the analog telephone line required one of these devices. Today, the use of the term has expanded to cover devices connected to special digital telephone lines—DSL modems—and to cable television lines—cable modems. Although these devices are not actually modems in the true sense of the word, the term has stuck through marketing efforts directed to consumers. DSL and cable modems offer broadband high-speed connections and the opportunity for continuous connections to the Internet. Along with these new desirable characteristics come some undesirable ones, however. Although they both provide the same type of service, cable and DSL modems have some differences. A DSL modem provides a direct connection between a subscriber's computer and an Internet connection at the local telephone company's switching station.

This private connection offers a degree of security, as it does not involve others sharing the circuit. Cable modems are set up in shared arrangements that theoretically could allow a neighbor to sniff a user's cable modem traffic.

Both cable and DSL services are designed for a continuous connection, which brings up the question of IP address life for a client. Although some services originally used static IP arrangement, virtually all have now adopted the Dynamic Host Configuration Protocol (DHCP) to manage their address space. A static IP has an advantage of being the same and enabling convenient DNS connections for outside users. As cable and DSL services are primarily designed for client services asopposed to host services, this is not a relevant issue. A security issue of static IP is that it is a stationary target for hackers. The move to DHCP has not significantly lessened this threat, however, for the typical IP lease on a cable modem DHCP is for days. This is still relatively stationary, and some form of firewall protection needs to be employed by the user.

Cable/DSL Security

The modem equipment provided by the subscription service converts the cable or DSL signal into a standard Ethernet signal that can then be connected to a NIC on the client device. This is still just a direct network connection, with no security device separating the two. The most common security device used in cable/DSL connections is a firewall. The firewall needs to be installed between the cable/DSL modem and client computers.

Telecom/PBX

Private branch exchanges (PBXs) are an extension of the public telephone network into a business. Although typically considered a separate entity from data systems, they are frequently interconnected and have security requirements as part of this interconnection as well as of their own. PBXs are computer-based switching equipment designed to connect telephones into the local phone system. Basically digital switching systems, they can be compromised from the outside and used by phone hackers (preachers) to make phone calls at the business' expense. Although this type of hacking has decreased with lower cost long distance, it has not gone away, and as several firms learn every year, voice mail boxes and PBXs can be compromised and the long-distance bills can get very high, very fast.

Another problem with PBXs arises when they are interconnected to the data systems, either by corporate connection or by rogue modems in the hands of users. In either case, a path exists for connection to outside data networks and the Internet. Just as a firewall is needed for security on data connections, one is needed for these connections as well. Telecommunications firewalls are a distinct type of firewall designed to protect both the PBX and the data connections. The functionality of a telecommunications firewall is the same as that of a data firewall: it is there to enforce security policies.

Telecommunication security policies can be enforced even to cover hours of phone use to prevent unauthorized long-distance usage through the implementation of access codes and/or restricted service hours.

RAS

Remote Access Service (RAS) is a portion of the Windows OS that allows the connection between a client and a server via a dial-up telephone connection. Although slower than cable/DSL connections, this is still a common method for connecting to a remote network. When a user dials into the computer system, authentication and authorization are performed through a series of remote access protocols. For even greater security, a callback system can be employed, where the server calls back to the client at a set telephone number for the data exchange. RAS can also meanRemote Access Server, a term for a server designed to permit remote users access to a network and to regulate their access. A variety of protocols and methods exist to perform this function.

VPN

A virtual private network (VPN) is a construct used to provide a secure communication channel between users across public networks such as the Internet. A variety of techniques can be employed to instantiate a VPN connection.

The use of encryption technologies allows either the data in a packet to be encrypted or the entire packet to be encrypted. If the data is encrypted, the packet header can still be sniffed and observed between source and destination, but the encryption protects the contents of the packet from inspection. If the entire packet is encrypted, it is then placed into another packet and sent via tunnel across the public network. Tunneling can protect even the identity of the communicating parties.

The most common implementation of VPN is via IPsec, a protocol for IP security. IPsec is mandated in IPv6 and is optionally back-fitted into IPv4. IPsec can be implemented in hardware, software, or a combination of both.

Intrusion Detection Systems

Intrusion detection systems (IDSs) are designed to detect, log, and respond to unauthorized network or host use, both in real time and after the fact. IDSs are available from a wide selection of vendors and are an essential part of network security. These systems are implemented in software, but in large systems, dedicated hardware is required as well. IDSs can be divided into two categories: network-based systems and host-based systems. Two primary methods of detection are used: signature-based and anomaly-based.

Network Access Control

Networks comprise connected workstations and servers. Managing security on a network involves managing a wide range of issues, from various connected hardware and the software operating these devices. Assuming that the network is secure, each additional connection involves risk. Managing the endpoints on a case-by-case basis as they connect is a security methodology known as *network access control.* Two main competing methodologies exist: Network Access Protection (NAP) is a Microsoft technology for controlling network access of a computer host, and Network Admission Control (NAC) is Cisco's technology for controlling network admission.

Both the Cisco NAC and Microsoft NAP are in their early stages of implementation. The concept of automated admission checking based on client device characteristics is here to stay, as it provides timely control in the ever-changing network world of today's enterprises.

Network Monitoring/Diagnostic

The computer network itself can be considered a large computer system, with performance and operating issues. Just as a computer needs management, monitoring,and fault resolution, so do networks. SNMP was developed to perform this function across networks. The idea is to enable a central monitoring and control center to maintain, configure, and repair network devices, such as switches and routers, as well as other network services such as firewalls, IDSs, and remote access servers. SNMP has some security limitations, and many vendors have developed software solutions that sit on top of SNMP to provide better security and better management tool suites.

The concept of a network operations center (NOC) comes from the old phone company network days, when central monitoring centers monitored the health of the telephone network and provided interfaces for maintenance and management. This same concept works well with computer networks, and companies with midsize and larger networks employ the same philosophy. The NOC allows operators to observe and interact with the network, using the self-reporting and in some cases self-healing nature of network devices to ensure efficient network operation. Although generally a boring operation under normal conditions, when things start to go wrong, as in the case of a virus or worm attack, the center can become a busy and stressful place as operators attempt to return the system to full efficiency while not interrupting existing traffic.

As networks can be spread out literally around the world, it is not feasible to have a person visit each device for control functions. Software enables controllers at NOCs to measure the actual performance of network devices and make changes to the configuration and operation of devices remotely. The ability to make remote connections with this level of functionality is both a blessing and a security issue. Although this allows efficient network operations management, it also provides an opportunity for unauthorized entry into a network. For this reason, a variety of security controls are used, from secondary networks to VPNs and advanced authentication methods with respect to network control connections.

Routers
Routers are network traffic management devices used to connect different network segments together. Routers operate at the network layer of the OSI model, routing traffic using the network address (typically an IP address) utilizing routing protocols to determine optimal routing paths across a network. Routers form the backbone of the Internet, moving traffic from network to network, inspecting packets from every communication as they move traffic in optimal paths.

Routers operate by examining each packet, looking at the destination address, and using algorithms and tables to determine where to send the packet next. This process of examining the header to determine the next hop can be done in quick fashion. Routers use access control lists (ACLs) as a method of deciding whether a packet is allowed to enter the network. With ACLs, it is also possible to examine the source address and determine whether or not to allow a packet to pass. This allows routers equipped with ACLs to drop packets according to rules built in the ACLs. This can be a cumbersome process to set up and maintain, and as the ACL grows in size, routing efficiency can be decreased. It is also possible to configure some routers to act as quasi–application gateways, performing stasteful packet inspection and using contentsas well as IP addresses to determine whether or not to permit a packet to pass. This can tremendously increase the time for a router to pass traffic and can significantly decrease router throughput.

Switches
Switches form the basis for connections in most Ethernet-based local area networks (LANs). Although hubs and bridges still exist, in today's high-performance network environment switches have replaced both. A switch has separate collision domains for each port. This means that for each port, two collision domains exist: one from the port to the client on the downstream side and one from the switch to the network upstream.

When *full duplex* is employed, collisions are virtually eliminated from the two nodes, host and client. This also acts as a security factor in that a sniffer can see only limited traffic, as opposed to a hub-based system, where a single sniffer can see all of the traffic to and from connected devices.

Switches operate at the data link layer, while routers act at the network layer. For intranets, switches have become what routers are on the Internet—the device of choice for connecting machines. As switches have become the primary network connectivity device, additional functionality has been added to them. A switch is usually a layer 2 device, but layer 3 switches incorporate routing functionality.

Switches can also perform a variety of security functions. Switches work by moving packets from inbound connections to outbound connections. While moving the packets, it is possible to inspect the packet headers and enforce security policies. Port address security based on MAC addresses can determine whether a packet is allowed or blocked from a connection. This is the very function that a firewall uses for its determination, and this same functionality is what allows an 802.1x device to act as an "edge device."

Load Balancers
Network Load Balancing, a clustering technology included in the Microsoft Windows 2000 Advanced Server and Datacenter Server operating systems, enhances the scalability and

availability of mission-critical, TCP/IP-based services, such as Web, Terminal Services, virtual private networking, and streaming media servers. This component runs within cluster hosts as part of the Windows 2000 operating system and requires no dedicated hardware support. To scale performance, Network Load Balancing distributes IP traffic across multiple cluster hosts. It also ensures high availability by detecting host failures and automatically redistributing traffic to the surviving hosts. Network Load Balancing provides remote controllability and supports rolling upgrades from the Windows NT 4.0 operating system.

The unique and fully distributed architecture of Network Load Balancing enables it to deliver very high performance and failover protection, especially in comparison with dispatcher-based load balancers. This white paper describes the key features of thistechnology and explores its internal architecture and performance characteristics in detail.

Internet server programs supporting mission-critical applications such as financial transactions, database access, corporate intranets, and other key functions must run 24 hours a day, seven days a week. And networks need the ability to scale performance to handle large volumes of client requests without creating unwanted delays. For these reasons, clustering is of wide interest to the enterprise. Clustering enables a group of independent servers to be managed as a single system for higher availability, easier manageability, and greater scalability.

The Microsoft Windows 2000 Advanced Server and Datacenter Server operating systems include two clustering technologies designed for this purpose: Cluster service, which is intended primarily to provide failover support for critical line-of-business applications such as databases, messaging systems, and file/print services; and Network Load Balancing, which serves to balance incoming IP traffic among multi-node clusters. We will treat this latter technology in detail here.

Network Load Balancing provides scalability and high availability to enterprise-wide TCP/IP services, such as Web, Terminal Services, proxy, Virtual Private Networking (VPN), and streaming media services. Network Load Balancing brings special value to enterprises deploying TCP/IP services, such as e-commerce applications, that link clients with transaction applications and back-end databases.

Network Load Balancing servers (also called hosts) in a cluster communicate among themselves to provide key benefits, including:

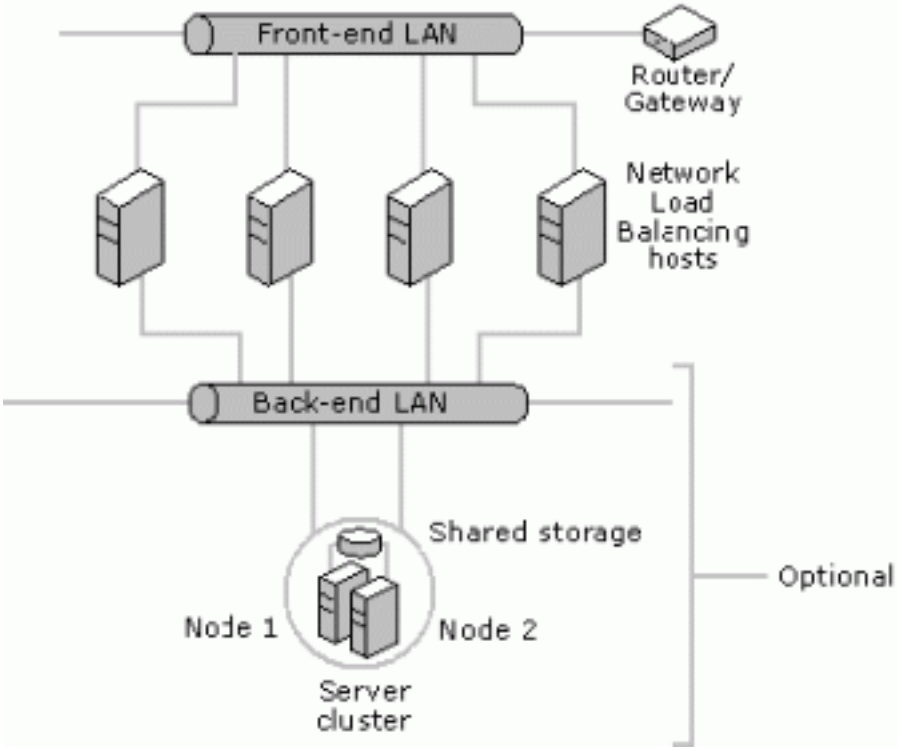
- **Scalability.** Network Load Balancing scales the performance of a server-based program, such as a Web server, by distributing its client requests across multiple servers within the cluster. As traffic increases, additional servers can be added to the cluster, with up to 32 servers possible in any one cluster.

- **High availability.** Network Load Balancing provides high availability by automatically detecting the failure of a server and repartitioning client traffic among the remaining servers within ten seconds, while providing users with continuous service.

Network Load Balancing distributes IP traffic to multiple copies (or instances) of a TCP/IP service, such as a Web server, each running on a host within the cluster. Network Load

Balancing transparently partitions the client requests among the hosts and lets the clients access the cluster using one or more "virtual" IP addresses. From the client's point of view, the cluster appears to be a single server that answers these client requests. As enterprise traffic increases, network administrators can simply plug another server into the cluster.

For example, the clustered hosts in the figure below work together to service network traffic from the Internet. Each server runs a copy of an IP-based service, such asInternet Information Services 5.0 (IIS), and Network Load Balancing distributes the networking workload among them. This speeds up normal processing so that Internet clients see faster turnaround on their requests. For added system availability, the back-end application (a database, for example) may operate on a two-node cluster running Cluster service.



A four-host cluster works as a single virtual server to handle network traffic. Each host runs its own copy of the server with Network Load Balancing distributing the work among the four hosts.

Advantages of Network Load Balancing

Network Load Balancing is superior to other software solutions such as round robin DNS (RRDNS), which distributes workload among multiple servers but does not provide a mechanism for server availability. If a server within the host fails, RRDNS, unlike Network Load Balancing, will continue to send it work until a network administrator detects the failure and removes the server from the DNS address list. This results in service disruption for clients. Network Load Balancing also has advantages over other load balancing solutions—both hardware- and software-based—that introduce single points of failure or performance bottlenecks by using a centralized dispatcher. Because Network Load Balancing has no proprietary hardware requirements, any industry-standard compatible computer can be used. This provides significant cost savings when compared to proprietary hardware load balancing solutions.

The unique and fully distributed software architecture of Network Load Balancing enables it to deliver the industry's best load balancing performance and availability.

Proxy Servers
Though not strictly a security tool, a *proxy server* can be used to filter out undesirable traffic and prevent employees from accessing potentially hostile web sites. A proxy server takes requests from a client system and forwards it to the destination server on behalf of the client. Proxy servers can be completely transparent (these are usually called *gateways* or *tunneling proxies*), or a proxy server can modify the client request before sending it on or even serve the client's request without needing to contact the destination server. Several major categories of proxy servers are in use:

- **Anonymizing proxy** An anonymizing proxy is designed to hide information about the requesting system and make a user's web browsing experience "anonymous." Individuals concerned often use this type of proxy service with the amount of personal information being transferred across the Internet and the use of tracking cookies and other mechanisms to track browsing activity.
- **Caching proxy** This type of proxy keeps local copies of popular client requests and is often used in large organizations to reduce bandwidth usage and increase performance. When a request is made, the proxy server first checks to see whether it has a current copy of the requested content in the cache; if it does, it services the client request immediately without having to contact the destination server. If the content is old or the caching proxy does not have a copy of the requested content, the request is forwarded to the destination server.
- **Content filtering proxy** Content filtering proxies examine each client request and compare it to an established acceptable use policy. Requests can usually be filtered in a variety of ways including the requested URL, destination system, or domain name or by keywords in the content itself. Content filtering proxies typically support user-level authentication so access can be controlled and monitored and activity through the proxy can be logged and analyzed. This type of proxy is very popular in schools, corporate environments, and government networks.
- **Open proxy** An open proxy is essentially a proxy that is available to any Internet user and often has some anonymizing capabilities as well. This type of proxy has been the subject of some controversy with advocates for Internet privacy and freedom on one side of the argument, and law enforcement, corporations, and government entities on the other side. As open proxies are often used to circumvent corporate proxies, many corporations attempt to block the use of open proxies by their employees.
- **Reverse proxy** A reverse proxy is typically installed on the server side of a network connection, often in front of a group of web servers. The reverse proxy intercepts all incoming web requests and can perform a number of functions including traffic filtering, SSL decryption, serving of common static content such as graphics, and performing load balancing.
- **Web proxy** A web proxy is solely designed to handle web traffic and is sometimes called a *web cache*. Most web proxies are essentially specialized caching proxies

Web Security Gateways
If your organization is like most, Web security gateways weren't high on your list of antimalware measures until pretty recently. Your attention to incoming Web traffic has focused largely on policy control--HR concerns over employee access to Internet

pornography, gambling, etc., and productivity, as users spend disproportionate time shopping online and checking up on their stocks and favorite teams.

Anti-malware largely meant anti-virus and was pretty well controlled by email screening and desktop antivirus. While Web security gateways are attracting increased attention, desktop antivirus vendors are scrambling to reinforce their products with improved heuristics, host-based IPS and application controls. The antivirus vendors are responding to the rapidly shifting threats from email-borne viruses to Web-based malware designed to steal confidential data and identities and take control of corporate computers.

The Web security gateway market is an interesting mix of appliance and software vendors, each expanding on their primary strengths--URL filtering vendors like Websense and Secure Computing; traditional AV vendors like McAfee, Trend Micro and Sophos; IM control specialists like FaceTime and email security vendors such as IronPort (recently purchased by Cisco) and MessageLabs--by development, acquisition or partnerships. Newer companies like Mi5 and Anchiva suggest room for growth. (Gartner identifies Blue Coat and Secure Computing as market leaders in a June Magic Quadrant report for this newly defined market.)

Managed Web security gateway services are another option. Although the market is still young, vendors are starting to offer their technology as a service. ScanSafe, the first company to offer antimalware and URL filtering and IM control as pure-play services, actually scans all their customers Web traffic. It OEMs for companies like Postini and AT&T. MessageLabs, which initially sold ScanSafe-based services, now offers managed services based on its own technology.

VPN concentrators

With the Internet, we had the ability to create a VPN, providing a secure connection for users dialing in to their ISP from wherever. As time has passed, the need for greater security over these VPNs has increased. Unfortunately, small businesses usually have a limited amount of funds and/or IT expertise. But that doesn't mean they should ignore the need to secure their VPNs properly. A VPN concentrator -- ideal when you require a single device to handle a large number of incoming VPN tunnels -- may be just what they need.

VPN concentrators typically arrive in one of two architectures: SSL VPNs and IPSec VPNs. Some concentrators only offer support of one protocol or the other, whereas Cisco and other vendors advertise the ability to utilize either with their concentrators.

The traditional tunnel for VPNs relies on IPSec, which resides at the network layer of the OSI model. At this level, a client is considered a virtual member of the connected network and can pretty much access the network as if locally connected. Therein lies a positive aspect of IPSec: Apps run without any awareness that the client is comingfrom outside the network. The drawback is that additional security controls have to be configured to reduce risks.

For a client to access the IPSec VPN, it must have the client-side software configured. While this adds security, it provides additional cost to implement and leads to additional time and energy spent by tech support. This is what leads many toward an SSL solution.

SSL is already built in to the capabilities of pretty much all computers through Web browsers. Thus, there is no additional work to install and configure the client side. In addition, rather than residing at the network layer, allowing access to all aspects of a network,

SSL lets admins allow access a bit more precisely toward applications that are Web-enabled. In addition, admins can establish a finer level of control over uses with SSL VPN connections.

On the negative angle, however, being that you can only utilize SSL VPNs through a Web browser, only Web-based applications will work. With a little bit of work, you can Web-enable additional applications, but this adds to the configuration time and may make SSL an unattractive solution for some.

In addition, SSL applications will not have centralized storage, shared access to resources (like printers), or files and other options that you can achieve through an IPSec connection. Some worry about Web caching with private information being left behind. Thus, you might want to choose a VPN concentrator that lists within its feature sets "automatic cache cleanup after session termination to ensure privacy of data," as the NetGear SSL device does.

Network-based IDSs

Network-based IDSs (NIDS) came along a few years after host-based systems. After running host-based systems for a while, many organizations grew tired of the time, energy, and expense involved with managing the first generation of these systems. The desire for a "better way" grew along with the amount of interconnectivity between systems and consequently the amount of malicious activity coming across the networks themselves.

This fueled development of a new breed of IDS designed to focus on the source for a great deal of the malicious traffic—the network itself.

The NIDS integrated very well into the concept of *perimeter security*. More and more companies began to operate their computer security like a castle or military base with attention and effort focused on securing and controlling the ways in and out—the idea being that if you could restrict and control access at the perimeter, you didn't have to worry as much about activity inside the organization. Even though the idea of a security perimeter is somewhat flawed (many security incidents originate inside the perimeter), it caught on very quickly, as it was easy to understand and devices such as firewalls, bastion hosts, and routers were available to define and secure that perimeter. The best way to secure the perimeter from outside attack is to reject all traffic fromexternal entities, but as this is impossible and impractical to do, security personnel needed a way to let traffic in but still be able to determine whether or not the traffic was malicious. This is the problem that NIDS developers were trying to solve.

Active vs. Passive NIDSs

Most NIDSs can be distinguished by how they examine the traffic and whether or not they interact with that traffic. On a *passive* system, the IDS simply watches the traffic, analyzes it, and generates alarms. It does not interact with the traffic itself in any way, and it does not modify the defensive posture of the system to react to the traffic. A passive IDS is very similar to a simple motion sensor—it generates an alarm when it matches a pattern much as the motion sensor generates an alarm when it sees movement. Active IDS will contain all the same components and capabilities of the passive IDS with one critical addition—the active IDS can *react* to the traffic it is analyzing.

These reactions can range from something simple, such as sending a TCP reset message to interrupt a potential attack and disconnect a session, to something complex, such as

dynamically modifying firewall rules to reject all traffic from specific source IP addresses for the next 24 hours.

Signatures
Signatures can be very simple or remarkably complicated, depending on the activity they are trying to highlight. In general, signatures can be divided into two main groups, depending on what the signature is looking for: context-based and context-based.

Content-based signatures are generally the simplest. They are designed to examine the content of such things as network packets or log entries. Content-based signatures are typically easy to build and look for simple things, such as a certain string of characters or a certain flag set in a TCP packet. Here are some example content-based signatures: • *Matching the characters* /etc/passwd *in a Telnet session.* On a UNIX system, the names of valid user accounts (and sometimes the passwords for those user accounts) are stored in a file called *passwd* located in the *etc* directory.

- *Matching a TCP packet with the synchronize, reset, and urgent flags all set within the same packet.* This combination of flags is impossible to generate undernormal conditions, and the presence of all of these flags in the same packet would indicate this packet was likely created by a potential attacker for aspecific purpose, such as to crash the targeted system.
- *Matching the characters* to: decode *in the header of an e-mail message.* On certain older versions of sendmail, sending an e-mail message to "decode" would cause the system to execute the contents of the e-mail.

Context-based signatures are generally more complicated, as they are designed to match large patterns of activity and examine how certain types of activity fit into the
Other activities going on around them. Context signatures generally address the question "How does this event compare to other events that have already happened or might happen in the near future?" Context-based signatures are more difficult to analyze and take more resources to match, as the IDS must be able to "remember" past events to match certain context signatures. Here are some examples of context-based signatures:

- ***Match a potential intruder scanning for open web servers on a specific network.***A potential intruder may use a port scanner to look for any systems accepting connections on port 80. To match this signature, the IDS must analyze all attempted connections to port 80 and then be able to determine which connection attempts are coming from the same source but are going to multiple, different destinations.
- ***Identify a Nessus scan.***Nessus is an open-source vulnerability scanner that allows security administrators (and potential attackers) to quickly examine systems for vulnerabilities. Depending on the tests chosen, Nessus will typically perform the tests in a certain order, one after the other. To be able to determine the presence of a Nessus scan, the IDS must know which tests Nessus runs as well as the typical order in which the tests are run.
- ***Identify a ping flood attack***. A single ICMP packet on its own is generally regarded as harmless, certainly not worthy of an IDS signature. Yet thousands of ICMP packets coming to a single system in a short period of time can have a devastating effect on the receiving system. By flooding a system with thousands of valid ICMP packets, an attacker can keep a target system so busy it doesn't have time to do anything else—a

very effective denial-of-service attack. To identify a ping flood, the IDS must recognize each ICMP packet and keep track of how many ICMP packets different systems have received in the recent past.

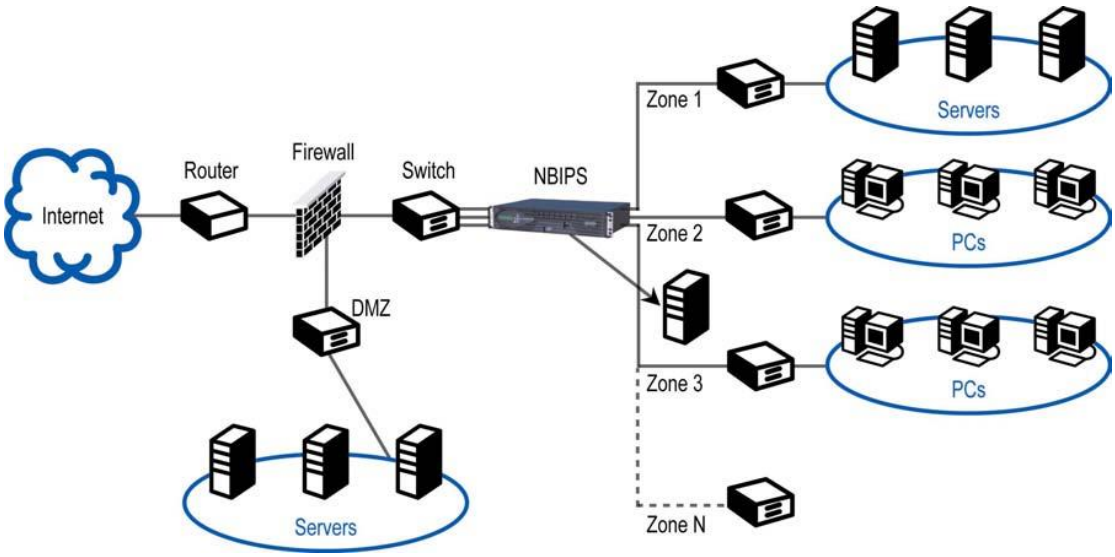
False Positives and Negatives
Viewed in its simplest form, an IDS is really just looking at activity (be it host-based or network-based) and matching it against a predefined set of patterns. When it matches an activity to a specific pattern, the IDS cannot know the true intent behind that activity—whether or not it is benign or hostile—and therefore it can react only as it has been programmed to do. In most cases, this means generating an alert that must then be analyzed by a human who tries to determine the intent of the traffic from whatever information is available.

When AN ID matches a pattern and generates an alarm for benign traffic, meaning the traffic was not hostile and not a threat, this is called a *falsepositive.* In other words, the IDS matched a pattern and raised an alarm when it didn't really need to do so. Keep in mind that the IDS can only match patterns and has no ability to determine intent behind the activity, so in some ways this is an unfair label. Technically, the IDS is functioning correctly by matching the pattern, but from a human standpoint this is not information the analyst needed to see, as it does not constitute a threat and does not require intervention.

NIPS (Network Intrusion Protection System)
The advent of Network-Based Intrusion Prevention heralds a new era of effective and efficient information security for corporations, educational institutions and government agencies. In effect, Network-Based Intrusion Prevention Systems (NBIPS) transform networks from a vulnerable and weak IT element to a tremendously powerful weapon against cyber-terrorism. The network becomes a potent and forceful instrument of protection – continuously defending every resource attached to it. Desktops, servers, operating systems, applications and Web services are aggressively protected from both external and internal attacks by Network-Based Intrusion Prevention Systems.

As well, the cost of securing your information assets declines dramatically with the deployment of Network-Based Intrusion Prevention. These efficient systems continuously filter attacks as they attempt to traverse the network and as a result, no damage occurs and no cleanup is required. Security administration is reduced and system downtime as a result of attack is eliminated.

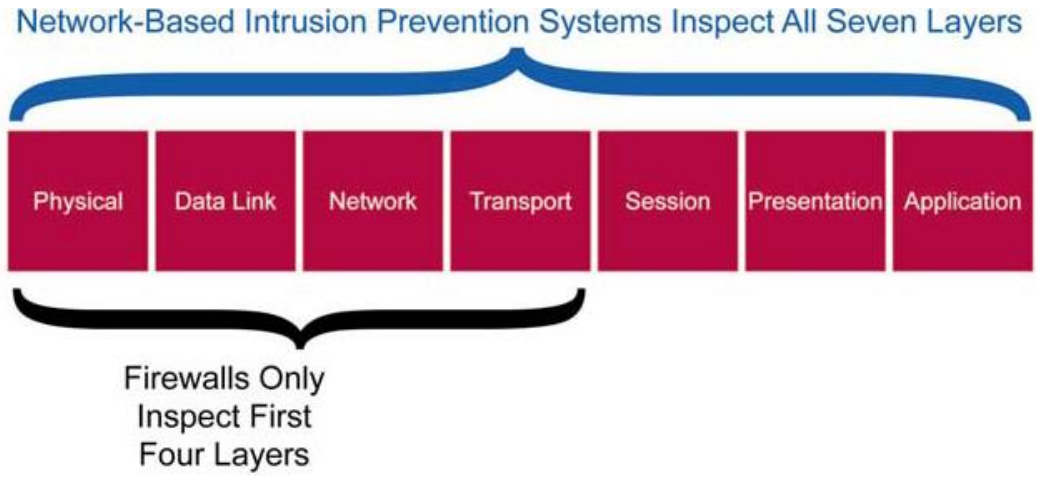
An NBIPS installs in the network and is used to create physical security zones.



In essence, the network becomes intelligent and is able to quickly and precisely discern good traffic from bad traffic. The Intrusion Prevention System becomes a "jail" for hostile traffic such as Worms, Trojans, Viruses, Blended Attacks and Polymorphic Threats.

NBIPS are made possible through the deft blending of high-speed Application Specific Integrated Circuits (ASICS) and newly available Network Processors. Network Processors are very different from microprocessors in that they are specifically designed to process a high-speed flow of network traffic by executing tens of thousands of instructions and comparisons in parallel. A microprocessor, such as the Pentium, was designed as a general-purpose processor for graphics and spreadsheets and only executes one instruction at a time.

Network-Based Intrusion Prevention Systems are an extension of today's Firewall technologies. To some extent, you can think of an NBIPS as a Seven- Layer Firewall. Today's Firewalls inspect only the first four layers of any packet of information flow. NBIPS inspect all 7 Layers, making it impossible to hide anything in the last four layers of a packet



Network-Based Intrusion Prevention Systems portend an immediate future where chaos, anxiety, cost and sweat are replaced with certainty, productivity and profitability. The nature of these systems creates a security posture never before seen and harmonizes the management of all security initiatives. We believe it is incumbent on all organizations, private and public, to deploy NBIPS for the following reasons:

- NBIPS will improve corporate productivity and profitability
- NBIPS will protect sensitive information from being stolen
- NBIPS will protect key infrastructure from imminent global cyber-attacks thus preserving standards of living and ways of life.
- NBIPS will limit copyright infringement liability

Protocol analyzers

A protocol analyzer (also known as a packet sniffer, network analyzer, or network sniffer) is a piece of software or an integrated software/hardware system that can capture and decode network traffic. Protocol analyzers have been popular with system administrators and security professionals for decades because they are such versatile and useful tools for a network environment. From a security perspective, protocol analyzers can be used for a number of activities, such as the following:

- Detecting intrusions or undesirable traffic (IDS/IPS must have some typeof capture and decode ability to be able to look for suspicious traffic)
- Capturing traffic during incident response or incident handling
- Looking for evidence of botnets, Trojans, and infected systems
- Looking for unusual traffic or traffic exceeding certain thresholds
- Testing encryption between systems or applications

From a network administration perspective, protocol analyzers can be used for activities such as these:

- Analyzing network problems
- Detecting misconfigured applications or misbehaving applications
- Gathering and reporting network usage and traffic statistics
- Debugging client/server communications

Regardless of the intended use, a protocol analyzer must be able to see network traffic in order to capture and decode it. A software-based protocol analyzer must be able to place the NIC it is going to use to monitor network traffic in promiscuous mode (sometimes called promiscuous mode). Promiscuous mode tells the NIC to process every network packet it sees regardless of the intended destination. Normally, a NIC will process only broadcast packets (that are going to everyone on that subnet) and packets with the NIC's Media Access Control (MAC) address as the destination address inside the packet. As a sniffer, the analyzer must process every packet crossing the wire, so the ability to place a NIC into promiscuous mode is critical.

Sniffers
The group of protocols that make up the TCP/IP suite was designed to work in a friendly environment where everybody who connected to the network used the protocols as they were designed. The abuse of this friendly assumption is illustrated by network traffic sniffing programs, sometimes referred to as sniffers.

A network sniffer is software or hardware device that is used to observe traffic as it passes through a network on shared broadcast media. The device can be used to view all traffic, or it can target a specific protocol, service, or even string of characters (looking for logins, for example). Normally, the network device that connects a computer to a network is designed to ignore all traffic that is not destined for that computer. Network sniffers ignore this friendly agreement and observe all traffic on the network, whether destined for that computer or others. A network card that is listening to all network traffic and not just its own is said to be in "promiscuous mode." Some network sniffers are designed not just to observe all traffic but to modify traffic as well.

Network administrators for monitoring network performance can use network sniffers. They can be used to perform traffic analysis, for example, to determine what type of traffic is most commonly carried on the network and to determine which
Segments are most active. They can also be used for network bandwidth analysis and to troubleshoot certain problems (such as duplicate MAC addresses).

Spoofing
Spoofing is nothing more than making data look like it has come from a different source. This is possible in TCP/IP because of the friendly assumptions behind the protocols. When the protocols were developed, it was assumed that individuals who had access to the network layer would be privileged users who could be trusted. When a packet is sent from one system to another, it includes not only the destination IP address and port but the source IP address as well. You are supposed to fill in the source with your own address, but nothing stops you from filling in another system's address. This is one of the several forms of spoofing.

Spoofing E-Mail
In e-mail spoofing, a message is sent with a From address that differs from that of the sending system. This can be easily accomplished in several different ways using several programs. To demonstrate how simple it is to spoof an e-mail address, you can Telnet to port 25 (the port associated with e-mail) on a mail server. From there, you can fill in any address for the From and To sections of the message, whether or not the addresses are yours and whether they actually exist or not.

Spam Filter
Spam filter is a program that is used to detect unsolicited and unwanted email and prevent those messages from getting to a user's inbox. Like other types of filtering programs, a spam filter looks for certain criteria on which it bases judgments. For example, the simplest and earliest versions (such as the one available with Microsoft's Hotmail) can be set to watch for particular words in the subject line of messages and to exclude these from the user's inbox. This method is not especially effective, too often omitting perfectly legitimate messages (these are called false positives) and letting actual spam through. More sophisticated programs, such as Bayesian filters or other heuristic filters, attempt to identify spam through suspicious word patterns or word frequency.

Bayesian Filtering
Bayesian spam filtering is the process of using a naive Bayes classifier to identify spam e-mail. It is based on the principle that most events are dependent and that the probability of an event occurring in the future can be inferred from the previous occurrences of that event. This same technique can be used to classify spam. If some piece of text occurs often in spam but not in legitimate mail, then it would be reasonable to assume that this email is probably spam.

Bayesian spam filtering has become a popular mechanism to distinguish illegitimate spam email from legitimate email. Nowadays many mail clients implement Bayesian spam filtering.

Bayesian filters must be 'trained' to work effectively. Particular words have certain probabilities (also known as likelihood functions) of occurring in spam email but not in legitimate email. For instance, most email users will frequently encounter the word

Viagra in spam email, but will seldom see it in other email. Before mail can be filtered using this method, the user needs to generate a database with words and tokens (such as the $ sign, IP addresses and domains, and so on), collected from a sample of spam mail and valid mail (referred to as 'ham'). For all words in each training email, the filter will adjust the probabilities that each word will appear in spam or legitimate email in its database.

After training, the word probabilities are used to compute the probability that an email with a particular set of words in it belongs to either category. If the total of word probabilities exceeds a certain threshold, the filter will mark the email as spam. Users can then decide whether to move email marked as spam to their spam folder or whether to just delete them.

Web application firewall
A web application firewall (WAF) is an appliance, server plugin, or filter that applies a set of rules to an HTTP conversation. Generally, these rules cover common attacks such as Cross-site Scripting (XSS) and SQL Injection. By customizing the rules to your application, many attacks can be identified and blocked. The effort to perform this customization can be significant and needs to be maintained as the application is modified.

A network-based application layer firewall is a computer networking firewall operating at the application layer of a protocol stack, and is also known as a proxy-based or reverse-proxy firewall. Application firewalls specific to a particular kind of network traffic may be titled with the service name, such as a web application firewall. They may be implemented through software running on a host or a stand-alone piece of network hardware. Often, it is a host using various forms of proxy servers to proxy traffic before passing it on to the client or server. Because it acts on the application layer, it may inspect the contents of the traffic, blocking specified content, such as certain websites, viruses, and attempts to exploit known logical flaws in client software.

Network-based application-layer firewalls work on the application level of the network stack (for example, all web browser, telnet, or ftp traffic), and may intercept all packets traveling to or from an application. In principle, application firewalls can prevent all unwanted outside traffic from reaching protected machines.

Modern application firewalls may also offload encryption from servers, block application input/output from detected intrusions or malformed communication, manage or consolidate authentication, or block content which violates policies.

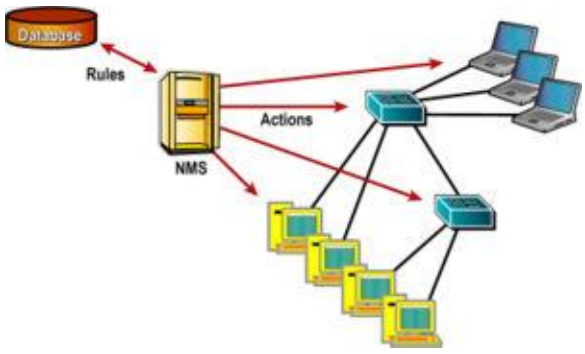
Section 1.2 Given a scenario, use secures network administration principles.

Rule-based management
Traditional network management systems are implemented in a rules-based environment. Here, the functional areas of network management are performed based on a set of guidelines established by the administrative staff. Fault events, configuration setup, accounting recording, performance thresholds and security rules are all preset, based on best practices.

One advantage of rules-based implementations is that the interaction between the managed device (agent) and the manager can be simplified to a very small set of actions. In the simplest case the agent responds to polls from the manager and sends alarms when certain conditions are met. This disassociation allows rules-based network management to operate easily in a multivendor environment.

> Hardware independence is another benefit of rules-based implementations. The manager characterizes the agents by using a set of database (management information base) entries. As long as the hardware manufacturer uses the same naming constructs for the internal design of the agent, the same manager can manage it. The manager need not know how the action is performed, only that it is performed at the agent.

Rules-based management systems manage the devices and connectivity of a network. All links, interconnection devices, and hosts can be part of the management scheme. Two examples of rules-based management rules are: Collect interface utilization once every five months, or apply a filter that filters protocol X.



Firewall rules
You create firewall rules to allow this computer to send traffic to, or receive traffic from, programs, system services, computers, or users. Firewall rules can be created to take one of three actions for all connections that match the rule's criteria:

- Allow the connection.

- Allow a connection only if it is secured through the use of Internet Protocol security (IPsec).
- Block the connection.

Rules can be created for either inbound traffic or outbound traffic. The rule can be configured to specify the computers or users, program, service, or port and protocol. You can specify which type of network adapter the rule will be applied to: local area network (LAN), wireless, remote access, such as a virtual private network (VPN) connection, or all types. You can also configure the rule to be applied when any profile is being used or only when a specified profile is being used.

As your IT environment changes, you might have to change, create, disable, or delete rules.

Firewall rule priority
Because you can make firewall rules that have apparent conflicts, it is important to understand the order in which the rules are processed:

- **Authenticated bypass**. These are rules in which the Override block rules option is selected. These rules allow matching network traffic that would otherwise be blocked. The network traffic must be authenticated by using a separate connection security rule. You can use these rules to permit access to the computer to authorized network administrators and authorized network troubleshooting devices.
- **Block connection.** These rules block all matching inbound network traffic.

- **Allow connection**. These rules allow matching inbound network traffic. Because the default behavior is to block unsolicited inbound network traffic, you must create an allow rule to support any network program or service that must be able to accept inbound connections.

- **Default profile behavior**. The default behavior is to block unsolicited inbound network traffic, but to allow all outbound network traffic. You can change the default behavior on the Domain Profile, Private Profile, and Public Profile tabs of the Windows Firewall with Advanced Security Properties dialog box.

As soon as a network packet matches a rule, that rule is applied, and processing stops. For example, an arriving network packet is first compared to the authenticated bypass rules. If it matches one, that rule is applied and processing stops. The packet is not compared to the block, allow, or default profile rules. If the packet does not match an authenticated bypass rule, then it is compared to the block rules. If it matches one, the packet is blocked, and processing stops, and so on.

Inbound Rule
Inbound rules explicitly allow, or explicitly block, inbound network traffic that matches the criteria in the rule. For example, you can configure a rule to explicitly allow traffic secured by IPsec for Remote Desktop through the firewall, but block the same traffic if IPsec does not secure it. When Windows is first installed, all unsolicited inbound traffic is blocked. To allow a certain type of unsolicited inbound traffic, you must create an inbound rule that describes that traffic. For example, if you want to run a Web server, then you must create a rule that allows unsolicited inbound network traffic on TCP port 80.

You can also configure the default action that Windows Firewall with Advanced Security takes, whether connections are allowed or blocked, when no inbound rule applies.

Outbound Rule
Outbound rules explicitly allow, or explicitly block, network traffic originating from the computer that matches the criteria in the rule. For example, you can configure a rule to explicitly block outbound traffic to a computer (by IP address) through the firewall, but allow the same traffic for other computers. Because outbound traffic is allowed by default, you typically use outbound rules to block network traffic that you do not want.

You can also configure the default action that Windows Firewall with Advanced Security takes, whether outbound connections are allowed or blocked, when no outbound rule applies.

VLAN management
A VLAN Management Policy Server or "VMPS" is a network switch that contains a mapping of device information to VLAN.

The primary goal of VMPS is VLAN assignment for general network management purposes, but can also be used for providing security through segregating clients with an unknown MAC address, or through further extension of the protocol to provide login for Cisco ACS. This last functionality is now deprecated by Cisco, in favour of 802.1x, and as the VMPS technology is Cisco only, the VLAN assignment can now be carried out in the 802.1x framework.

Client switches query the VMPS server using the VLAN Query Protocol, or VQP. Only Cisco produces hardware with VMPS client functionality, and is currently fully supported across their IOS switching lines. Cisco officially only supports the use of Catalyst 4000, 5000 and 6500 switch platforms (with appropriate firmware) as VMPS servers, but these have limited functionality, and only support a static text file transferred into them with tftp.

Secure router configuration
Then it comes to an enterprise's network, routers are at the top of the food chain. Clients request information, servers provide information, and switches connect clients and servers together. But routers run the network.

The security you add when managing routers can make the difference between providing a functional and responsive network or an isolated intranet that provides services to no one. Let's look at some steps you can take to maintain router security.

Managing your routers starts with how you configure them. If you don't have a baseline document that details your routers' configurations, you need to create one.

Establishing and documenting a router's configuration brings you to the first crucial step in securely managing that configuration: Loading and storing the initial baseline configuration in a secure manner is essential.

Ideally, you should perform the initial configuration from the console and store it on a network drive. Most important, do not store it on the local drive of a laptop! Portable

computing devices (i.e., laptops, PDAs, memory sticks, etc.) have a way of getting lost or stolen, which can compromise the integrity and functionality of your entire network.

After you've loaded the configuration, your next step is to synchronize the running configuration with the startup configuration. But don't think you're finished once the router is up and running on the network — you need to maintain that configuration and make changes periodically.

Some administrators like to make changes online, while others prefer making changes offline and then uploading the configuration. Both have their benefits.

When making online changes, you can get immediate feedback as well as syntax checking. For example, the router will alert you if you misspell a command. In addition, if you make a change that causes problems with your network, you'll generally know right away.

On the other hand, if you make offline changes, you have the opportunity to add comments and use router configuration editors. However, this method provides no syntax checking or feedback on changes.

If you decide to use the offline approach, make sure you use a secure method of configuration delivery. Trivial File Transfer Protocol (TFTP) is not a recommended method for delivery, as it provides no security for connection or delivery of your configuration. File Transfer Protocol (FTP) — as long as you configure a username and password — or Secure Copy Protocol (SCP) are the most secure methods of delivering a new configuration.

Regardless of how you manage the updates of your router configurations, it's essential that you save each configuration change and document all modifications. This enables you and others to better understand the changes and review them if something goes awry.

Access control lists

An access control list (ACL) is a list of access control entries (ACE). Each ACE in an ACL identifies a trustee and specifies the access rights allowed, denied, or audited for that trustee. The security descriptor for a securable object can contain two types of ACLs: a DACL and a SACL.

A discretionary access control list (DACL) identifies the trustees that are allowed or denied access to a securable object. When a process tries to access a securable object, the system checks the ACEs in the object's DACL to determine whether to grant access to it. If the object does not have a DACL, the system grants full access to everyone. If the object's DACL has no ACEs, the system denies all attempts to access the object because the DACL does not allow any access rights. The system checks the ACEs in sequence until it finds one or more ACEs that allow all the requested access rights, or until any of the requested access rights are denied.

A system access control list (SACL) enables administrators to log attempts to access a secured object. Each ACE specifies the types of access attempts by a specified trustee that cause the system to generate a record in the security event log. An ACE in a SACL can generate audit records when an access attempt fails, when it succeeds, or both. In future releases, a SACL will also be able to raise an alarm when an unauthorized user attempts to gain access to an object.

Do not try to work directly with the contents of an ACL. To ensure that ACLs are semantically correct, use the appropriate functions to create and manipulate ACLs.

ACLs also provide access control to Microsoft Active Directory directory service objects. Active Directory Service Interfaces (ADSI) includes routines to create and modify the contents of these ACLs.

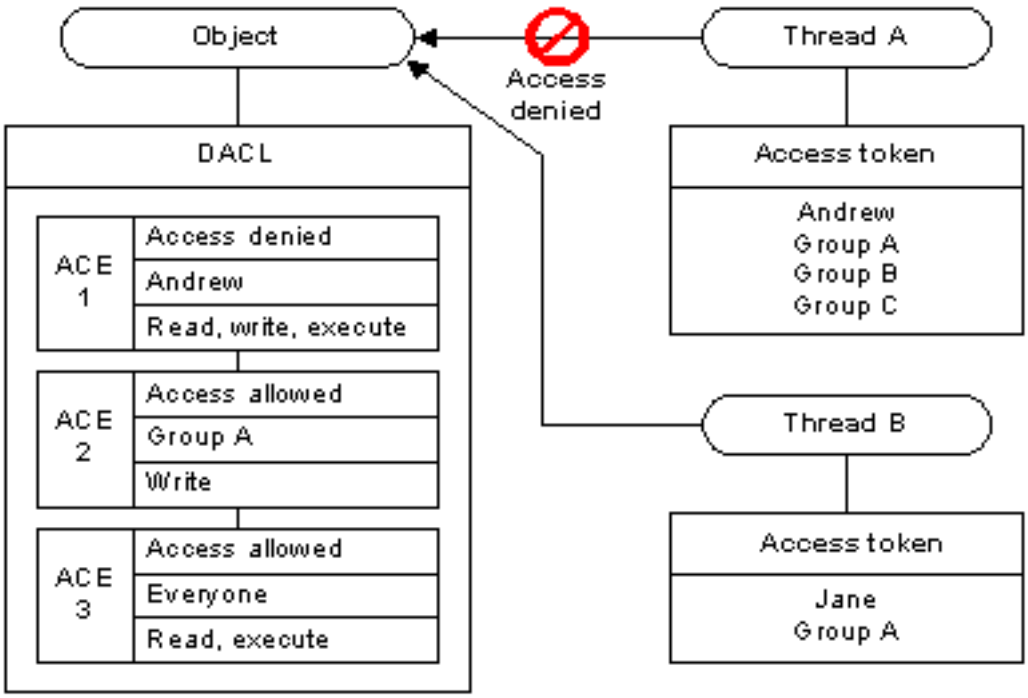
How DACLs control an object
When a thread tries to access a securable object, the system either grants or denies access. If the object does not have a discretionary access control list (DACL), the system grants access; otherwise, the system looks for Access Control Entries (ACEs) in the object's DACL that apply to the thread. Each ACE in the object's DACL specifies the access rights allowed or denied for a trustee, which can be a user account, a group account, or a logon session.

The system compares the trustee in each ACE to the trustees identified in the thread's access token. An access token contains security identifiers (SIDs) that identify the user and the group accounts to which the user belongs. A token also contains a logon SID that identifies the current logon session. During an access check, the system ignores group SIDs that are not enabled.

Typically, the system uses the primary access token of the thread that is requesting access. However, if the thread is impersonating another user, the system uses the thread's impersonation token.

The system examines each ACE in sequence until one of the following events occurs:

- An access-denied ACE explicitly denies any of the requested access rights to one of the trustees listed in the thread's access token.
- One or more access-allowed ACEs for trustees listed in the thread's access token explicitly grant all the requested access rights.
- All ACEs have been checked and there is still at least one requested access right that has not been explicitly allowed, in which case, access is implicitly denied.
- The following illustration shows how an object's DACL can allow access to one thread while denying access to another.



For Thread A, the system reads ACE 1 and immediately denies access because the access-denied ACE applies to the user in the thread's access token. In this case, the system does not check ACEs 2 and 3. For Thread B, ACE 1 does not apply, so the system proceeds to ACE 2, which allow™s write access, and ACE 3 which allows read and execute access.

Because the system stops checking ACEs when the requested access is explicitly granted or denied, the order of ACEs in a DACL is important. The ACE order were different in the example, the system might have granted access to Thread A. For system objects, the operating system defines a preferred order of ACEs in a DACL.

Port Security
You can use the port security feature to restrict input to an interface by limiting and identifying MAC addresses of the workstations that are allowed to access the port. When you assign secure MAC addresses to a secure port, the port does not forward packets with source addresses outside the group of defined addresses. If you limit the number of secure MAC addresses to one and assign a single secure MAC address, the workstation attached to that port is assured the full bandwidth of the port.

If a port is configured as a secure port and the maximum number of secure MAC addresses is reached, when the MAC address of a workstation attempting to access the port is different from any of the identified secure MAC addresses, a security violation occurs.

After you have set the maximum number of secure MAC addresses on a port, the secure addresses are included in an address table in one of these ways:

- You can configure all secure MAC addresses by using **the switchport port-security mac-address** mac_address interface configuration command.
- You can allow the port to dynamically configure secure MAC addresses with the MAC addresses of connected devices.

- You can configure a number of addresses and allow the rest to be dynamically configured.

After the maximum number of secure MAC addresses is configured, they are stored in an address table. To ensure that an attached device has the full bandwidth of the port, configure the MAC address of the attached device and set the maximum number of addresses to one, which is the default.

A security violation occurs if the maximum number of secure MAC addresses has been added to the address table and a workstation whose MAC address is not in the address table attempts to access the interface.

You can configure the interface for one of these violation modes, based on the action to be taken if a violation occurs:

Restrict—A port security violation restricts data, causes the SecurityViolation counter to increment, and causes an SNMP Notification to be generated. The rate at which SNMP traps are generated can be controlled by the snmp-server enable traps port-security trap-rate command. The default value ("0") causes an SNMP trap to be generated for every security violation.

Shutdown—A port security violation causes the interface to shut down immediately. When a secure port is in the error-disabled state, you can bring it out of this state by entering the errdisable recovery cause psecure_violation global configuration command or you can manually reenable it by entering the shutdown and no shut down interface configuration commands. This is the default mode.

You can also customize the time to recover from the specified error disable cause (default is 300 seconds) by entering the err-disable recovery interval command.

802.1x

The IEEE 802.1x standard manages port-based network access. It authenticates devices attached to a LAN port by initiating a connection and requesting login details. Access is prevented if authentication fails.

As well as being valuable for authenticating and controlling user traffic to a protected network, 802.1x is effective for dynamically varying encryption keys. 802.1x attaches the Extensible Authentication

Protocol (EAP) to both wired and wireless LAN media, and supports multiple authentication methods, such as token cards, one-time passwords, certificates, and public key authentication.

The Main Elements of the 802.1x System

Supplicant - The supplicant is the client that wishes to access services offered by the authenticator's system.The supplicant is responsible for answering any requests from the authenticator for informationthat establishes the supplicant's identity.

Port - A port is where a device is attached to the LAN, either directly into the switch or a wirelessaccess point.

Authenticator - The authenticator challenges the supplicant for appropriate authentication before it allowsaccess to the services available via the port. The authenticator communicates with thesupplicant and submits the information received from the supplicant to a suitable authenticationserver. This allows the verification of user credentials to determine the consequent portauthorization state.The authenticator's functionality is independent of the authenticationmethod. It acts as a go-between for the supplicant and authentication server.

Extensible Authentication Protocol - 802.1X uses Extensible Authentication Protocol (EAP) as an authentication tool. EAP carries out the authentication exchange between the supplicant and the authentication server. Noother devices such as access points and proxy servers take part in this exchange.

Extensible Authentication Protocol Over LAN- The Extensible Authentication Protocol Over LAN (EAPOL) captures the EAP messages sothey can be managed directly by a LAN MAC service. Management functions such as start,logoff, and key distribution are also provided by EAPOL.

Remote Access Dial In User Service - The Remote Authentication Dial In User Service (RADIUS) server:

- Manages a database of users.
- Provides authentication by verifying username and password.
- Optionally, provides authorization such as dynamic VLAN assignment.
- Optionally, provides accounting information about how long a user was connected, and how much data they transferred.

Unified Threat Management
In the broadest sense of the term, any freestanding device that operates in a largely self-contained manner is considered to be an appliance. An all-in-one appliance, also known as Unified Threat Management (UTM) and Next Generation Firewall (NGFW), is one that provides a good foundation for security. When you combine a firewall with other abilities (intrusion prevention, antivirus, content filtering, etc.), what used to be called an all-in-one appliance is now known as a UTM. The advantages of combining everything into one include a reduced learning curve (you only have one product to learn), a single vendor to deal with, and—typically—reduced complexity. The disadvantages of combining everything into one include a potential single point of failure, and the dependence on the one vendor.

Section 1.3 - Explain network design elements and components.

DMZ
In computer networking, DMZ is a firewall configuration for securing local area networks (LANs).

In a DMZ configuration, most computers on the LAN run behind a firewall connected to a public network like the Internet. One or more computers also run outside the firewall, in the

DMZ. Those computers on the outside intercept traffic and broker requests for the rest of the LAN, adding an extra layer of protection for computers behind the firewall.

Traditional DMZs allow computers behind the firewall to initiate requests outbound to the DMZ. Computers in the DMZ in turn respond, forward or re-issue requests out to the Internet or other public network, as proxy servers do. (Many DMZ implementations, in fact, simply utilize a proxy server or servers as the computers within the DMZ.) The LAN firewall, though, prevents computers in the DMZ from initiating inbound requests.

DMZ is a commonly touted feature of home broadband routers. However, in most instances these features are not true DMZs. Broadband routers often implement a DMZ simply through additional firewall rules, meaning that incoming requests reach the firewall directly. In a true DMZ, incoming requests must first pass through a DMZ computer before reaching the firewall.

Sub-netting
Sub-netting an IP Network can be done for a variety of reasons, including organization, use of different physical media (such as Ethernet, FDDI, WAN, etc.), preservation of address space, and security. The most common reason is to control network traffic. In an Ethernet network, all nodes on a segment see all the packets transmitted by all the other nodes on that segment. Performance can be adversely affected under heavy traffic loads, due to collisions and the resulting retransmissions. A router is used to connect IP networks to minimize the amount of traffic each segment must receive.

Subnet Masking
Applying a subnet mask to an IP address allows you to identify the network and node parts of the address. The 1s in the mask represents the network bits, and the 0s represents the node bits. Performing a bitwise logical AND operation between the IP address and the subnet mask results in the Network Address or Number.

VLAN
The other security feature that can be enabled in some switches is the concept of *virtual local area networks (VLANs)*. Cisco defines a VLAN as a "broadcast domain within aswitched network," meaning that information is carried in broadcast mode only todevices within a VLAN. Switches that allow multiple VLANs to be defined enable broadcastmessages to be segregated into the specific VLANs. If each floor of an office, forexample, were to have a single switch and you had accounting functions on two floors,engineering functions on two floors, and sales functions on two floors, then separateVLANs for accounting, engineering, and sales would allow separate broadcast domainsfor each of these groups, even those that spanned floors. This configuration increasesnetwork segregation, increasing throughput and security.

Unused switch ports can be preconfigured into empty VLANs that do not connect to the rest of the network. This significantly increases security against unauthorized network connections. If, for example, a building is wired with network connections in all rooms, including multiple connections for convenience and future expansion, these unused ports become open to the network. One solution is to disconnect the connection at the switch, but this merely moves the network opening into the switch room.

The better solution is to disconnect it and disable the port in the switch. This can be accomplished by connecting all unused ports into a VLAN that isolates them from the rest of the network.

NAT

In computer networking, network address translation (NAT) is the process of modifying IP address information in IP packet headers while in transit across a traffic routing device.

The simplest type of NAT provides a one to one translation of IP addresses. RFC 2663 refers to this type of NAT as basic NAT. It is often also referred to as one-to-one NAT. In this type of NAT only the IP addresses, IP header checksum and any higher level checksums that include the IP address need to be changed. The rest of the packet can be left untouched (at least for basic TCP/UDP functionality, some higher level protocols may need further translation). Basic NATs can be used when there is a requirement to interconnect two IP networks with incompatible addressing.

However it is common to hide an entire IP address space, usually consisting of private IP addresses, behind a single IP address (or in some cases a small group of IP addresses) in another (usually public) address space. To avoid ambiguity in the handling of returned packets a one-to-many NAT must alter higher level information such as TCP/UDP ports in outgoing communications and must maintain a translation table so that return packets can be correctly translated back. RFC 2663 uses the term NAPT (network address and port translation). Other names for this type of NAT include PAT (port address translation), IP masquerading, NAT Overload and many-to-one NAT. Since this is the most common type of NAT it is often referred to simply as NAT.

As described, the method enables communication through the router only when the conversation originates in the masqueraded network, since this establishes the translation tables. For example, a web browser in the masqueraded network can browse a website outside, but a web browser outside could not browse a web site in the masqueraded network. However, most NAT devices today allow the network administrator to configure translation table entries for permanent use. This feature is often referred to as "static NAT" or port forwarding and allows traffic originating in the "outside" network to reach designated hosts in the masqueraded network.

In the mid-1990s NAT became a popular tool for alleviating the consequences of IPv4 address exhaustion. It has become a standard, indispensable feature in routers for home and small-office Internet connections. Most systems using NAT do so in order to enable multiple hosts on a private network to access the Internet using a single public IP address Network address translation has serious drawbacks on the quality of Internet connectivity and requires careful attention to the details of its implementation. In particular all types of NAT break the originally envisioned model of IP end-to-end connectivity across the Internet and NAPT makes it difficult for systems behind a NAT to accept incoming communications. As a result, NAT traversal methods have been devised to alleviate the issues encountered.

One to many NATs

The majority of NATs map multiple private hosts to one publicly exposed IP address. In a typical configuration, a local network uses one of the designated "private" IP address subnets (RFC 1918). A router on that network has a private address in that address space. The router

is also connected to the Internet with a "public" address assigned by an Internet service provider. As traffic passes from the local network to the Internet, the source address in each packet is translated on the fly from a private address to the public address. The router tracks basic data about each active connection (particularly the destination address and port). When a reply returns to the router, it uses the connection tracking data it stored during the outbound phase to determine the private address on the internal network to which to forward the reply.

All Internet packets have a source IP address and a destination IP address. Typically packets passing from the private network to the public network will have their source address modified while packets passing from thepublic network back to the privatenetwork will have their destination address modified. More complex configurations are also possible.

To avoid ambiguity in how to translate returned packets further modifications to the packets are required. The vast bulk of Internet traffic is TCP and UDP packets and for these protocols the port numbers are changed so that the combination of IP and port information on the returned packet can be unambiguously mapped to the corresponding private address and port information. Protocols not based on TCP or UDP require other translation techniques. ICMP packets typically relate to an existing connection and need to be mapped using the same IP and port mappings as that connection.

Type of NAT and NAT Traversal
The NAT traversal problem arises when two peers behind distinct NAT try to communicate. One way to solve this problem is to use port forwarding, another way is to use various NAT traversal techniques. The most popular technique for TCP NAT traversal is TCP hole punching, which requires the NAT to follow the port preservation design for TCP, as explained below.

Many NAT implementations follow the port preservation design especially for TCP, which is to say that they use the same values as internal and external port numbers. NAT port preservation for outgoing TCP connections is especially important for TCP NAT traversal, because programs usually bind distinct TCP sockets to ephemeral ports for distinct TCP connections, rendering NAT port prediction impossible for TCP. On the other hand, for UDP, NATs do not need to have port preservation because applications usually reuse the same UDP socket to send packets to distinct hosts, making port prediction straightforward, as it is the same source port for each packet. Furthermore, port preservation in NAT for TCP allows P2P protocols to offer less complexity and less latency because there is no need to use a third party to discover the NAT port since the application already knows the NAT port. However, if two internal hosts attempt to communicate with the same external host using the same port number, the external port number used by the second host will be chosen at random. Such NAT will be sometimes perceived as (address) restricted cone NAT and other times as symmetric NAT.

Remote Access
Remote Access Service (RAS) is a portion of the Windows OS that allows the connection between a client and a server via a dial-up telephone connection. Although slower than cable/DSL connections, this is still a common method for connecting to a remote network. When a user dials into the computer system, authentication and authorization are performed through a series of remote access protocols. For even greater security, a callback system can be employed, where the server calls back to the client at a set telephone number for the data exchange. RAS can also mean Remote Access Server, a term for a server designed to permit

remote users access to anetwork and to regulate their access. A variety of protocols and methods exist to perform this function.

Telephony
Data and voice communications have coexisted in enterprises for decades. Recent connections inside the enterprise of Voice over IP and traditional PBX solutions increase both functionality and security risks. Specific firewalls to protect against unauthorized traffic over telephony connections are available to counter the increased risk.

NAC
Networks comprise connected workstations and servers. Managing security on a network involves managing a wide range of issues, from various connected hardware and the software operating these devices. Assuming that the network is secure, each additional connection involves risk. Managing the endpoints on a case-by-case basis as they connect is a security methodology known as network access control. Two main competing methodologies exist: Network Access Protection (NAP) is a Microsoft technology for controlling network access of a computer host, and Network Admission Control (NAC) is Cisco's technology for controlling network admission.

Both the Cisco NAC and Microsoft NAP are in their early stages of implementation. The concept of automated admission checking based on client device characteristics is here to stay, as it provides timely control in the ever-changing network world of today's enterprises.

Virtualization
Virtualization, in computing, is the creation of a virtual (rather than actual) version of something, such as a hardware platform, operating system, a storage device or network resources. Virtualization can be viewed as part of an overall trend in enterprise IT that includes autonomic computing, a scenario in which the IT environment will be able to manage itself based on perceived activity, and utility computing, in which computer processing power is seen as a utility that clients can pay for only as needed. The usual goal of virtualization is to centralize administrative tasks while improving scalability and workloads.

Types of virtualization

Hardware
Hardware virtualization or platform virtualization refers to the creation of a virtual machine that acts like a real computer with an operating system. Software executed on these virtual machines is separated from the underlying hardware resources. For example, a computer that is running Microsoft Windows may host a virtual machine that looks like a computer with Mac OS X operating system. Subsequently, Mac OS X-based software can be run on that virtual machine.

In hardware virtualization, the term host machine refers to the actual machine on which the virtualization takes place; the term guest machine, however, refers to the virtual machine. Likewise, the adjectives host and guest are used to help distinguish the software that runs on the actual machine from the software that runs on the virtual machine. The software or firmware that creates a virtual machine on the host hardware is called a hypervisor or Virtual Machine Monitor.

Different types of hardware virtualization include:

1. Full virtualization: Almost complete simulation of the actual hardware to allow software, which typically consists of a guest operating system, to run unmodified

2. Partial virtualization: Some but not all of the target environment is simulated. Some guest programs, therefore, may need modifications to run in this virtual environment.

3. Para-virtualization: A hardware environment is not simulated; however, the guest programs are executed in their own isolated domains, as if they are running on a separate system. Guest programs need to be specifically modified to run in this environment.

Hardware-assisted virtualization is a way of improving the efficiency of hardware virtualization. It involves employing specially-designed CPUs and hardware components that help improve the performance of a guest environment.
Hardware virtualization must not be mistaken with hardware emulation: In hardware emulation, a piece of hardware imitates another, while in hardware virtualization, a hypervisor (a piece of software) imitates a particular piece of computer hardware or the whole computer altogether. Furthermore, a hypervisor must not be mistaken with an emulator. These two are defined similarly: Both are computer programs that imitate hardware, but their domain of use in language differs.

Software

- Operating system-level virtualization, hosting of multiple virtualized environments within a single OS instance
- Application virtualization and Workspace virtualization, the hosting of individual applications in an environment separated from the underlying OS

Memory

- Memory virtualization, aggregating RAM resources from networked systems into a single memory pool

- Virtual memory, giving an application program the impression that it has contiguous working memory, isolating it from the underlying physical memory implementation

Storage
Storage virtualization, the process of completely abstracting logical storage from physical storage

Data virtualization, the presentation of data as an abstract layer, independent of underlying database systems, structures and storage

Database virtualization, the decoupling of the database layer, which lies between the storage and application layers within the application stack

Network

Desktop virtualization, the concept of separating a desktop environment from its physical computer (and its associated operating system) and storing it on another machine across a network, such as a center server. Thin clients employ desktop virtualization.

Network virtualization, creation of a virtualized network addressing space within or across network subnets

Cloud Computing
Cloud computing is a general term for anything that involves delivering hosted services over the Internet. These services are broadly divided into three categories: Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) and Software-as-a-Service (SaaS). The name cloud computing was inspired by the cloud symbol that's often used to represent the Internet in flowcharts and diagrams.

A cloud service has three distinct characteristics that differentiate it from traditional hosting. It is sold on demand, typically by the minute or the hour; it is elastic -- a user can have as much or as little of a service as they want at any given time; and the service is fully managed by the provider (the consumer needs nothing but a personal computer and Internet access). Significant innovations in virtualization and distributed computing, as well as improved access to high-speed Internet and a weak economy, have accelerated interest in cloud computing.
A cloud can be private or public. A public cloud sells services to anyone on the Internet. (Currently, Amazon Web Services is the largest public cloud provider.) A private cloud is a proprietary network or a data center that supplies hosted services to a limited number of people. When a service provider uses public cloud resources to create their private cloud, the result is called a virtual private cloud. Private or public, the goal of cloud computing is to provide easy, scalable access to computing resources and IT services.

Infrastructure-as-a-Service like Amazon Web Services provides virtual server instanceAPI) to start, stop, access and configure their virtual servers and storage. In the enterprise, cloud computing allows a company to pay for only as much capacity as is needed, and bring more online as soon as required. Because this pay-for-what-you-use model resembles the way electricity, fuel and water are consumed; it's sometimes referred to as utility computing.

Platform-as-a-service in the cloud is defined as a set of software and product development tools hosted on the provider's infrastructure. Developers create applications on the provider's platform over the Internet. PaaS providers may use APIs, website portals or gateway software installed on the customer's computer. Force.com, (an outgrowth of Salesforce.com) and GoogleApps are examples of PaaS. Developers need to know that currently, there are not standards for interoperability or data portability in the cloud. Some providers will not allow software created by their customers to be moved off the provider's platform.

In the software-as-a-service cloud model, the vendor supplies the hardware infrastructure, the software product and interacts with the user through a front-end portal. SaaS is a very broad market. Services can be anything from Web-based email to inventory control and database processing. Because the service provider hosts both the application and the data, the end user is free to use the service from anywhere.

Layered security / Defense in depth
Two terms synonymous with each other are layered security and defense in depth. All these terms mean is that you should not rely on a single entity for protection but instead implement multiple layers of security. In a physical environment, for example, it is all well and good to have a guard posted at the entrance of the office building, but to keep the servers secure; you should also put a lock on the server room door. From a technology standpoint,a firewall is a great thing to restrict traffic into the network from the outside, but you will also want to have antivirus software, intrusion detection, and as many other layers of security as you can to truly protect the systems.

Section 1.4 -Given a scenario, implement common protocols and services.

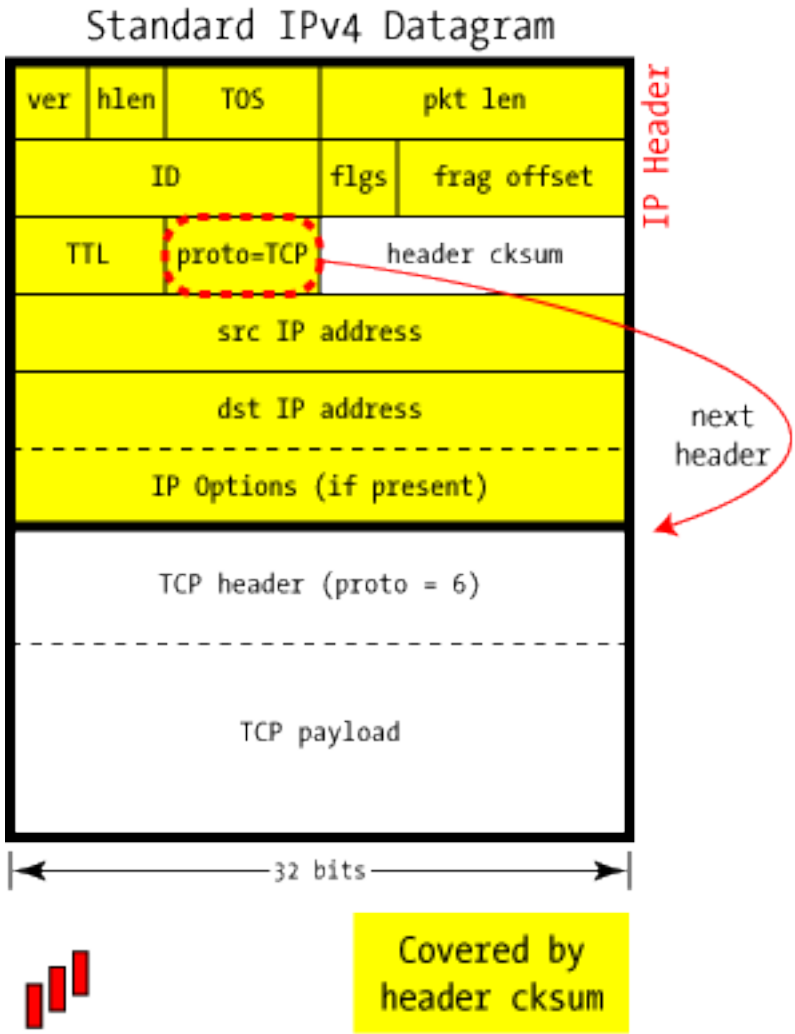
IPSec
Internet Protocol security (IPsec) uses cryptographic security services to protect communications over Internet Protocol (IP) networks. IPsec supports network-level peer authentication, data origin authentication, data integrity, data confidentiality (encryption), and replay protection. The Microsoft implementation of IPsec is based on Internet Engineering Task Force (IETF) standards.

In Windows 7, Windows Server 2008 R2, Windows Vista and Windows Server 2008, you can configure IPsec behavior by using the Windows Firewall with Advanced Security snap-in. In earlier versions of Windows, IPsec was a stand-alone technology separate from Windows Firewall.

One of the first things that one notices when trying to set up IPsec is that there are so many knobs and settings: even a pair of entirely standards-conforming implementations sports a bewildering number of ways to impede a successful connection. It's just an astonishingly complex suite of protocols.

One cause of the complexity is that IPsec provides mechanism, not policy: rather than define such-and-such encryption algorithm or a certain authentication function, it provides a framework that allows an implementation to provide nearly anything that both ends agree upon.

The IP Datagram
Since we're looking at IPsec from the bottom up, we must first take a brief detour to revisit the IP Header itself, which carries all of the traffic we'll be considering.

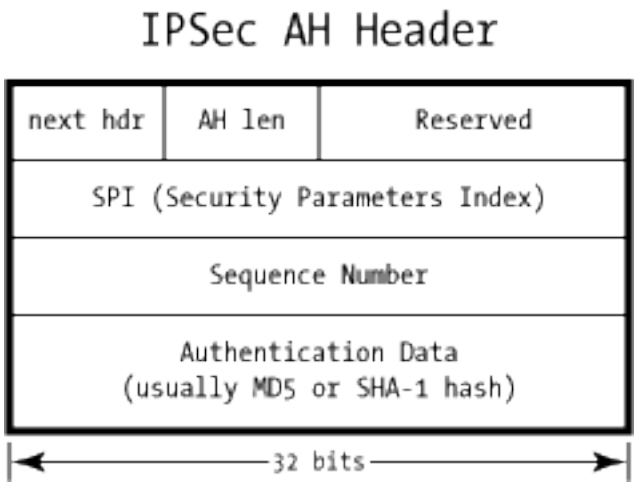


AH: Authentication Only

AH is used to authenticate — but not encrypt — IP traffic, and this serves the treble purpose of ensuring that we're really talking to who we think we are, detecting alteration of data while in transit, and (optionally) to guard against replay by attackers who capture data from the wire and attempt to re-inject that data back onto the wire at a later date.

Authentication is performed by computing a cryptographic hash-based message authentication code over nearly all the fields of the IP packet (excluding those which might be modified in transit, such as TTL or the header checksum), and stores this in a newly added AH header and sent to the other end.

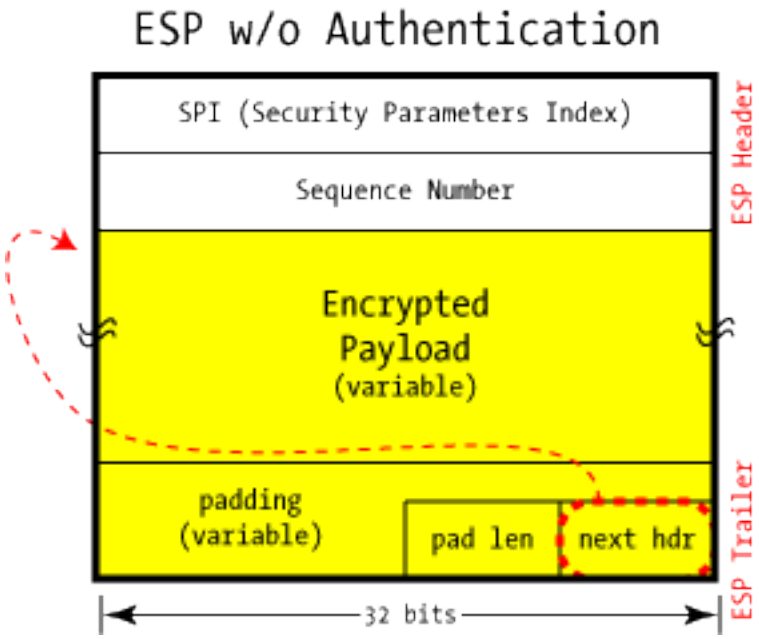
This AH header contains just five interesting fields, and it's injected between the original IP header and the payload. We'll touch on each of the fields here, though their utility may not be fully apparent until we see how they're used in the larger picture.



ESP — Encapsulating Security Payload
Adding encryption makes ESP a bit more complicated because the encapsulation surrounds the payload rather than precedes it as with AH: ESP includes header and trailer fields to support the encryption and optional authentication. It also provides Tunnel and Transport modes, which are used in by-now familiar ways.
The IPsec RFCs don't insist upon any particular encryption algorithms, but we find DES, triple-DES, AES, and Blowfish in common use to shield the payload from prying eyes. The Security Association specifies the algorithm used for a particular connection, and this SA includes not only the algorithm, but the key used.

Unlike AH, which provides a small header before the payload, ESP surrounds the payload it's protecting. The Security Parameters Index and Sequence Number serve the same purpose as in AH, but we find padding, the next header, and the optional Authentication Data at the end, in the ESP Trailer.

It's possible to use ESP without any actual encryption (to use a NULL algorithm), which nonetheless structures the packet the same way. This provides no confidentiality, and it only makes sense if combined with ESP authentication. It's pointless to use ESP without either encryption or authentication (unless one is simply doing protocol testing).

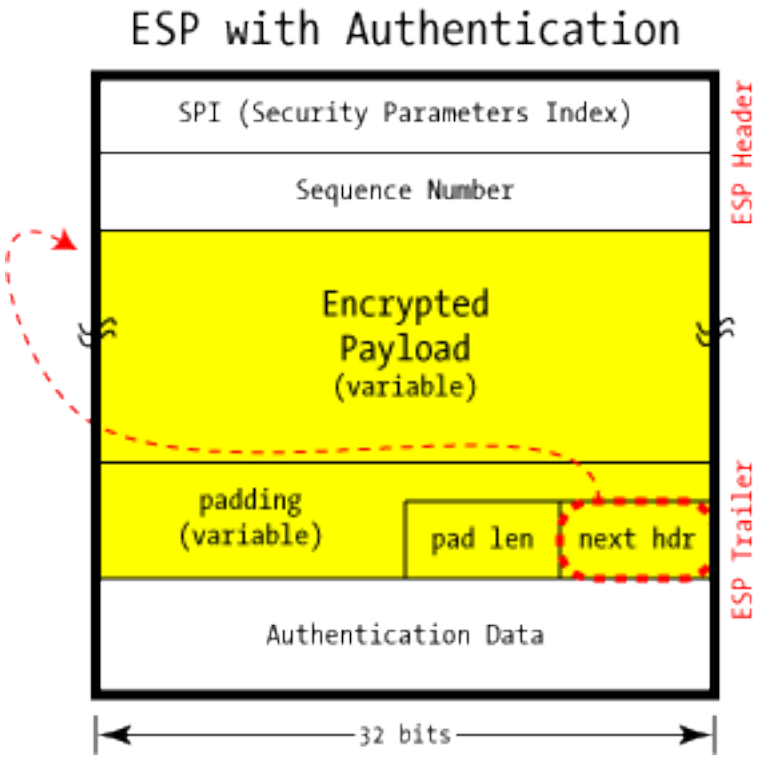


Padding is provided to allow block-oriented encryption algorithms room for multiples of their blocksize, and the length of that padding is provided in the pad len field. The next hdr field gives the type (IP, TCP, UDP, etc.) of the payload in the usual way, though it can be thought of as pointing "backwards" into the packet rather than forward as we've seen in AH.

In addition to encryption, ESP can also optionally provide authentication, with the same HMAC as found in AH. Unlike AH, however, this authentication is only for the ESP header and encrypted payload: it does not cover the full IP packet. Surprisingly, this does not substantially weaken the security of the authentication, but it does provide some important benefits.

When an outsider examines an IP packet containing ESP data, it's essentially impossible to make any real guesses about what's inside save for the usual data found in the IP header (particularly the source and destination IP addresses). The attacker will certainly know that it's ESP data — that's also in the header — but the type of the payload is encrypted with the payload.

Even the presence or absence of Authentication Data can't be determined by looking at the packet itself (this determination is made by using the Security Parameters Index to reference the pre-shared set of parameters and algorithms for this connection).



However, it should be noted that sometimes the envelope provides hints that the payload does not. With more people sending VoIP inside ESP over the Internet, the QoS tagging are in the outside header and is fairly obvious what traffic is VoIP signaling (IP precedence 3) and what is RTP traffic (IP precedence 5). It's not a sure thing, but it might be enough of a clue to matter in some circumstances.

SNMP

Since its creation in 1988 as a short-term solution to manage elements in the growing Internet and other attached networks, SNMP has achieved widespread acceptance. SNMP was derived from its predecessor SGMP (Simple Gateway Management Protocol) and was intended to be replaced by a solution based on the CMIS/CMIP (Common Management Information Service/Protocol) architecture. This long-term solution, however, never received the widespread acceptance of SNMP.

SNMP is based on the manager/agent model consisting of an SNMP manager, an SNMP agent, a database of management information, managed SNMP devices and the network protocol. The SNMP manager provides the interface between the human network manager and the management system. The SNMP agent provides the interface between the manager and the physical device(s) being managed (see the illustration above).

The SNMP manager and agent use an SNMP Management Information Base (MIB) and a relatively small set of commands to exchange information. The SNMP MIB is organized in a tree structure with individual variables, such as point status ordescription, being represented as leaves on the branches. A long numeric tag or object identifier (OID) is used to distinguish each variable uniquely in the MIB and in SNMP messages.

SNMP uses five basic messages (GET, GET-NEXT, GET-RESPONSE, SET, and TRAP) to communicate between the SNMP manager and the SNMP agent. The GET and GET-NEXT messages allow the manager to request information for a specific variable.

The agent, upon receiving a GET or GET-NEXT message, will issue a GET-RESPONSE message to the SNMP manager with either the information requested or an error indication as to why the request cannot be processed. A SET message allows the SNMP manager to request a change be made to the value of a specific variable in the case of an alarm remote that will operate a relay. The SNMP agent will then respond with a GET-RESPONSE message indicating the change has been made or an error indication as to why the change cannot be made. The SNMP TRAP message allows the agent to spontaneously inform the SNMP manager of an "important" event.

As you can see, most of the messages (GET, GET-NEXT, and SET) are only issued by the SNMP manager. Because the TRAP message is the only message capable of being initiated by an SNMP agent, it is the message used by DPS Remote Telemetry Units (RTUs) to report alarms. This notifies the SNMP manager as soon as an alarm condition occurs, instead of waiting for the SNMP manager to ask.

The small number of commands used is only one of the reasons SNMP is "simple." The other simplifying factor is the SNMP protocol's reliance on an unsupervised or connectionless communication link. This simplicity has led directly to the widespread use of SNMP, specifically in the Internet Network Management Framework. Within this framework, it is considered "robust" because of the independence of the SNMP managers from the agents, e.g. if an SNMP agent fails, the SNMP manager will continue to function, or vice versa.

SSH

Secure Shell or SSH is a network protocol that allows data to be exchanged using a secure channel between two networked devices. The two major versions of the protocol are referred to as SSH1 or SSH-1 and SSH2 or SSH-2. Used primarily on Linux and Unix based systems to access shell accounts, SSH was designed as a replacement for Telnet and other insecure remote shells, which send information, notably passwords, in plaintext, rendering them susceptible to packet analysis. The encryption used by SSH is intended to provide confidentiality and integrity of data over an unsecured network, such as the Internet.

DNS

The DNS translates Internet domain and host names to IP addresses. DNS automatically converts the names we type in our Web browser address bar to the IP addresses of Web servers hosting those sites.
DNS implements a distributed database to store this name and address information for all public hosts on the Internet. DNS assumes IP addresses do not change (are statically assigned rather than dynamically assigned).

The DNS database resides on a hierarchy of special database servers. When clients like Web browsers issue requests involving Internet host names, a piece of software called the DNS resolver (usually built into the network operating system) first contacts a DNS server to determine the server's IP address. If the DNS server does not contain the needed mapping, it will in turn forward the request to a different DNS server at the next higher level in the hierarchy. After potentially several forwarding and delegation messages are sent within the

DNS hierarchy, the IP address for the given host eventually arrives at the resolver, that in turn completes the request over Internet Protocol.

DNS additionally includes support for caching requests and for redundancy. Most network operating systems support configuration of primary, secondary, and tertiary DNS servers, each of which can service initial requests from clients. ISPs maintain their own DNS servers and use DHCP to automatically configure clients, relieving most home users of the burden of DNS configuration.

Encryption (SSL and TLS)

Secure Sockets Layer (SSL) is a general-purpose protocol developed by Netscape for managing the encryption of information being transmitted over the Internet. It began as a competitive feature to drive sales of Netscape's web server product, which could then send information securely to end users. This early vision of securing the transmission channel between the web server and the browser became an Internet standard.

Today, SSL is almost ubiquitous with respect to e-commerce—all browsers support it as do web servers, and virtually all sensitive financial traffic from e-commerce web sites uses this method to protect information in transit between web servers and browsers. The Internet Engineering Task Force (IETF) embraced SSL in 1996 through a series of RFCs and named the group Transport Layer Security (TLS). Starting with SSL 3.0, in 1999 the IETF issued RFC 2246, "TLS Protocol Version 1.0," followed by RFC 2712, which added Kerberos authentication, and then RFCs 2817 and 2818, which extended TLS to HTTP version 1.1 (HTTP/1.1). Although SSL has been through several versions,

TLS begins with an equivalency to SSL 3.0, so today SSL and TLS are essentially the same although not inter-changeable. SSL/TLS is a series of functions that exist in the OSI (Open System Interconnection) model between the application layer and the transport and network layers. The goal of TCP is to send an unauthenticated error-free stream of information between two computers. SSL/TLS adds message integrity and authentication functionality to TCP through the use of cryptographic methods. Because cryptographic methods are an ever-evolving field, and because both parties must agree on an implementation method, SSL/TLS has embraced an open, extensible, and adaptable method to allow flexibility and strength.

When two programs initiate an SSL/TLS connection, one of their first tasks is to compare available protocols and agree on an appropriate common cryptographic protocol for use in this particular communication. As SSL/TLS can use separate algorithms and methods for encryption, authentication, and data integrity, each of these is negotiated and determined depending upon need at the beginning of a communication.

How SSL/TLS Works

SSL/TLS uses a wide range of cryptographic protocols. To use these protocols effectively between a client and a server, an agreement must be reached on which protocol to use via the SSL handshake process. The process begins with a client request for a secure connection and a server's response. The questions asked and answered are which protocol and which cryptographic algorithm will be used. For the client and server to communicate, both sides must agree on a commonly held protocol (SSL v1, v2, v3, or TLS v1). Commonly available cryptographic algorithms include Diffie-Hellman and RSA. The next step is to exchange certificates and keys as necessary to enable authentication. Authentication was a one-way

process for SSL v1 and v2 with only the server-providing authentication. In SSL v3/TLS, mutual authentication of both client and server is possible.

The certificate exchange is via X.509 certificates, and public key cryptography is used to establish authentication. Once authentication is established, the channel is secured with symmetric key cryptographic methods and hashes, typically RC4 or 3DES for symmetric key and MD5 or SHA-1 for the hash functions.

The Web (HTTP and HTTPS)
HTTP is used for the transfer of hyperlinked data over the Internet, from web servers to browsers. When a user types a URL such as http://www.example.com into a browser, the http:// portion indicates that the desired method of data transfer is HTTP. Although it was initially created just for HTML pages, today many protocols deliver content over this connection protocol. HTTP traffic takes place over TCP port 80 by default, and this port is typically left open on firewalls because of the extensive use of HTTP.

One of the primary drivers behind the development of SSL/TLS was the desire to hide the complexities of cryptography from end users. When using an SSL/TLS-enabled browser, simply requesting a secure connection from a web server instead of non-secure connection can do this.

When a browser is SSL/TLS-aware, the entry of an SSL/TLS-based protocol will cause the browser to perform the necessary negotiations with the web server to establish the required level of security. Once these negotiations have been completed and a session key secures the session, a closed padlock icon is displayed in the lower right of the screen to indicate that the session is secure. If the protocol is https:, your connection is secure; if it is http:, then the connection is carried by plaintext for anyone to see. As the tiny padlock placed in the lower-right corner of the screen could have been missed, Microsoft moved it to an obvious position next to the URL in Internet Explorer 7. Another new security feature that begins with Internet Explorer 7 and Firefox 3 is the use of high assurance SSL, a combination of an extended validation SSL certificate and a high security browser. If a high security browser, Internet Explorer 7 or Firefox 3 and beyond, establish a connection with a vendor that has registered with a certificate authority for an extended validation SSL certificate, then the URL box will be colored green and the box next to it will display the registered entity and additional validation information when clicked. These improvements were in response to phishing sites and online fraud, and although they require additional costs and registration on the part of the vendors, this is a modest up-front cost to help reduce fraud and provide confidence to customers.

One important note on SSL certificate-based security is the concept of single- versus dual-sided authentication. The vast majority of SSL connections are single-sided, meaning that only the identity of the server side is vouched for via a certificate. The client is typically not identified by certificate, mainly because of the number of clients and corresponding PKI issues. A single-sided SSL secured conversation can be attacked using a man-in-the-middle attack by capturing all the traffic and relaying responses. Dual-sided SSL would prevent this attack mechanism, yet the management of every client needing to obtain and maintain a certificate makes this practically infeasible with the current PKI available to most end users.

The objective of enabling cryptographic methods in this fashion is to make it easy for end users to use these protocols. SSL/TLS is designed to be protocol agnostic. Although designed

to run on top of TCP/IP, it can operate on top of other lower level protocols, such as X.25. SSL/TLS requires a reliable lower level protocol, so it is not designed and cannot properly function on top of a non-reliable protocol such as the User Datagram Protocol (UDP). Even with this limitation, SSL/TLS has been used to secure many common TCP/IP-based services

TCP/IP

As with all other communications protocol, TCP/IP is composed of layers:

- **IP** - is responsible for moving packet of data from node to node. IP forwards each packet based on a four byte destination address (the IP number). The Internet authorities assign ranges of numbers to different organizations. The organizations assign groups of their numbers to departments. IP operates on gateway machines that move data from department to organization to region and then around the world.
- **TCP** - is responsible for verifying the correct delivery of data from client to server. Data can be lost in the intermediate network. TCP adds support to detect errors or lost data and to trigger retransmission until the data is correctly and completely received.
- **Sockets** - is a name given to the package of subroutines that provide access to TCP/IP on most systems.

The Internet Protocol was developed to create a Network of Networks (the "Internet"). Individual machines are first connected to a LAN (Ethernet or Token Ring). TCP/IP shares the LAN with other uses (a Novell file server, Windows for Workgroups peer systems). One device provides the TCP/IP connection between the LAN and the rest of the world.

To insure that all types of systems from all vendors can communicate, TCP/IP is absolutely standardized on the LAN. However, larger networks based on long distances and phone lines are more volatile. In the US, many large corporations would wish to reuse large internal networks based on IBM's SNA. In Europe, the national phone companies traditionally standardize on X.25. However, the sudden explosion of high-speed microprocessors, fiber optics, and digital phone systems has created a burst of new options: ISDN, frame relay, FDDI, Asynchronous Transfer Mode (ATM). New technologies arise and become obsolete within a few years. With cable TV and phone companies competing to build the National Information Superhighway, no single standard can govern citywide, nationwide, or worldwide communications.

The original design of TCP/IP as a Network of Networks fits nicely within the current technological uncertainty. TCP/IP data can be sent across a LAN, or it can be carried within an internal corporate SNA network, or it can piggyback on the cable TV service. Furthermore, machines connected to any of these networks can communicate to any other network through gateways supplied by the network vendor.

Addresses

Each technology has its own convention for transmitting messages between two machines within the same network. On a LAN, messages are sent between machines by supplying the six byte unique identifier (the "MAC" address). In an SNA network, every machine has Logical Units with their own network address. DECNET, Appletalk, and Novell IPX all have a scheme for assigning numbers to each local network and to each workstation attached to the network.

On top of these local or vendor specific network addresses, TCP/IP assigns a unique number to every workstation in the world. This "IP number" is a four byte value that, by convention, is expressed by converting each byte into a decimal number (0 to 255) and separating the bytes with a period. For example, the PC Lube and Tune server is 130.132.59.234.

It is still possible for almost anyone to get assignment of a number for a small "Class C" network in which the first three bytes identify the network and the last byte identifies the individual computer. The author followed this procedure and was assigned the numbers 192.35.91.* for a network of computers at his house. Larger organizations can get a "Class B" network where the first two bytes identify the network and the last two bytes identify each of up to 64 thousand individual workstations. Yale's Class B network is 130.132, so all computers with IP address 130.132.*.* are connected through Yale.

The organization then connects to the Internet through one of a dozen regional or specialized network suppliers. The network vendor is given the subscriber network number and adds it to the routing configuration in its own machines and those of the other major network suppliers.

There is no mathematical formula that translates the numbers 192.35.91 or 130.132 into "Yale University" or "New Haven, CT." The machines that manage large regional networks or the central Internet routers managed by the National Science Foundation can only locate these networks by looking each network number up in a table. There is potentially thousands of Class B networks, and millions of Class C networks, but computer memory costs are low, so the tables are reasonable.

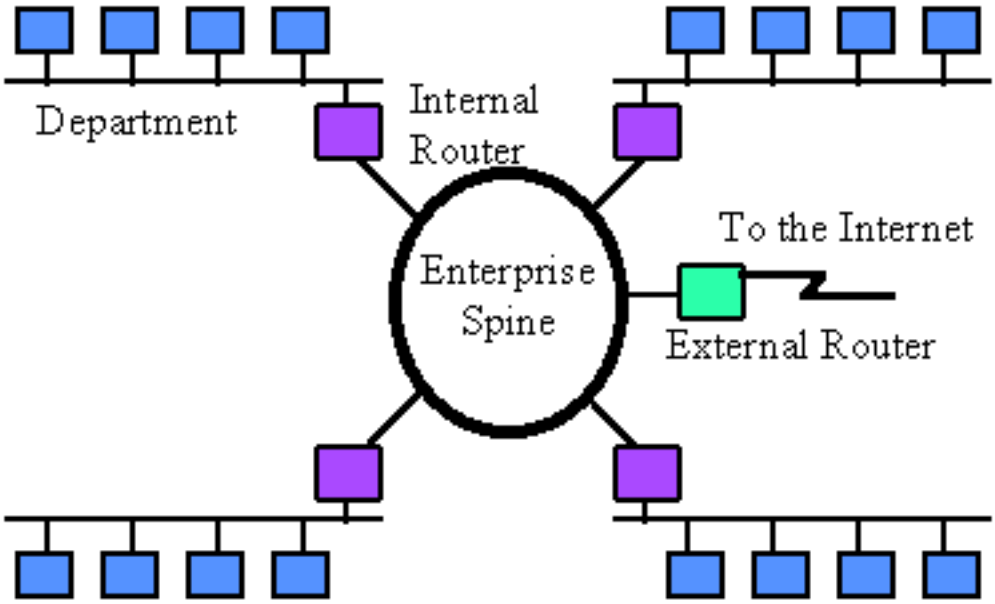
Customers that connect to the Internet, even customers as large as IBM, do not need to maintain any information on other networks. They send all external data to the regional carrier to which they subscribe, and the regional carrier maintains the tables and does the appropriate routing.

New Haven is in a border state split 50-50 between the Yankees and the Red Sox. In this spirit, Yale recently switched its connection from the Middle Atlantic regional network to the New England carrier. When the switch occurred, tables in the other regional areas and in the national spine had to be updated, so that traffic for 130.132 was routed through Boston instead of New Jersey. The large network carriers handle the paperwork and can perform such a switch given sufficient notice. During a conversion period, the university was connected to both networks so that messages could arrive through either path.

Subnets

Although the individual subscribers do not need to tabulate network numbers or provide explicit routing, it is convenient for most Class B networks to be internally managed as a much smaller and simpler version of the larger network organizations.

It is common to subdivide the two bytes available for internal assignment into a one byte department number and a one byte workstation ID.



The enterprise network is built using commercially available TCP/IP router boxes. Each router has small tables with 255 entries to translate the one byte department number into selection of a destination Ethernet connected to one of the routers. Messages to the PC Lube and Tune server (130.132.59.234) are sent through the national and New England regional networks based on the 130.132 part of the number. Arriving at Yale, the 59 departments ID selects an Ethernet connector in the C& IS building. The 234 select a particular workstation on that LAN. The Yale network must be updated as new Ethernets and departments are added, but it is noteffected by changes outside the university or the movement of machines within the department.

Uncertain Path

Every time a message arrives at an IP router, it makes an individual decision about where to send it next. There is concept of a session with a preselected path for all traffic. Consider a company with facilities in New York, Los Angeles, Chicago and Atlanta. It could build a network from four phone lines forming a loop (NY to Chicago to LA to Atlanta to NY). A message arriving at the NY router could go to LA via either Chicago or Atlanta. The reply could come back the other way.

How does the router make a decision between routes? There is no correct answer. Traffic could be routed by the "clockwise" algorithm (go NY to Atlanta, LA to Chicago). The routers could alternate, sending one message to Atlanta and the next to Chicago. More sophisticated routing measures traffic patterns and sends data through the least busy link.

If one phone line in this network breaks down, traffic can still reach its destination through a roundabout path. After losing the NY to Chicago line, data can be sent NY to Atlanta to LA to Chicago. This provides continued service though with degraded performance. This kind of recovery is the primary design feature of IP. The routers in NY and Chicago immediately detect the loss of the line, but somehow this information must be sent to the other nodes.

Otherwise, LA could continue to send NY messages through Chicago, where they arrive at a "dead end." Each network adopts some Router Protocol, which periodically updates the routing tables throughout the network with information about changes in route status.

If the size of the network grows, then the complexity of the routing updates will increase as will the cost of transmitting them. Building a single network that covers the entire US would be unreasonably complicated. Fortunately, the Internet is designed as a Network of Networks. This means that loops and redundancy are built into each regional carrier. The regional network handles its own problems and reroutes messages internally. Its Router Protocol updates the tables in its own routers, but no routing updates need to propagate from a regional carrier to the NSF spine or to the other regions (unless, of course, a subscriber switches permanently from one region to another).

Undiagnosed Problems

IBM designs its SNA networks to be centrally managed. If any error occurs, it is reported to the network authorities. By design, any error is a problem that should be corrected or repaired. IP networks, however, were designed to be robust. In battlefield conditions, the loss of a node or line is a normal circumstance. Casualties can be sorted out later on, but the network must stay up. So IP networks are robust. They automatically (and silently) reconfigure themselves when something goes wrong. If there is enough redundancy built into the system, then communication is maintained.

In 1975 when SNA was designed, such redundancy would be prohibitively expensive, or it might have been argued that only the Defense Department could afford it. Today, however, simple routers cost no more than a PC. However, the TCP/IP design that, "Errors are normal and can be largely ignored," produces problems of its own.

Data traffic is frequently organized around "hubs," much like airline traffic. One could imagine an IP router in Atlanta routing messages for smaller cities throughout the Southeast. The problem is that data arrives without a reservation. Airline companies experience the problem around major events, like the Super Bowl. Just before the game, everyone wants to fly into the city. After the game, everyone wants to fly out. Imbalance occurs on the network when something new gets advertised. Adam Curry announced the server at "mtv.com" and his regional carrier was swamped with traffic the next day. The problem is that messages come in from the entire world over high-speed lines, but they go out to mtv.com over what was then a slow speed phone line.

Occasionally a snowstorm cancels flights and airports fill up with stranded passengers. Many go off to hotels in town. When data arrives at a congested router, there is no place to send the overflow. Excess packets are simply discarded. It becomes the responsibility of the sender to retry the data a few seconds later and to persist until it finally gets through. This recovery is provided by the TCP component of the Internet protocol.

TCP was designed to recover from node or line failures where the network propagates routing table changes to all router nodes. Since the update takes some time, TCP is slow to initiate recovery. The TCP algorithms are not tuned to optimally handle packet loss due to traffic congestion. Instead, the traditional Internet response to traffic problems has been to increase the speed of lines and equipment in order to say ahead of growth in demand.

TCP treats the data as a stream of bytes. It logically assigns a sequence number to each byte. The TCP packet has a header that says, in effect, "This packet starts with byte 379642 and contains 200 bytes of data." The receiver can detect missing or incorrectly sequenced packets. TCP acknowledges data that has been received and retransmits data that has been lost. The TCP design means that error recovery is done end-to-end between the Client and Server machine. There is no formal standard for tracking problems in the middle of the network, though each network has adopted some ad hoc tools.

Need to Know

There are three levels of TCP/IP knowledge. Those who administer a regional or national network must design a system of long distance phone lines, dedicated routing devices, and very large configuration files. They must know the IP numbers and physical locations of thousands of subscriber networks. They must also have a formal network monitor strategy to detect problems and respond quickly.

Each large company or university that subscribes to the Internet must have an intermediate level of network organization and expertise. A half dozen routers might be configured to connect several dozen departmental LANs in several buildings. All traffic outside the organization would typically be routed to a single connection to a regional network provider.

However, the end user can install TCP/IP on a personal computer without any knowledge of either the corporate or regional network. Three pieces of information are required:

1. The IP address assigned to this personal computer
2. The part of the IP address (the subnet mask) that distinguishes other machines on the same LAN (messages can be sent to them directly) from machines in other departments or elsewhere in the world (which are sent to a router machine)
3. The IP address of the router machine that connects this LAN to the rest of the world.

In the case of the PCLT server, the IP address is 130.132.59.234. Since the first three bytes designate this department, a "subnet mask" is defined as 255.255.255.0 (255 is the largest byte value and represents the number with all bits turned on). It is a Yale convention (which we recommend to everyone) that the router for each department have station number 1 within the department network. Thus the PCLT router is 130.132.59.1. Thus the PCLT server is configured with the values:

- My IP address: 130.132.59.234
- Subnet mask: 255.255.255.0
- Default router: 130.132.59.1

The subnet mask tells the server that any other machine with an IP address beginning 130.132.59.* is on the same department LAN, so messages are sent to it directly. Any IP address beginning with a different value is accessed indirectly by sending the message through the router at 130.132.59.1 (which is on the departmental LAN).

File Transfer (FTP and SFTP)

One of the original intended uses of the Internet was to transfer files from one machine to another in a simple, secure, and reliable fashion, which was needed by scientific researchers. Today, file transfers represent downloads of music content, reports, and other data sets from other computer systems to a PC-based client. Until 1995, the majority of Internet traffic was file transfers. With all of this need, a protocol was necessary so that two computers could agree on how to send and receive data. As such, FTP is one of the older protocols.

FTP

FTP is an application-level protocol that operates over a wide range of lower level protocols. FTP is embedded in most operating systems and provides a method of transferring files from a sender to a receiver. Most FTP implementations are designed to operate both ways, sending and receiving, and can enable remote file operations over a TCP/IP connection. FTP clients are used to initiate transactions and FTP servers are used to respond to transaction requests. The actual request can be either to upload (send data from client to server) or download (send data from server to client).

Clients for FTP on a PC can range from an application program to the command line ftp program in Windows/DOS to most browsers. To open an FTP data store in a browser, you can enter **ftp://url** in the browser's address field to indicate that you want to see the data associated with the URL via an FTP session—the browser handles the details. File transfers via FTP can be either binary or in text mode, but in either case, they are in plaintext across the network.

Blind FTP (Anonymous FTP)

To access resources on a computer, an account must be used to allow the operating system level authorization function to work. In the case of an FTP server, you may not wish to control who gets the information, so a standard account called anonymous exists.

This allows unlimited public access to the files and is commonly used when you want to have unlimited distribution. On a server, access permissions can be established to allow only downloading or only uploading or both, depending on the system's function. As FTP can be used to allow anyone access to upload files to a server, it is considered a security risk and is commonly implemented on specialized servers isolated from other critical functions. As FTP servers can present a security risk, they are typically not permitted on workstations and are disabled on servers without need for this functionality.

HTTPS

HTTPS (HTTP over SSL or HTTP Secure) is the use of Secure Socket Layer (SSL) or Transport Layer Security (TLS) as a sub-layer under regular HTTP application layering. HTTPS encrypts and decrypts user page requests as well as the pages that are returned by the Web server. The use of HTTPS protects against eavesdropping and man-in-the-middle attacks. Netscape developed HTTPS.

HTTPS and SSL support the use of X.509 digital certificates from the server so that, if necessary, a user can authenticate the sender. Unless a different port is specified, HTTPS uses port 443 instead of HTTP port 80 in its interactions with the lower layer, TCP/IP.

Suppose you visit a Web site to view their online catalog. When you're ready to order, you will be given a Web page order form with a Uniform Resource Locator (URL) that starts with https://. When you click "Send," to send the page back to the catalog retailer, your browser's HTTPS layer will encrypt it. The acknowledgement you receive from the server will also travel in encrypted form, arrive with an https:// URL, and be decrypted for you by your browser's HTTPS sub-layer.

The effectiveness of HTTPS can be limited by poor implementation of browser or server software or a lack of support for some algorithms. Furthermore, although HTTPS secures data as it travels between the server and the client, once the data is decrypted at its destination, it is only as secure as the host computer. According to security expert Gene Spafford, that level of security is analogous to "using an armored truck to transport rolls of pennies between someone on a park bench and someone doing business from a cardboard box."

HTTPS is not to be confused with S-HTTP, a security-enhanced version of HTTP developed and proposed as a standard by EIT.

SFTP

FTP operates in a plaintext mode, so an eavesdropper can observe the data being passed. If confidential transfer is required, Secure FTP (SFTP) utilizes both the Secure Shell (SSH) protocol and FTP to accomplish this task. SFTP is an application program that encodes both the commands and the data being passed and requires SFTP to be on both the client and the server. SFTP is not interoperable with standard FTP—the encrypted commands cannot be read by the standard FTP server program. To establish SFTP data transfers, the server must be enabled with the SFTP program, and then clients can access the server provided they have the correct credentials. One of the first SFTP operations is the same as that of FTP: an identification function that uses a username and an authorization function that uses a password. There is no anonymous SFTP account by definition, so access is established and controlled from the server using standard access control lists (ACLs), IDs, and passwords.

SCP

Several heavily used Internet applications such as FTP, GOPHER, and HTTP use a protocol model in which every transaction requires a separate TCP connection. Since clients normally issue multiple requests to the same server, this model is quite inefficient, as it incurs all the connection start up costs for every single request.
SCP is a simple protocol, which lets a server and client have multiple conversations over a single TCP connection. The protocol is designed to be simple to implement, and is modeled after TCP.

SCP's main service is dialogue control. This service allows either end of the connection to establish a virtual session over a single transport connection. SCP also allows a sender to indicate message boundaries, and allows a receiver to reject an incoming session.

Protocol Operation

Session ID allocation

Each session is allocated a session identifier. Session Identifiers below 1024 are reserved. Session IDs allocated by clients are even; those allocated by servers, odd.

Session establishment
A session is established by setting the SYN bit in the first message sent on that channel.

Graceful release
Sending a message with the FIN bit set ends a session. Each end of a connection may be closed independently.

Disgraceful release
Sending a message with the RST bit set may terminate a session. All pending data for that session should be discarded

Message boundaries
Sending a message with the PUSH bit set marks a message boundary. The boundary is set at the final octet in this message, including that octet.

ICMP
Routers, intermediary devices, or hosts to communicate updates or error information to other routers, intermediary devices, or hosts use iCMPs.

Each ICMP message contains three fields that define its purpose and provide a checksum. They are TYPE, CODE, and CHECKSUM fields. The TYPE field identifies the ICMP message, the CODE field provides further information about the associated TYPE field, and the CHECKSUM provides a method for determining the integrity of the message.

The TYPES defined are:

TYPE	Description
0	Echo Reply
3	Destination Unreachable
4	Source Quench
5	Redirect Message
8	Echo Request
11	Time Exceeded
12	Parameter Problem
13	Timestamp Request
14	Timestamp Reply
15	Information Request (No Longer Used)
16	Information Reply (No Longer Used)
17	Address Mask Request
18	Address Mask Reply

Echo Request & Echo Reply
This is the ICMP most used to test IP connectivity commonly known as PING. The Echo Request ICMP will have a Type field of 8 and a Code field of 0. Echo Replies have a Type field of 0 and a Code field of 0.

Destination Unreachable
When a packet is undeliverable, a Destination Unreachable, Type 3, ICMP is generated. Type 3 ICMPs can have a Code value of 0 to 15:
Type 3
Code

Value	Description
0	Network Unreachable
1	Host Unreachable
2	Protocol Unreachable
3	Port Unreachable
4	Fragmentation needed and DF (Don't Fragment) set
5	Source route failed
6	Destination Network unknown
7	Destination Host unknown
8	Source Host isolated
9	Communication with Destination Network Administratively Prohibited
10	Communication with Destination Host Administratively Prohibited
11	Network Unreachable for Type Of Service
12	Host Unreachable for Type Of Service
13	Communication Administratively Prohibited by Filtering
14	Host Precedence Violation
15	Precedence Cutoff in Effect

Source Quench
An ICMP Source Quench message has a Type field of 4 and Code 0. Source Quench messages are sent when the destination is unable to process traffic as fast as the source is sending it. The Source Quench ICMP tells the source to cut back the rate at which it is sending data. The destination will continue to generate Source Quench ICMPs until the source is sending at an acceptable speed.

Redirect Message
An intermediary device will generate an ICMP Redirect Message when it determines that a route being requested can be reached either locally or through a better path.

Time Exceeded
If a router or host discards a packet due to a time-out, it will generate a Time Exceeded Type 11 ICMP. The Time Exceeded ICMP will have a Code value of either 0 or 1. A Code 0 is generated when the hop count of a datagram is exceeded and the packet is discarded. A Code 1 is generated when the reassemble of a fragmented packet exceeds the time-out value.

Parameter Problem
When an intermediary device or host discards a datagram due to inability to process, an ICMP 12 is generated. Common causes of this ICMP are corrupt header information or missing options. If the reason for the ICMP is a required missing option, the ICMP will have a Code value of 1. If the Code value is 0, the Pointer field will contain the octet of the discarded datagram's header where the error was detected.

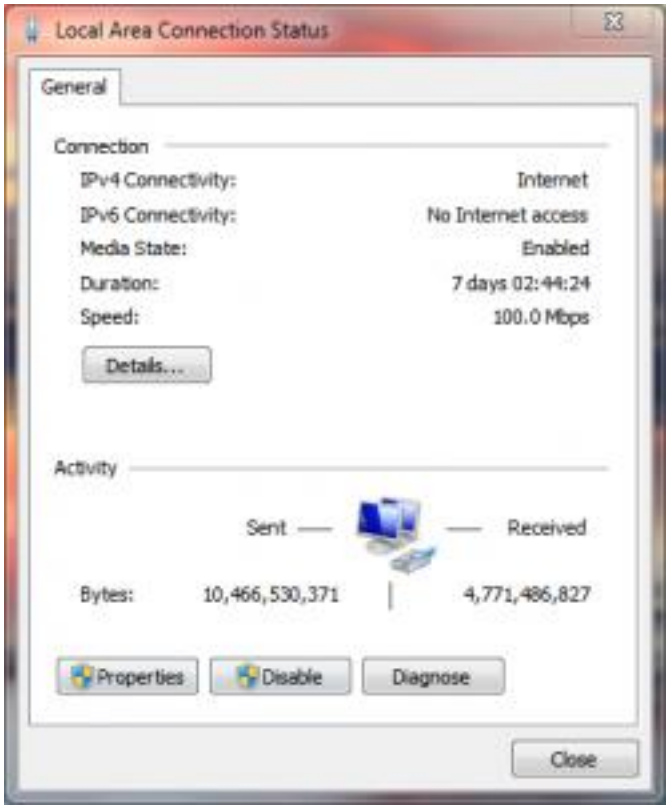
Timestamp Request & Timestamp Reply
Timestamp Request and Timestamp Reply is a rudimentary method for synchronizing the time maintained on different devices. The Request has a Type field of 13 and the Reply is Type 14. This method for time synchronization is crude and unreliable. Therefore, it is not heavily used.

Information Request & Information Reply
These ICMP types were originally designed to allow a booting host to discover an IP address. This method is obsolete and is no longer used. Most common methods for IP address discovery are BOOTP (bootstrap protocol) and DHCP (dynamic host configuration protocol). BOOTP is defined by RFC1542, and DHCP is defined by RFC1541.

Address Mask Request & Address Mask Reply
A booting computer to determine the subnet mask in use on the local network uses the Address Mask Request ICMP Type 17. An intermediary device or computer acting as an intermediary device will reply with a Type 18 ICMP Address Mask Reply ICMP.

IPv4 vs. IPv6

The "I" and "P" in "IPv" stands for "Internet Protocol" which directly refers to the communication protocol, or packet transfer procedure of the internet.
Every device that connects to the Internet uses a unique address called an IP address, which works very similar to a home/location address. Pieces of data, called "packets", are transferred via the Internet between machines, which in turn gives us the fully functioning interior workings of the online community. In order for two machines, or devices to communicate via the internet, they must transfer these "packets" of data back and forth. Unfortunately the data "packets" can not be transferred if the devices do not each have their own unique address.



Think of it basically as a home address. You can't send a mail correctly if you don't list a proper return address, because basically if the mail doesn't reach its destination it must have a way of returning back to you. Also, the mail receiver would have no possible way of responding considering they have no idea what address the should reply to.

While the Internet does not necessarily return data "packets" that don't reach their destination, like undelivered mail, proper use or protocol requires two devices to have unique addresses to even begin communications.
The "v" and number ("4″ or "6″) in "IPv4 vs IPv6″ refers to the related protocol version number. "IPv4″ is of course "Internet Protocol version 4″, and "IPv6″ is subsequently "Internet Protocol version 6″.

IPv4 is of course the older, more supported version of the internet address procedure. But ultimately, there are no longer any free IPv4 addresses, meaning all of them have been occupied or taken up. What does this mean exactly?

In a general sense, there will no longer be any alternative IPv4 addresses, directly meaning they will all be occupied and new users will not be able to venture into cyberspace. Although the realistic situation is not quite as dire.

Queue in IPv6, the latest Internet Protocol or address procedure. The older IPv4 only supports a maximum 32 bit internet address, which translates to 2^32 IP addresses available for assignment (about 4.29 billion total). IPv6 utilizes 128 bit web addresses, allowing a maximum 2^128 available addresses:

340,282,366,920,938,000,000,000,000,000,000,000,000,000; which if you couldn't already tell is a very big number.

So basically the IPv4 protocol has run out of available addresses, which is why most websites or internet servers are adopting the newer IPv6 protocol. In most cases, the two versions are compatible. This contrast between the two protocol versions is exactly what' s being referred to when "IPv4 vs IPv6″ is mentioned.

Ports

Ports identify how a communication process occurs. Ports are special addresses that allow communication between hosts. A port number is added from the origi- nator, indicating which port to communicate with on a server. If a server has a port defined and available for use, it will send back a message accepting the request. If the port isn't valid, the server will refuse the connection. The Internet Assigned Numbers Authority (IANA) has defined a list of ports called well-known ports.

A port address or number is nothing more than a bit of additional information added either to the TCP or UDP message. This information is added in the header of the packet. The layer below it encapsulates the message with its header.

Many of the services you'll use in the normal course of using the Internet use the TCP port numbers identified in following table. The other table identifies some of the more common, well-known UDP ports. You will note that some services use both TCP and UDP ports, whereas many use only one or the other.

TCP Port Number	Service
20	FTP (data channel)
21	FTP (Control Channel)
22	SSH and SCP
23	Telnet
25	SMTP
49	TACACS authentication service
80	HTTP (used for world wide web)
110	POP3
115	SFTP
119	NNTP
137	NetBIOS name service
138	NetBIOS datagram service
143	IMAP
389	LDAP
443	HTTPS (used for secure web connections)
989	FTPS (data channel)
3389	MS WBT Server

Well-known UDP ports

UDP Port Number	Service
22	SSH and SCP
49	TACACS authentication service
53	DNS name queries
69	Trivial File Transfer Protocol (TFTP)

80	HTTP (used for world wide web)
137	NetBIOS name Service
138	NetBIOS datagram service
139	NetBIOS session service
143	IMAP
161	SNMP
389	LDAP
989	FTPS (data channel)
990	FTPS (control channel)
3389	MS WBT server

The early documentation for these ports specified that ports below 1024 were restricted to administrative uses. However, enforcement of this restriction has been voluntary, and it is creating problems for computer security professionals. As you can see, each of these ports potentially requires different security considerations, depending on the application to which it's assigned. All of the ports allow access to your network; even if you establish a firewall, you must have these ports open if you want to provide email or web services.

OSI relevance
When discussing networking, most experts refer to the seven-layer OSI model—long considered the foundation for how networking protocols should operate. This model is the most common one used, and the division between layers is well defined.

TCP/IP precedes the creation of the OSI model. Although it carries out the same operations, it does so with four layers instead of seven.

TCP/IP Suite
The TCP/IP suite is broken into four architectural layers:

- Application layer
- Host-to-Host, or Transport layer
- Internet layer
- Network Access layer (also known as the Network Interface layer or the Link layer)

Computers using TCP/IP use the existing physical connection between the systems. TCP/IP doesn't concern itself with the network topology, or physical connections. The network controller that resides in a computer or host deals with the physical protocol, or topology. TCP/IP communicates with that controller and lets the controller worry about the network topology and physical connection.

In TCP/IP parlance, a computer on the network is a *host*. A host is any device connected to the network that runs a TCP/IP protocol suite, or stack. Figure 3.1 shows the four layers in a TCP/IP protocol stack. Note that this drawing includes the physical, or network topology. Although it isn't part of TCP/IP, the topology is essential to conveying information on a network.

The four layers of TCP/IP have unique functions and methods for accomplishing work. Each layer talks to the layers that reside above and below it. Each layer also has its own rules and capabilities.

The Application Layer
The Application layer is the highest layer of the suite. It allows applications to access services or protocols to exchange data. Most programs, such as web browsers, interface with TCP/IP at this level. The most commonly used Application layer protocols are as follows:

Hypertext Transfer Protocol Hypertext Transfer Protocol (HTTP) is the protocol usedfor web pages and the World Wide Web. HTTP applications use a standard language called Hypertext Markup Language (HTML). HTML files are normal text files that contain special coding that allows graphics, special fonts, and characters to be displayed by a web browser or other web-enabled applications. The default port is 80, and the URL begins with http://.

HTTP Secure HTTP Secure (HTTPS) is the protocol used for "secure" web pages that users should see when they must enter personal information such as credit card numbers, passwords, and other identifiers. It combines HTTP with SSL/TLS to provide encrypted communication. The default port is 443, and the URL begins with https:// instead of http://. Netscape originally created the protocol for use with their browser, and it became an accepted standard with RFC 2818.

File Transfer Protocol *File Transfer Protocol (FTP)* is an application that allows connections to FTP servers for file uploads and downloads. FTP is a common application that uses ports 20 and 21 by default. It is used to transfer files between hosts on the Internet but is inherently insecure. A number of options have been released to try to create a more secure protocol, including *FTP over SSL (FTPS)*, which adds support for SSL cryptography, and *SSH File Transfer Protocol (SFTP)*, which is also known as *Secure FTP*.

An alternative utility for copying files is *Secure Copy (SCP)*, which uses port 22 by default and combines an old remote copy program (RCP) from the first days of TCP/IP with SSH. On the opposite end of the spectrum from a security standpoint is the *Trivial File Transfer Protocol (TFTP)*, which can be configured to transfer files between hosts without any user interaction (unattended mode). It should be avoided anywhere there are more secure alternatives.

Simple Mail Transfer Protocol *Simple Mail Transfer Protocol (SMTP)* is the standard protocol for email communications. SMTP allows email clients and servers to communicate with each other for message delivery. The default port is 25.

Telnet *Telnet* is an interactive terminal emulation protocol. It allows a remote user to conduct an interactive session with a Telnet server. This session can appear to the client as if it were a local session.

Domain Name System *Domain Name System (DNS)* allows hosts to resolve hostnames to an Internet Protocol (IP) address. The default port used by name queries for this service is 53.

Remote Desktop Protocol The *Remote Desktop Protocol (RDP)* is becoming more common in the workplace, and it allows Windows-based terminal servers to run on port 3389 by default.

Simple Network Management Protocol *Simple Network Management Protocol (SNMP)* is a management tool that allows communications between network devices and a management

console. Most routers, bridges, and intelligent hubs can communicate using SNMP.

Post Office Protocol *Post Office Protocol (POP)* is a protocol used for receiving email. It enables the implementation of advanced features, and it is a standard interface in many email servers. The default port for version 3 (POP3) is 110. In its place, many systems now use the *Internet Message Access Protocol (IMAP)*, which uses port 143 by default. The primary difference between the two is that POP was originally created to move email to your client machine and not keep it on the server, whereas IMAP was intended to store the email on the server and allow you to access it from there. Although those remain default options, today you can configure POP not to delete from the server automatically and IMAP to do so. For this reason, most email providers allow you to use either POP or IMAP and even change between them.

The Host-to-Host or Transport Layer

The *Host-to-Host layer*, also called the *Transport layer*, provides the Application layer with session and datagram communications services. The *Transmission Control Protocol (TCP)* and *User Datagram Protocol (UDP)* operate at this layer. These two protocols provide a huge part of the functionality of the TCP/IP network:

TCP TCP is responsible for providing a reliable, one-to-one, connection-oriented ses- sion. TCP establishes a connection and ensures that the other end receives any packets sent. Two hosts communicate packet results with each other. TCP also ensures that packets are decoded and sequenced properly. This connection is persistent during the session. When the session ends, the connection is torn down.

UDP UDP provides an unreliable connectionless communication method between hosts. UDP is considered a best-effort protocol, but it's considerably faster than TCP. The ses- sions don't establish a synchronized session like the kind used in TCP, and UDP doesn't guarantee error-free communications. The primary purpose of UDP is to send small pack- ets of information. The application is responsible for acknowledging the correct reception of the data.

The Internet Layer

The *Internet layer* is responsible for routing, IP addressing, and packaging. The Internet layer protocols accomplish most of the behind-the-scenes work in establishing the ability to exchange information between hosts. The following is an explanation of the four standard protocols of the Internet layer:

Internet Protocol *Internet Protocol (IP)* is a routable protocol that is responsible for IP addressing. IP also fragments and reassembles message packets. IP only routes information; it doesn't verify it for accuracy. Accuracy checking is the responsibility of TCP. IP deter- mines if a destination is known and, if so, routes the information to that destination. If the destination is unknown, IP sends the packet to the router, which sends it on.

Address Resolution Protocol *Address Resolution Protocol (ARP)* is responsible for resolving IP addresses to Network Interface layer addresses, including hardware addresses. ARP can resolve an IP address to a *Media Access Control (MAC)* address. MAC addresses are used to identify hardware network devices, such as a network interface card (NIC).

Internet Control Message Protocol *Internet Control Message Protocol (ICMP)* provides maintenance and reporting functions. The Ping program uses it. When a user wants to test connectivity to another host, they can enter the PING command with the IP address, and the user's system will test connectivity to the other host's system. If connectivity is good, ICMP will return data to the originating host. ICMP will also report if a destination is unreachable. Routers and other network devices report path information between hosts with ICMP.

The Network Access Layer

The lowest level of the TCP/IP suite is the *Network Access (or Interface) layer*. This layer is responsible for placing and removing packets on the physical network through communications with the network adapters in the host. This process allows TCP/IP to work with virtually any type of network topology or technology with little modification. If a new physical network topology were installed—say, a 10 GB Fiber Ethernet connection— TCP/IP would only need to know how to communicate with the network controllerin order to function properly. TCP/IP can also communicate with more than one network topology simultaneously. This allows the protocol to be used in virtually any environment.

Section 1.5 -Given a scenario, troubleshoot security issues related to wireless networking.

WPA
WPA is Wi-Fi Protected Access, one of several popular standards for wireless network security. This WPA is not to be confused with Windows XP Product Activation, a separate technology that is also included with the Microsoft Windows operating system.

Before being able to use Wi-Fi WPA with Windows XP, you may need to upgrade one or more components of your network including the XP operating system and network adapters on some computers as well as the wireless access point.

Follow these instructions to set up WPA on Wi-Fi networks having Windows XP clients.

Difficulty: Average

Time Required: 30 minutes

Here's How:

1. Verify each Windows computer on the network is running Windows XP Service Pack 1 (SP1) or greater. WPA cannot be configured on older versions of Windows XP or older versions of Microsoft Windows.

2. For any Windows XP computer running SP1 or SP2, update the operating system to XP Service Pack 3 or newer for best WPA/WPA2 support.

   XP Service Pack 1 computers do not support WPA by default and cannot support WPA2. To upgrade an XP SP1 computer to support WPA (but not WPA2), either

- install the Windows XP Support Patch for Wi-Fi Protected Access from Microsoft, or
- upgrade the computer to XP SP2

XP Service Pack 2 computers by default support WPA but not WPA2. To upgrade an XP SP2 computer to also support WPA2, install the Wireless Client Update for Windows XP SP2 from Microsoft.

1. Verify your wireless network router (or other access point) supports WPA. Because some older wireless access points do not support WPA, you many need to replace yours. If necessary, upgrade the firmware on the access point according to the manufacturer's directions to enable WPA on it.

2. Verify each wireless network adapter also supports WPA. Obtain a device driver upgrade from the adapter manufacturer if necessary. Because some wireless network adapters cannot support WPA, you may need to replace them.

3. On each Windows computer, verify that its network adapter is compatible with the Wireless Zero Configuration (WZC) service. Consult the adapter's product documentation, manufacturer's Web site, or appropriate customer service department for details on WZC. Upgrade the network adapter driver and configuration software to support WZC on clients if necessary.

4. Apply compatible WPA settings on each Wi-Fi device. These settings cover network encryption and authentication.

The WPA encryption keys (or passphrases) chosen must match exactly between devices.

For authentication, two versions of Wi-Fi Protected Access exist called WPA and WPA2. To run both versions on the same network, ensure the access point is configured for WPA2 mixed mode. Otherwise, you must set all devices to WPA or WPA2 mode exclusively.

Wi-Fi products use a few different naming conventions to describe types of WPA authentication. Set all equipment to use either Personal/PSK or Enterprise/*EAP options.

WPA2

In July 2004, the IEEE approved the full IEEE 802.11i specification, which was quickly followed by a new interoperability testing certification from the WiFi Alliance known as WPA2. WPA2 is based on the Robust Security Network (RSN) mechanism, which provided support for all of the mechanisms available in WPA, as well as:

- Strong encryption and authentication support for infrastructure and ad-hoc networks (WPA is limited to infrastructure networks);

- Reduced overhead in key derivation during the wireless LAN authentication exchange;

- Support for opportunistic key caching to reduce the overhead in roaming between access points;

- Support for pre-authentication, where a station completes the IEEE 802.1X authentication exchange before roaming;

- Support for the CCMP (Counter Mode with Cipher Block Chaining Message Authentication Code Protocol) encryption mechanism based on the Advanced Encryption Standard (AES) cipher as an alternative to the TKIP protocol.

As of March 2006, the WPA2 certification became mandatory for all new equipment certified by the Wi-Fi Alliance, ensuring that any reasonably modern hardware will support both WPA and WPA2.

By leveraging the RC4 cipher (also used in the WEP protocol), the IEEE 802.11i task group was able to improve the security of legacy networks with TKIP while the IEEE 802.11i amendment was completed. It is important to note, however, that TKIP was designed as an interim solution for wireless security, with the goal of providing sufficient security for 5 years while organizations transitioned to the full IEEE 802.11i security mechanism. While there have not been any catastrophic weaknesses reported in the TKIP protocol, organizations should take this design requirement into consideration and plan to transition WPA networks to WPA2 to take advantage of the benefits provided by the RSN architecture.

WEP
The privacy protocol specified in IEEE 802.11 to provide wireless LAN users protection against casual eavesdropping, WEP refers to the intent to provide a privacy service to wireless LAN users similar to that provided by the physical security inherent in a wired LAN.

When WEP is active in a wireless LAN, each 802.11 packet is encrypted separately with an RC4 cipher stream generated by a 64-bit RC4 key. This key is composed of a 24-bit initialization vector (IV) and a 40-bit WEP key. The encrypted packet is generated with a bitwise exclusive OR (XOR) of the original packet and the RC4 stream. The IV is chosen by the sender and can be changed periodically so every packet won't be encrypted with the same cipher stream. The IV is sent in the clear with each packet. An additional 4-byte Integrity Check Value (ICV) is computed on the original packet and appended to the end. The ICV (be careful not to confuse this with the IV) is also encrypted with the RC4 cipher stream.

WEP has been widely criticized for a number of weaknesses:

- **Key management and key size**
  Key management is not specified in the WEP standard; without interoperable key management, keys will tend to be long-lived and of poor quality. Most wireless networks that use WEP have one single WEP key shared between every node on the network. Access points and client stations must be programmed with the same WEP key. Since synchronizing the change of keys is tedious and difficult, keys are seldom

changed. Also, the 802.11 standard does not specify any WEP key sizes other than 40 bits.

- **The IV is too small**

  WEP's IV size of 24 bits provides for 16,777,216 different RC4 cipher streams for a given WEP key, for any key size. Remember that the RC4 cipher stream is XOR-ed with the original packet to give the encrypted packet that is transmitted, and the IV is sent in the clear with each packet. The problem is IV reuse. If the RC4 cipher stream for a given IV is found, an attacker can decrypt subsequent packets that were encrypted with the same IV or can forge packets.

- **Weakness: The ICV algorithm is not appropriate**
  The WEP ICV is based on CRC-32, an algorithm for detecting noise and common errors in transmission. CRC-32 is an excellent checksum for detecting errors, but an awful choice for a cryptographic hash. Better-designed encryption systems use algorithms such as MD5 or SHA-1 for their ICVs.
- **Authentication messages can be easily forged**

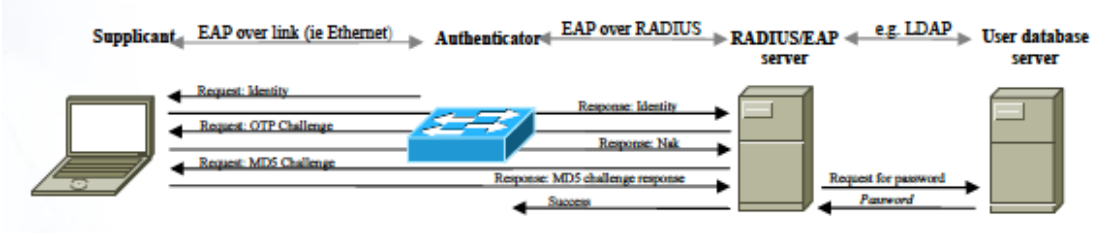
EAP
The Extensible Authentication Protocol (EAP) is best considered as a framework for transporting authentication protocols, rather than as an authentication protocol itself. EAP can be used for authenticating dial-up and VPN connections, and also Local Area Network (LAN) ports in conjunction with IEEE 802.1X.

In EAP, the party demanding proof of authentication is called the authenticator and the party being authenticated is called the supplicant. EAP defines four types of packet: request, response, success and failure. The authenticator issues request packets and they solicit a response packet from the supplicant. Any number of request-response exchanges may be used to complete the authentication. If the authentication is successful, a success packet is sent to the supplicant; if not, a failure packet is sent.

The basic EAP packet format is simple. A type field indicates the type of packet, such as a response or a request. An Identifier field is used to match requests and responses. Response and request packets have two further fields. The first, confusingly called ‘type’, indicates the type of data being transported (such as an authentication protocol), and the second, type-data, consists of that data. Note that EAP method is synonymous with type, and both are used frequently.

The EAP specification defines three ‘basic’ authentication EAP types (MD5-Challenge, OTP and GTC) and three non-authentication types (Identity, Nak,* and Notification). The three ‘basic’ authentication types are not considered secure for typical use, particularly in wireless environments. The authenticator to request the user name claimed by the supplicant uses the Identity type, and is typically the first packet transmitted. The Nak type is used by the peer to indicate that a type proposed by the authenticator is unacceptable (for example, the authenticator has proposed an authentication protocol that is unsupported by the peer, or policy forbids its use). If this happens then the authenticator may choose to try another, thereby allowing supplicant and authenticator to negotiate a mutually acceptable authentication protocol. The Notification type, which is rarely used, returns a message that must be displayed to the user.

Finally, EAP permits pass-through authentication. This allows the authenticator to forward all responses, using the RADIUS protocol, to a remote EAP server (in practice, most RADIUS servers also understand EAP). This server assumes the role of the authenticator for the remainder of the EAP session, and attempts to authenticate the supplicant against a user database server. Pass-through authentication, therefore, permits centralized management of authentication against large numbers of authenticators. Another advantage is that the authenticator does not need to support the type negotiated by the peer and the EAP server. An example EAP exchange is shown in the figure below. The peer refuses OTP authentication, but agrees to MD5-Challenge and is authenticated.

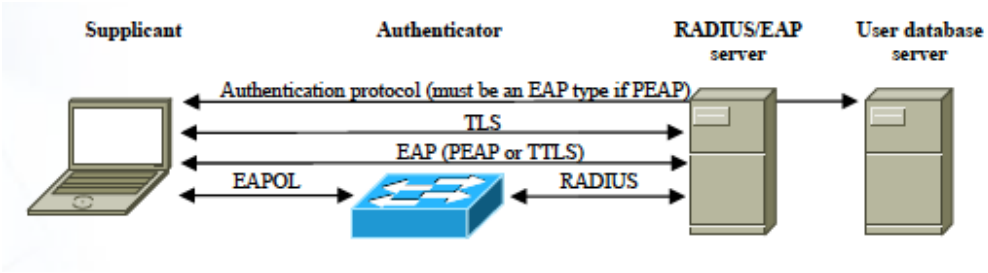


EAP Types TLS, TTLS and PEAP

As previously mentioned, the 'basic' authentication types should not be used. They do not provide sufficient protection for use on a shared network and, in particular, do not allow negotiation of the keying material required for IEEE 802.11 wireless LAN encryption. Consequently, a number of more secure types have been developed. Of these, only three have been widely implemented: TLS, TTLS and PEAP.

The TLS EAP type is based on the Transport Layer Security (TLS) protocol, which uses public key cryptography for authentication and negotiation of keys that can beused to encrypt data. TLS is also the protocol used for securing HTTPS. The main difference is that HTTPS is transported over TCP, whereas
EAP TLS is transported over the EAP session between the supplicant and EAP server. As in HTTPS, the supplicant authenticates the server's identity using a locally stored root certificate. However, unlike most HTTPS transactions, EAP TLS uses a user certificate to authenticate the supplicant to the server.

This means TLS can only be used by organizations with a Certificate Authority (CA) that issues user certificates; as such, although it offers excellent security, it is not widely deployed. Instead, two further EAP types, Protected EAP (PEAP) and Tunneled TLS (TTLS), work around this problem. Both of these types also use TLS for server authentication and encryption, but avoid the need for user certificates by using a second authentication protocol between the supplicant and the server that is protected by theTLS encryption. This is very similar to conventional HTTPS authentication, where the user's plain-text credentials are protected by TLS. The main difference between the types is that PEAP can only protect other EAP types, whereas TTLS can protect almost any authentication protocol. An overview of the protocol layering is shown in the figure below.



LEAP

The Lightweight Extensible Authentication Protocol (LEAP) is a proprietary wireless LAN authentication method developed by Cisco Systems. Important features of LEAP are dynamic WEP keys and mutual authentication (between a wireless client and a RADIUS server). LEAP allows for clients to re-authenticate frequently; upon each successful authentication, the clients acquire a new WEP key (with the hope that the WEP keys don't live long enough to be cracked). LEAP may be configured to use TKIP instead of dynamic WEP.

Cisco LEAP, similar to WEP, has had well-known security weaknesses since 2003 involving offline password cracking. LEAP uses a modified version of MS-CHAP, an authentication protocol in which user credentials are not strongly protected. Stronger authentication protocols employ a salt to strengthen the credentials against eavesdropping during the authentication process. Cisco's response to the weaknesses of LEAP suggests that network administrators either force users to have stronger, more complicated passwords or move to another authentication protocol also developed by Cisco, EAP-FAST, to ensure security. Automated tools like ASLEAPdemonstrate the simplicity of getting unauthorized access in networks protected by LEAP implementations.

MAC filter

Most Wi-Fi access points and routers ship with a feature called hardware or MAC address filtering. The manufacturer normally turns “off” this feature, because it requires a bit of effort to set up properly. However, to improve the security of your Wi-Fi LAN (WLAN), strongly consider enabling and using MAC address filtering.

Without MAC address filtering, any wireless client can join (authenticate with) a Wi-Fi network if they know the network name (also called the SSID) and perhaps a few other security parameters like encryption keys. When MAC address filtering is enabled, however, the access point or router performs an additional check on a different parameter. Obviously the more checks that are made, the greater the likelihood of preventing network break-ins.

To set up MAC address filtering, you as a WLAN administrator must configure a list of clients that will be allowed to join the network. First, obtain the MAC addresses of each client from its operating system or configuration utility. Then, they enter those addresses into a configuration screen of the wireless access point or router. Finally, switch on the filtering option.

Once enabled, whenever the wireless access point or router receives a request to join with the WLAN, it compares the MAC address of that client against the administrator's list. Clients on the list authenticate as normal; clients not on the list are denied any access to the WLAN.

MAC addresses on wireless clients can't be changed as they are burned into the hardware. However, some wireless clients allow their MAC address to be "impersonated" or "spoofed" in software. It's certainly possible for a determined hacker to break into your WLAN by configuring their client to spoof one of your MAC addresses. Although MAC address filtering isn't bulletproof, still it remains a helpful additional layer of defense that improves overall Wi-Fi network security.

Do not confuse MAC address filtering with content filtering. Content filtering on a wireless access point or router allows administrators to maintain a list of Web site URLs or addresses that should not be accessed from the home WLAN.

SSID broadcast
SSID (service set identifier) is a function performed by an Access Point that transmits its name so that wireless stations searching for a network connection can 'discover' it. It's what allows your wireless adapter's client manager program or Windows XP's built-in wireless software to give you a list of the Access Points in range.

Having SSID broadcast disabled essentially makes your Access Point invisible unless a wireless client already knows the SSID, or is using tools that monitor or 'sniff' traffic from an AP's associated clients.

Using the default SSIDs poses a security risk even if the AP is not broadcasting it, here are some standard ones that can possibly be probed by potenital attackers:

- 101 (3Com)
- Compaq (Compaq)
- compex (Compex)
- Default SSID
- intel (Intel)
- linksys (Linksys)
- RoamAbout Default Network Name (Lucent/Cabletron)
- tsunami (Cisco)
- Wireless
- WLAN (Addtron)

Note that turning off SSID broadcast does not effectively protect the network from attacks, as network-monitoring tools like Kismet and airodump-ng can still easily find the SSID, often within minutes.

TKIP
Temporal Key Integrity Protocol (TKIP), as defined by the IEEE 802.11i specification, addresses the encryption part of the wireless security equation. (A different part of 802.11i addresses the per-message integrity problem) TKIP was designed with a very difficult constraint in place: it had to operate on existing hardware, and therefore it could not require computationally advanced encryption.

TKIP is a "wrapper" that goes around the existing WEP encryption. TKIP comprises the same encryption engine and RC4 algorithm defined for WEP. However, the key used for

encryption in TKIP is 128 bits long. This solves the first problem of WEP: a too-short key length.

An important part of TKIP is that it changes the key used for each packet. This is the "Temporal" part of the picture. The key is created by mixing together a combination of things, including a base key (called a Pairwise Transient Key in TKIP parlance), the MAC address of the transmitting station, and the serial number for the packet. The mixing operation is designed to put a minimum demand on the stations and access points, yet have enough cryptographic strength so that it cannot easily be broken.

Each packet transmitted using TKIP has a unique 48-bit serial number that is incremented every time a new packet is transmitted and used both as the Initialization Vector and part of the key. Putting a sequence number into the key ensures that the key is different for every packet. This solves another problem of WEP, called "collision attacks," which can occur when the same key is used for two different packets. With different keys, there are no collisions.

Having the serial number of the packet also be the initialization vector helps to reduce yet another WEP problem, called "replay attacks." Because a 48-bit sequence number will take thousands of years to repeat itself, no one can replay old packets from a wireless connection---they will be detected as out of order because the sequence numbers won't be right.

The last, and most important, piece that is mixed into the TKIP key is the base key. Without a way to generate unique base keys, TKIP would solve many of WEP's problems, but not its worst one: the constant reuse of a well-known key by everyone on the wireless LAN. To deal with this, TKIP generates the base key that is mixed into the per-packet key. Each time a wireless station associates to an access point, a new base key is created. This base key is built by hashing together a special session secret with some random numbers (called nonce) generated by the access point and the station as well as the MAC address of the access point and the station. With 802.1X authentication, the session secret is unique and transmitted securely to the station by the authentication server; when using TKIP with pre-shared keys, the session secret is the same for everyone and never changes---hence the vulnerability of using TKIP with pre-shared keys.

Captive portals

Most public networks, including Wi-Fi hotspots, use a captive portal, which requires users to agree to some condition before they use the network or Internet. The condition could be to agree to the acceptable use policy, payment charges for the time they are using the network, and so forth.

One of the most popular implementations of captive portals is a Cisco application in their Identity Services Engine. However, there have been vulnerabilities identified with it, which allow attackers to intercept cleartext values:

Antenna types

Just as important as antenna placement is the type of antenna used. The default antenna on many (but not all) APs can be replaced to increase or decrease transmission range. The proper antenna can work around obstacles, minimize the effects of interference, increase signal strength, and focus the transmission (which can increase signal speed).

The antenna can be completely internal on an AP, or it can consist of one, two, or three external poles.

An *omnidirectional* antenna is designed to provide a 360-degree pattern and an even signal in all directions, so you usually want to locate the AP in the middle of the area to be covered. A *directional* antenna, on the other hand, forces the signal in one direction, and since it is focusing the signal, it can cover a greater distance with a stronger signal.

All antennas are rated in terms of *gain value*, which is expressed in dBi numbers. A wire- less antenna advertised with a 20 dBi would be 20 times stronger than the base of 0 dBi. As a general rule, every 3 dB added to an antenna effectively doubles the power output.

Site surveys

An additional aspect of wireless systems is the site survey. Site surveys involve listening in on an existing wireless network using commercially available technologies. Doing so allows intelligence, and possibly data capture, to be performed on systems in your wireless network.

The term site survey initially meant determining whether a proposed location was free from interference. When used by an attacker, a site survey can determine what types of systems are in use, the protocols used, and other critical information about your network. It's the primary method used to gather information about wireless networks. Virtually all wireless networks are vulnerable to site surveys.

As for interference, it can be unintentional (caused by other devices in the vicinity, for example) or intentional. When it is intentional, then it is referred to as jamming, as the intent is to jam the signal and keep the legitimate device from communicating.
If wirelesses APs are installed in a building, the signals will frequently radiate past the inside of the building, and they can be detected and decoded outside the building using inexpensive equipment. The term war driving refers to driving around town with a laptop looking for APs to communicate with. The network card on the intruder's laptop is set to promiscuous mode, and it looks for signals coming from anywhere. After intruders gain access, they may steal Internet access or corrupt your data.

Once weaknesses have been discovered in a wireless network, war chalking can occur. War chalking involves those who discover a way into the network leaving signals (oftenwritten in chalk) on, or outside, the premise to notify others that vulnerability exists there. The marks can be on the sidewalk, the side of the building, a nearby signpost, and so on.

Topic 2 - Compliance and Operational Security

Section 2.1-Explain the importance of risk related concepts.

Control types

To prepare for the certification exam, it often helps to use analogies to put topics in con- text. In light of that, consider a residential home this author owns in the middle of town. I grow prized tomato plants in the backyard, and it is very important to me that no one goes back there for fear that they might do something to harm the tomatoes. Thus, I implement the following controls:

Administrative: I establish a number of policies to keep the tomatoes safe:

Preventive: I instruct every member of my family that they are not to go into thebackyard and they are not to let anyone else go back there either.

Deterrent: I tell the kids that if I ever hear of any of them—or their friends—being the backyard, I will take away their allowance for month.

Detective: As a matter of routine, I want each member of the family to look out the window on a regular basis to see if anyone has wandered into the yard.

Compensating: Every member of the family is instructed on how to call the police the minute they see anyone in the yard.

Technical: Not trusting that the administrative controls will do the job without fail, I implement a number of technical controls:

Preventive: I put up a fence around the yard, and the door that leads out from the garage is locked.

Deterrent: "Beware of Dog" signs are posted all over the fence (although I have no dog).

Detective: Sensors are placed on the gate to trigger an alarm if the gate is opened.

Compensating: Triggered alarms turn on the backyard sprinklers at full volume to douse any intruder who wanders in.

These controls work in conjunction with one another to help keep individuals who should not be there out of the backyard and away from the tomatoes. Naturally, as the owner/administrator, I have the ability to override all of them as needed. I can ignore the warning signs, turn off the sprinklers, and get full access to the garden when I desire. The controls are not in place to hinder my access, but only to obstruct and prevent others from accessing the yard.

False positives/False negatives

False positives are events that aren't really incidents. Event flagging is often based on established rules of acceptance (deviations from which are known as anomalies) and things such as attack signatures. If the rules aren't set up properly, normal traffic may set off an analyzer and generate an event. You don't want to declare an emergency unless you're sure that you have one. The opposite of a false positive is a false negative. With a false negative, you are not alerted to a situation when you should be alerted. In this case, you miss some- thing crucial and it slips right by.

Many IDSs trigger false positives when reporting incidents. False positives are events that aren't really incidents. Remember that an IDS is based on established rules of acceptance (deviations from which are known as anomalies) and attack signatures. If the rules aren't set up properly, normal traffic may set off the analyzer and generate an event. Be sure to double-

check your results because you don't want to declare a false emergency.

One problem that can occur with manual network monitoring is overload. Over time, a slow attack may develop that increases in intensity. Manual processes typically will adapt, and they may not notice the attack until it's too late to stop it. Personnel tend to adapt to changing environments if the changes occur over a long period of time.

An automated monitoring system, SUCH AS IDS, will sound the alarm when a certain threshold or activity level occurs.

When a suspected incident pops up, first responders are those individuals who must ascertain whether it truly is an incident or a false alarm. Depending on your organization, the first responder may be the main security administrator or it could consist of a team of network and system administrators.

Importance of policies in reducing risk

Privacy Policies
Privacy policies define what controls are required to implement and maintain the sanctity of data privacy in the work environment. For now, however, think of the privacy policy as a legal document that outlines how data collected is secured. Google endorses a great example: www.google.com/privacy/privacy-policy.html. It outlines exactly what information the company collects, privacy choices you have based on your account, potential information sharing of your data with other parties, security measures in place, and enforcement. The last paragraph of the policy should appear in every privacy policy and addresses the fact that the policy may change. The verbiage, as currently written, is succinct and clear: "Please note that this Privacy Policy may change from time to time. We will not reduce your rights under this Privacy Policy without your explicit consent. We will post any Privacy Policy changes on this page and, if the changes are significant, we will provide a more prominent notice (including, for certain services, email notification of Privacy Policy changes). We will also keep prior versions of this Privacy Policy in an archive for your review."

Acceptable Use Policies
Acceptable use policies (AUPs) describe how the employees in an organization can use company systems and resources, both software and hardware. This policy should also outline the consequences for misuse. In addition, the policy (also known as a *use policy*) should address the installation of personal software on company computers and the use of personal hardware such as USB devices. When portable devices are plugged directly into a machine, they bypass the network security measures (such as firewalls) and allow data to be copied in what is known as *pod slurping*. This can also be done if employee's start using free cloud drives instead, and that scenario should be addressed in the AUP.

Although a smartphone is a convenience for employees (they can now more easily receive and make personal calls at work), it can be a head- ache for the security administrator. Most smartphones can store files in the same way as any USB device, and they can be used to copy files to and from a workstation. Additionally, the camera feature on most phones makes it possible for a user to take pictures of things such as company documents, servers, and physical security implementation, among many other things that the company may not want to share. For this reason, most secure facilities have stringent restrictions on the presence of

smartphones within the vicinity.

Security Policies
Security policies define what controls are required to implement and maintain the security of systems, users, and networks. This policy should be used as a guide in system implementations and evaluations.

Mandatory Vacations
A *mandatory vacation policy* requires all users to take time away from work to refresh. As contradictory as it may seem, an employee who doesn't take their vacation time can be detrimental to the health, not only of the employee, but to the company's health as well. If the company becomes too dependent on one person, they can end up in a real bind if some- thing should happen to that person. Not only does mandatory vacation give the employee a chance to refresh, but it also gives the company a chance to make sure that others can fill in any gaps in skills and satisfies the need to have replication or duplication at all levels. Mandatory vacations also provide an opportunity to discover fraud.

Job Rotation
A *job rotation policy* defines intervals at which employees must rotate through positions. Similar in purpose to mandatory vacations, it helps to ensure that the company does not become too dependent on one person (who then has the ability to do enormous harm). Rotate jobs on a frequent enough basis so that you are not putting yourself—and your data—at the mercy of any one administrator. Just as you want redundancy in hardware, you want redundancy in abilities.
When one person fills in for another, such as for mandatory vacations, it provides an opportunity to see what the person is doing and potentially uncover any fraud.

Least Privilege
A *least privilege policy* should be used when assigning permissions. Give users only the permissions that they need to do their work and no more. For example, a temporary employee should never have the right to install software, a receptionist does not need the right to make backups, and so on. Every operating system includes the ability to limit users based on groups and individual permissions, and your company should adhere to the policy of always applying only those permissions users need and blocking those that they do not.

Succession Planning
Succession planning outlines those internal to the organization who have the ability to step into positions when they open up. By identifying key roles that cannot be left unfilled and associating internal employees who can step into those roles, you can groom those employees to make sure that they are up to speed when it comes time for them to fill those positions.

Risk calculation
For purposes of risk assessment, both in the real world and for the exam, you should familiarize yourself with a number of terms to determine the impact an event could have:

- ALE is the annual loss expectancy value. This is a monetary measure of how much loss you could expect in a year.
- SLE is another monetary value, and it represents how much you expect to lose at any one time: the single loss expectancy. SLE can be divided into two components:

    - AV (asset value)
    - EF (exposure factor)
- ARO is the likelihood, often drawn from historical data, of an event occurring within a year" the annualized rate of occurrence

When you compute risk assessment, remember this formula:

SLE × ARO = ALE

As an example, if you can reasonably expect that every SLE, which is equal to asset value (AV) times exposure factor (EF), will be the equivalent of $1,000 and that there will be seven such occurrences a year (ARO), then the ALE is $7,000. Conversely, if there is only a 10 percent chance of an event occurring within a year time period (ARO = 0.1), then the ALE drops to $100.

Quantitative vs. qualitative
Risk assessment can be either qualitative (opinion-based and subjective) or quantitative (cost-based and objective), depending on whether you are focusing on dollar amounts. The formulas for single loss expectancy (SLE), annual loss expectancy (ALE), and annualized rate of occurrence (ARO) are all based on doing assessments that lead to dollar amounts and are thus quantitative.

To understand the difference between quantitative and qualitative, it helps to use a simple example. Imagine that you get an emergency call to help a small company that you have never heard from before. It turns out that their one and only server has crashed and that their backups are useless. One of the lost files was the only copy of the company's his- tory. This file detailed the company from the day it began to the present day and had the various iterations of the mission statement as it changed over time. As painful a loss as this file represents to the company's culture, it has nothing to do with filling orders and keeping customers happy, and thus its loss is qualitative in nature.

Another loss was the customer database. This held customer contact information as well as the history of all past orders, charge numbers, and so on. The company cannot function without this file, and it needs to be re-created by pulling all of the hard copy invoices from storage and re-entering them into the system. This loss can be calculated by the amount of business lost and the amount of time it takes to find/re-enter all the data, and thus it is a quantitative loss.

Vulnerabilities
Many security experts view vulnerability scanning as separate from penetration testing. However, it should be either part of the penetration test or done alongside it. Vulnerability scanning allows you to identify specific vulnerabilities in your network, and most penetration testers will start with this procedure so that they can identify likely targets to attack. A penetration test is essentially an attempt to exploit these vulnerabilities.

Once you have identified the vulnerabilities, it is time to attempt to exploit them. Of course the most egregious vulnerability is any aspect of your system where vulnerability scanning reveals a lack of security controls. Some of the more common vulnerabilities involve misconfiguration. In fact, popular vulnerability scanners, such as Nessus will help identify common misconfigurations.

Threat vectors
The term threat vector is the way in which an attacker poses a threat. This can be a particular tool that they can use against you (a vulnerability scanner, for example) or the path(s) of attack that they follow. Under that broad definition, a threat vector can be anything from a fake email that lures you into clicking a link (phishing) or an unsecured hotspot (rouge access point) and everything in between.

Probability / threat likelihood
The meaning of the word likelihood is usually self-explanatory; however, there are actual values that can be assigned to likelihood. The National Institute of Standards and Technology (NIST) recommends viewing likelihood as a score representing the possibility of threat initiation. In this way, it can be expressed either in qualitative or quantitative terms.

Risk-avoidance, transference, acceptance, mitigation, deterrence
Once you've identified and assessed the risks that exist, for the purpose of the exam you have five possible actions that you can choose to follow:

Risk Avoidance Risk avoidance involves identifying a risk and making the decision not to engage any longer in the actions associated with that risk. For example, a company may decide that many risks are associated with email attachments and choose to forbid any email attachments from entering the network.
Risk Transference Risk transference, contrary to what the name may imply, does not mean that you shift the risk completely to another entity. What you do instead is share some of the burden of the risk with someone else, such as an insurance company. A typical policy would pay you a cash amount if all of the steps were in place to reduce risk and your system was still harmed.
Risk Mitigation Risk mitigation is accomplished any time you take steps to reduce risk. This category includes installing antivirus software, educating users about possible threats, monitoring network traffic, adding a firewall, and so on. In Microsoft's SecurityIntelligence Report, Volume 13, the following suggestions for mitigating risk through user awareness training are listed:

- Keep security messages fresh and in circulation.
- Target new employees and current staff members.
- Set goals to ensure that a high percentage of the staff is trained on security best practices.
- Repeat the information to raise awareness.

CompTIA is fond of risk mitigation and confronting it through the use of routine audits that address user rights and permission reviews, change management—the structured approach that is followed to secure a company's assets—and incident management—the steps followed when events occur (making sure controls are in place to prevent unauthorized access to, and changes of, all IT assets). Policies addressing data loss or theft need to be in place, and technology controls should be enforced.

Risks associated with Cloud Computing and Virtualization
The term cloud computing has grown in popularity recently, but few agree on what it truly means. For the purpose of the Security+ exam, cloud computing means hosting services and

data on the Internet instead of hosting it locally. Some examples of this include running office suite applications such as Office 365 or Google Docs from the Web instead of having similar applications installed on each workstation; storing data on server space, suchas Google Drive, Sky Drive, or Amazon Web Services; and using cloud-based sites such as Salesforce.com.

If cloud computing has grown in popularity, virtualization has become the technology du jour. Virtualization consists of allowing one set of hardware to host multiple virtual machines. It is in use at most large corporations, and it is also becoming more common at smaller businesses.

Some of the possible security risks associated with virtualization include the following:

Breaking Out of the Virtual Machine If a disgruntled employee could break out of the virtualization layer and were able to access the other virtual machines, they could access data that they should never be able to access.
Network and Security Controls Can Intermingle The tools used to administer the virtual machine may not have the same granularity as those used to manage the network. This could lead to privilege escalation and a compromise of security.

Most virtualization-specific threats focus on the hypervisor. Hypervisor is the virtual machine monitor; that is, the software that allows the virtual machines to exist. If the hypervisor can be successfully attacked, the attacker can gain root-level access to all virtual systems. Although this is a legitimate issue, and one that has been demonstrated as possible in most systems (including VMware, Xen, and Microsoft Virtual Machine), it is one that has been patched each time it has arisen. The solution to most virtualization threats is always to apply the most recent patches and keep the system(s) up to date. Be sure to look for and implement suggestions that the vendor of your virtualization system may have published in a hardening guide.

Recovery time objective and recovery point objective
The recovery time objective (RTO) is the maximum amount of time that a process or service is allowed to be down and the consequences still be considered acceptable. Beyond this time, the break in business continuity is considered to affect the business negatively. The RTO is agreed on during BIA creation.

The recovery point objective (RPO) is similar to RTO, but it defines the point at which the system needs to be restored. This could be where the system was two days before it crashed (whip out the old backup tapes) or five minutes before it crashed (requiring complete redundancy). As a general rule, the closer the RPO matches the item of the crash, the more expensive it is to obtain.
Most SLAs that relate to risk management stipulate the definitions of these terms and how they apply to the agreement.

Section 2.2 - Summarize the security implications of integrating systems and data with third parties

On-boarding/off-boarding business partners
Transitioning with a business partner occurs either during the on-boarding or off-boarding of a business partner. Both the initialization and the termination of a close business relationship have serious security issues.

During the on boarding of a new business partner, it is important to determine whether the security policies of both organizations are compatible, at least in areas where the two companies' networks will interact. One area that usually does get adequate attention from most companies is the issue of interoperability agreements. These are documents that define how the two organizations' systems will interoperate and what the minimum requirements and expectations are.
Just as important is the issue of who owns the data and how that data will be backed up. In a joint enterprise, data may be combined from both organizations. It must be determined, in advance, who is responsible for that data and how the data backups will be managed.

Data backup issues include how frequently to back up, how and where to store backup media, and how to test the backup media.

It is also critical to consider privacy considerations. Certain businesses, such as medical-related companies, have specific, legally mandated privacy requirements. However, any business that has any personal data must take into consideration the security of that data. When two different organizations are interoperating, ensuring that both organizations maintain a minimum level of privacy protection is important.

Social media networks and/or applications
Many companies allow full use of social media in the workplace, believing that the marketing opportunities it holds outweigh any loss in productivity. What they are unknowingly minimizing are the threats that exist. Rather than being all new threats, the social networking/media threats tend to fall in the categories of the same old tricks used elsewhere but in a new format. A tweet can be sent with a shortened URL so that it does not exceed the 140-character limit set by Twitter; unfortunately, the user has no idea what the shortened URL leads to—it could be to a phishing site, a downloadable Trojan, or just about anything else.

Interoperability agreements
There are some specific documents that need to be part of any interoperability agreement. Those are described here:

SLA: The Service-Level Agreement The SLA defines the level of service to be provided. For example, with a company providing technical support, the SLA will determine the response time (for example, will a tech be on site within 4 hours? 8 hours?) and the level of response (will there be a replacement part if needed?).

BPO: The Blanket Purchase Order This is usually applicable to government agencies. It is an agreement between a government agency and a private company for ongoing purchases of goods or services.

MOU: The Memorandum of Understanding This document is used in many settings in the information industry. It is a brief summary of which party is responsible for what portion of the work. For example, Company A may be responsible for maintaining the database server and Company B may be responsible for telecommunications.

ISA: The Interconnection Security Agreement This is an agreement between two organizations that have connected systems. The agreement documents the technical requirements of the connected systems.

All documented standards are subject to verification. The documents represent the standards that are agreed upon, but it is necessary to periodically verify compliance and performance standards. This can be done through a review of procedures, an actual audit, or even a vulnerability scan or penetration test. The level of review is up to the two parties involved.

Risk awareness
Communication and awareness help ensure that security information is conveyed to the appropriate people in a timely manner. Most users aren't aware of current security threats. If you set a process in place to explain concisely and clearly what is happening and what is being done to correct current threats, you'll probably find the acceptance of your efforts to be much higher.

Communication methods that have proven to be effective for disseminating information include internal security websites, news servers, and emails. You might want to consider a regular notification process to convey information about security issues and changes.

In general, the more you communicate in a routine manner, the more likely people will internalize the fact that security is everybody's responsibility.

Unauthorized data sharing
One of the major reasons to implement a cryptographic system is to ensure the confidentiality of the information being used. Confidentiality may be intended to prevent the unauthorized dis- closure of information in a local network or to prevent the unauthorized disclosure of information across a network. A cryptographic system must do this effectively in order to be of value.

The need to keep records secure from internal disclosure may be just as great as the need to keep records secure from outside attacks. The effectiveness of a cryptographic system in preventing unauthorized decryption is referred to as its *strength*: A strong cryptographic system is difficult to crack. Strength is also referred to as the algorithm's *work factor*: The *work factor* describes an estimate of the amount of time and effort that would be needed to break a system.

The system may be considered weak if it allows weak keys, has defects in its design, or is easily decrypted. Many systems available today are more than adequate for business and personal use, but they are inadequate for sensitive military or governmental applications.

Cipher suites, for example, work with SSL/TLS to combine authentication, encryption, and message authentication. Most vendors allow you to set cipher suite preferences on a server to determine the level of strength required by client connections. With Sybase, for example, you set the cipher suite preference to Weak, Strong, FIPS, or All. If you choose Strong, you are

limiting the choices to only encryption algorithms that use keys of 64 bits or more. Choosing Weak adds all the encryption algorithms that are less than 64 bits, while choosing FIPS requires encryptions, hash and key exchange algorithms to be FIPS- compliant (AES, 3DES, DES, and SHA1). Apache offers similar choices but instead of the words Strong and Weak, the names are changed to High, Medium, and Low.

Data ownership
Data ownership becomes an issue with BYOD. If the device is personally owned but used for company business, who owns the data on the device? The company or the individual? Related to that is the issue of support ownership. Is the individual responsible for support or the company? Patch management is closely related to support ownership. Who will be responsible for ensuring the personal device has patches updated? Antivirus management is another related issue. What antivirus software will be used? How will it be updated? These are all important questions that will need to be answered.

Adherence to corporate policies is an obvious issue. If individuals own their own devices, which they have purchased with their own funds, ensuring the user and the device adheres to corporate policies will be a challenge. Related to that issue is legal concerns. When a device is owned by the individual but used for company business, a number of legal issues arise. As just one example, what if the device is used to send spam? Is the company responsible? Another example would involve the employee leaving the company. How does the organization verify the device does not have any proprietary data on it? Forensics is another legal issue. If there is, for example, litigation against the company, usually computer records are subpoenaed, but the data that might reside on a personal device is a legal gray area.

Data backups
At the most basic level, fault tolerance for a server means a data backup. *Backups* are simply the periodic archiving of the data so that if there is a server failure you can retrieve the data. Although database administrators may use a number of different types of data back- ups (for example, transaction log backups), from a security point of view there are three primary backup types with which we are concerned:

Full All changes to the data are archived.

Differential All changes since the last full backup are archived.

Incremental All changes since the last backup of any type are archived.

So consider a scenario where you do a full backup at 2 a.m. each morning. You are concerned about the possibility of a server crash before the next full backup, so you want to do a backup every two hours. The type of backup you choose will determine the efficiency of doing those frequent backups and the time needed to restore data. Let's consider each scenario and what would happen if the system crashes at 10:05 a.m.

Full Backup In this scenario, assume a full backup is done every two hours beginning at 2 a.m. When the system crashes at 10:05 a.m., you simply need to restore the 10:00 a.m. full backup. However, running a full backup every two hours is very time consuming and resource intensive, and it will have a significant negative impact on your server's performance.

Differential Backup In this scenario, you do a full backup at 2 a.m. and then perform a

differential every two hours thereafter. When the system crashes at 10:05 a.m., you have to restore the full backup from 2 a.m. and the differential backup from 10 a.m. This takes just one more step than restoring the full backup. Keep in mind, however, that the differential backups are going to get larger each time you do them and thus more time consuming and resource intensive. Although they won't have the impact of the full backups, they will still slow down your network.

Incremental Backup In this scenario you do a full backup at 2 a.m. and then an incremental backup every two hours. When the system crashes at 10:05 a.m., you need to restore the last full backup done at 2 a.m. and then each incremental backup done since then—and they must be restored in order. This is much more complex to restore, but each incremental backup is small and does not take much time, nor do they consume many resources.

Follow security policy and procedures
An organization's security management policies don't exist in a vacuum. Regulatory and governmental agencies are key components of a security management policy. These agencies have made large improvements over the last several years to ensure the privacy of information; several laws have been passed to help ensure that information isn't disclosed to unauthorized parties. As a security professional, you must stay current with these laws because you're one of the primary agents to ensure compliance.

Section 2.3 Given a scenario, implement appropriate risk mitigation strategies.

The selection and specification of security controls for an information system is accomplished as part of an organization-wide information security program that involves the management of organizational risk---that is, the risk to the organization or to individuals associated with the operation of an information system. The management of organizational risk is a key element in the organization's information security program and provides an effective framework for selecting the appropriate security controls for an information system---the security controls necessary to protect individuals and the operations and assets of the organization.

Risk-Based Approach
The risk-based approach to security control selection and specification considers effectiveness, efficiency, and constraints due to applicable laws, directives, Executive Orders, policies, standards, or regulations. The following activities related to managing organizational risk (also known as the Risk Management Framework) are paramount to an effective information security program and can be applied to both new and legacy information systems within the context of the system development life cycle and the Federal Enterprise Architecture:

Step 1: Categorize
Categorize the information system and the information processed, stored, and transmitted by that system based on an impact analysis.

Step 2: Select
Select an initial set of baseline security controls for the information system based on the

security categorization; tailoring and supplementing the security control baseline as needed based on organization assessment of risk and local conditions.

Step 3: Implement
Implement the security controls and document how the controls are deployed within the information system and environment of operation.

Step 4: Assess
Assess the security controls using appropriate procedures to determine the extent to which the controls are implemented correctly, operating as intended, and producing the desired outcome with respect to meeting the security requirements for the system .

Step 5: Authorize
Authorize information system operation based upon a determination of the risk to organizational operations and assets, individuals, other organizations and the Nation resulting from the operation of the information system and the decision that this risk is acceptable.

Step 6: Monitor
Monitor and assess selected security controls in the information system on an ongoing basis including assessing security control effectiveness, documenting changes to the system or environment of operation, conducting security impact analyses of the associated changes, and reporting the security state of the system to appropriate organizational officials.

Change management
The purpose of change management is to ensure proper procedures are followed when modifications to the IT infrastructure are made. These modifications can be prompted by a number of different reasons including new legislation, updated versions of software or hardware, implementation of new software or hardware, or improvements to the infrastructure. The term “management” implies that this process should be controlled in some systematic way, and that is indeed the purpose. Changes to the infrastructure can have a detrimental impact on operations. New versions of operating systems or application software can be incompatible with other software or hardware the organization is using. Without a process to manage the change, an organization can suddenly find itself unable to conduct business.

Incident management
'Real World' definition of Incident Management: IM is the way that the Service Desk puts out the 'daily fires'.

An 'Incident' is any event which is not part of the standard operation of the service and which causes, or may cause, an interruption or a reduction of the quality of the service.

The objective of Incident Management is to restore normal operations as quickly as possible with the least possible impact on either the business or the user, at a cost-effective price.

Inputs for Incident Management mostly come from users, but can have other sources as well like management Information or Detection Systems. The outputs of the process are RFC’s

(Requests for Changes), resolved and closed Incidents, management information and communication to the customer.

Activities of the Incident Management process:

- Incident detection and recording
- Classification and initial support
- Investigation and diagnosis
- Resolution and recovery
- Incident closure
- Incident ownership, monitoring, tracking and communication

Perform routine audits
System administrators should fulfill a number of periodical operations which are required to ensure that the network infrastructure is running and all computer systems work fine. VIP Task Manager can facilitate planning and tracking of IT administrators' routine tasks.

IT routine audits and maintenance tasks planning
Administrators should plan and perform daily technical checkups of hardware. These tasks are routine, so are regulated by certain procedure. VIP Task Manager is groupware product, which allows IT specialists to schedule and mark out their routine operations, and procedures, which are related to IT infrastructure maintenance and audits. For this purpose system administrator can create recurrence tasks, which should be done every day at certain time and set pop-up sound reminders for these tasks. IT specialists are able to plan their tasks by date, time and duration, set priority, order and use Scheduler (Calendar) mode to see the workloads for day, week or month.

IT routine audits and maintenance tasks tracking
The senior administrator should control the execution of routine tasks, be aware if any issues appear and keep the registration of tasks performed. The senior system administrator can plan tasks for subordinates, give instructions and control the work performance for IT system administrative team - get immediate notifications about completion of certain tasks and get acquainted with work results (reports attached to task - files, links, notes).VIP Task Manager is customizable and flexible groupware with access rights management, so each internal team (administrators, developers etc.) within IT department can work confidentially and safely and even do not know about tasks of each other, or contrary - easily share common projects.

IT routine audits and maintenance tasks reporting
With applying operating system and software updates, patches, and configuration changes, performing backups and restores, hardware monitoring and troubleshooting, system administrators should make reports and store them. Most reports are generated by special professional software and can be stored as files of different formats. After some checkup is done, the report can be attached to task, so IT specialist can easily find and open it next time. Besides system administrators can attach hyperlinks to files that are stored within the LAN in shared folder, for example user can set path to some system log that is placed on any computer within LAN and open it any time by one click on hyperlink. VIP Task Manager allows making printable reports and building charts to assess the state of work performance, to see volumes of work done and to learn IT staff performance indicators.

Enforce technology controls
CompTIA is fond of risk mitigation and confronting it through the use of routine audits that address user rights and permission reviews, change management—the structured approach that is followed to secure a company's assets—and incident management—the steps followed when events occur (making sure controls are in place to prevent un-authorized access to, and changes of, all IT assets). Policies addressing data loss or theft need to be in place, and technology controls should be enforced.

Data loss prevention (DLP) systems monitor the contents of systems (workstations, servers, and networks) to make sure that key content is not deleted or removed. They also monitor who is using the data (looking for unauthorized access) and transmitting the data. DLP systems share commonality with network intrusion prevention systems.

Section 2.4-Given a scenario, implement basic forensic procedures

The five steps outlined here will help in all incident response situations. For the exam, however, there are a number of procedures and topics about which CompTIA wants you to be aware that are relevant to a forensic investigation. We strongly recommend that you familiarize yourself with these topics as you prepare for the exam.

Act in Order of Volatility
When dealing with multiple issues, address them in order of volatility (OOV); always deal with the most volatile first. Volatility can be thought of as the amount of time that you have to collect certain data before a window of opportunity is gone. Naturally, in an investigation you want to collect everything, but some data will exist longer than others, and you cannot possibly collect all of it once. As an example, the OOV in an investigation may be RAM, hard drive data, CDs/DVDs, and printouts.

Capture System Image
A system image is a snapshot of what exists. Capturing an image of the operating system in its exploited state can be helpful in revisiting the issue after the fact to learn more about it. As an analogy, think of germ samples that are stored in labs after major outbreaks so that scientists can revisit them later and study them further.

Document Network Traffic and Logs
Look at network traffic and logs to see what information you can find there. This information can be useful in identifying trends associated with repeated attacks.

Capture Video
Capture any relevant video that you can. Video can later be analyzed manually in individual frames as well as run through a number of programs that can create indices of the contents.

Record Time Offset
It is quite common for workstation times to be off slightly from actual time, and that can happen with servers as well. Since a forensic investigation is usually dependent on a step-by-step account of what has happened, being able to follow events in the correct time sequence is critical. Because of this, it is imperative to record the time offset on each affected machine

during the investigation. One method of assisting with this is to add an entry to a log file and note the time that this was done and the time associated with it on the system.

Take Hashes It is important to collect as much data as possible to be able to illustrate the situation, and hashes must not be left out of the equation. NIST (the National Institute of Standards and Technology) maintains a National Software Reference Library (NSRL). One of the purposes of the NSRL is to collect "known, traceable software applications" through their hash values and store them in a Reference Data Set (RDS). The RDS can then be used by law enforcement, government agencies, and businesses to determine which files are important as evidence in criminal investigations.

Capture Screenshots
Just like video, capture all relevant screenshots for later analysis. One image can often parlay the same information that it would take hundreds of log entries to equal.

Talk to Witnesses
It is important to talk to as many witnesses as possible to learn exactly what happened and to do so as soon as possible after the incident. Over time, details and reflections can change, and you want to collect their thoughts before such changes occur. If at all possible, document as much of the interview as you can with video recorders, digital recorders, or whatever recording tools you can find.

Track Man Hours and Expenses
Make no mistake about it; an investigation is expensive. Track total man-hours and expenses associated with the investigation, and be prepared to justify them if necessary to superiors, a court, or insurance agents.

Chain of custody
An important concept to keep in mind when working with incidents is the chain of custody, which covers how evidence is secured, where it is stored, and who has access to it. When you begin to collect evidence, you must keep track of that evidence at all times and show who has it, who has seen it, and where it has been. The evidence must always be within your custody, or you're open to dispute about possible evidence tampering.

Big data analysis
One issue that will be tested with the first three (document review, walkthrough, and simulation) is called Big Data analysis. Big Data refers to data that is too large to be dealt with by traditional database management means. As of this writing, this usually means exabytes of data (a terabyte is a thousand gigabytes, a petabyte is a thousand terabytes, and an exabyte is a thousand petabytes). When systems are this large, obviously the system being down has a wide-ranging impact. However, doing a cutover test is very difficult, and in some cases it is just not practical. That does not mean, however, that you can simply ignore those systems in your disaster-recovery planning.

Section 2.5 Summarize common incident response procedures

Preparation

Preparing for incident response involves multiple factors. The first step is outlining how you intend to respond to specific incidents. Formulating an IRP is part of that preparation. You also will need to identify the personnel and resources needed for your response. For example, if you intend to take a server offline in the event that it is breached, do you have a backup server available? In the event of a suspected computer crime, which of your personnel are qualified to perform the initial forensic processes? If no one is qualified, you need to identify a third party that you can contact.

Incident identification

Incident identification is the first step in determining what has occurred in your organization. An internal or external attack may have been part of a larger attack that has just surfaced, or it may be a random probe or scan of your network.

An event is often an IDS-triggered signal. Operations personnel will deter- mine if an event becomes an incident. An easy way to think of the two is that an event is anything that happens, whereas an incident is any event that endangers a system or network.

Many IDSs trigger false positives when reporting incidents. False positives are events that aren't really incidents. Remember that an IDS is based on established rules of acceptance (deviations from which are known as anomalies) and attack signatures. If the rules aren't set up properly, normal traffic may set off the analyzer and generate an event. Be sure to double-check your results because you don't want to declare a false emergency.

One problem that can occur with manual network monitoring is overload. Over time, a slow attack may develop that increases in intensity. Manual processes typically will adapt, and they may not notice the attack until it's too late to stop it. Personnel tend to adaptto changing environments if the changes occur over a long period of time. An automated monitoring system, SUCH AS IDS, will sound the alarm when a certain threshold or activity level occurs.

When a suspected incident pops up, first responders are those individuals who must ascertain whether it truly is an incident or a false alarm. Depending on your organization, the first responder may be the main security administrator or it could consist of a team of network and system administrators.

The very first step, even with a suspected incident, is isolation. If you think, for example, a given machine is infected with a virus, you must isolate that machine, and even before you are sure it is indeed infected. That involves quarantining the machine(s) that you suspect of being infected. Literally disconnect them from the network while you analyze the situation. In some cases this is accomplished with simple device removal: Just remove the device from the network by unplugging the network cable.

After you've determined that you indeed have an incident on your hands, you need to consider how to handle it. This process, called escalation, involves consulting policies, consulting appropriate management, and determining how best to conduct an investigation into the incident. Make sure that the methods you use to investigate the incident are consistent with corporate and legal requirements for your organization. Bring your Human Resources and Legal departments into the investigation early, and seek their guidance whenever questions involving their areas of expertise arise.

A key aspect, often overlooked by system professionals, involves information control. When an incident occurs, who is responsible for managing the communications about the incident? Employees in the company may naturally be curious about a situation. A single spokesperson needs to be designated. Remember, what one person knows runs a risk of one hundred others also finding out.

Mitigation steps
CompTIA is fond of risk mitigation and confronting it through the use of routine audits that address user rights and permission reviews, change management—the structured approach that is followed to secure a company's assets—and incident management—the steps followed when events occur (making sure controls are in place to prevent un-authorized access to, and changes of, all IT assets). Policies addressing data loss or theft need to be in place, and technology controls should be enforced.

Reporting
During the entire process of responding to an incident, you should document the steps you take to identify, detect, and repair the system or network. This information is valuable; it needs to be captured in case an attack like this occurs again. The documentation should be accessible by the people most likely to deal with this type of problem. Many help-desk software systems provide detailed methods that you can use to record procedures and steps. These types of software products allow for fast access.
If appropriate, you should report/disclose the incident to legal authorities and CERT (www.cert.org) so that others can be aware of the type of attack and help to look for pro-active measures to prevent it from happening again.

You might also want to inform the software or system manufacturer of the problem and how you corrected it. Doing so might help them inform or notify other customers of the threat and save time for someone else.

First responder
When a suspected incident pops up, first responders are those individuals who must ascertain whether it truly is an incident or a false alarm. Depending on your organization, the first responder may be the main security administrator or it could consist of a team of network and system administrators.

Incident isolation
Forensics refers to the process of identifying what has occurred on a system by examining the data trail. It involves an analysis of evidence found in computers and on digital storage media. Incident response encompasses forensics and refers to the process of identifying, investigating, repairing, documenting, and adjusting procedures to prevent another incident. An incident is the occurrence of any event that endangers a system or network. We need to discuss responses to two types of incidents: internal incidents and incidents involving law enforcement professionals

Damage and loss control
One of your first considerations after an incident is to determine how to restore access to resources that have been compromised. Then, of course, you must reestablish control of the system. Most operating systems provide the ability to create a disaster-recovery process using distribution media or system state files.

After a problem has been identified, what steps will you take to restore service? In the case of a DoS attack, a system reboot may be all that is required. Your operating system manufacturer will typically provide detailed instructions or documentation on how to restore services in the event of an attack.

If a system has been severely compromised, as in the case of a worm, it might not be possible to repair it. It may need to be regenerated from scratch. Fortunately, antivirus soft- ware packages can repair most of the damage done by the viruses you encounter. But what if you come across something new? You might need to start over with a new system. In that case, we strongly advise you to do a complete disk drive format or repartition to ensure that nothing is lurking on the disk, waiting to infect your network again.

Section 2.6 Explain the importance of security related awareness and training

Security Policy Training and Procedures
Security awareness and training are critical to the success of a security effort. They include explaining policies, procedures, and current threats to both users and management.

A security awareness and training program can do much to assist in your efforts to improve and maintain security. Such efforts need to be ongoing, and they should be part of the organization's normal communications to be effective.

Communicating with Users to Raise Awareness
Communication and awareness help ensure that security information is conveyed to the appropriate people in a timely manner. Most users aren't aware of current security threats. If you set a process in place to explain concisely and clearly what is happening and what is being done to correct current threats, you'll probably find the acceptance of your efforts to be much higher.

Communication methods that have proven to be effective for disseminating information include internal security websites, news servers, and emails. You might want to consider a regular notification process to convey information about security issues and changes.

In general, the more you communicate in a routine manner, the more likely people will internalize the fact that security is everybody's responsibility.

Providing Education and Training
Your efforts in education and training must help users clearly understand prevention, enforcement, and threats. In addition to the efforts of the IT staff, the security department will probably be responsible for a security awareness program. Your organization's training and educational programs need to be tailored for at least three different audiences:

- The organization as a whole (the so-called rank and file employees)
- Management
- Technical staff

These three organizational roles have different considerations and concerns. For example, with organization-wide training everyone understands the policies, procedures, and resources available to deal with security problems, so it helps to ensure that all employees are on the same page.

Role-Based Training
Ideally, a security awareness-training program for the entire organization should cover the following areas:

- Importance of security
- Responsibilities of people in the organization
- Policies and procedures
- Usage policies
- Account and password-selection criteria
- Social engineering prevention

You can accomplish this training either by using internal staff or by hiring outside trainers. We recommend doing much of this training during new-employee orientation and staff meetings. To stay at their forefront of employees' minds, though, the training needs to be repeated periodically (once a year often works well). Also, don't forget to have employees sign that they received the training and are aware of the policies.

Management Managers are concerned with more global issues in the organization, including enforcing security policies and procedures. Managers will want to know the how's and whys of a security program: how it works and why it is necessary. They should receive additional training or exposure that explains the issues, threats, and methods of dealing with threats. Management will also be concerned about productivity impacts and enforcement and how the various departments are affected by security policies.

Technical Staff The technical staff needs special knowledge about the methods, implementations, and capabilities of the systems used to manage security. Network administrators should evaluate how to manage the network, best practices, and configuration issues associated with the technologies they support. Developers and implementers must evaluate the impact that these measures have on existing systems and new development projects. The training that both administrators and developers need will be vendor specific; vendors have their own methods of implementing security.

Personally Identifiable Information
Personally identifiable information (PII) is a catchall for any data that can be used to uniquely identify an individual. This data can be anything from the person's name to a fingerprint (think biometrics), credit card number, or patient record. The term became mainstream when the NIST (National Institute of Standards and Technology) began issuing guides and recommendations regarding it.

Information Classification
Information classification is a key aspect of a secure network. Again, the process of developing a classification scheme is both a technical and a human issue. The technologies you use must be able to support your organization's privacy requirements. People and processes must be in place and working effectively to prevent unauthorized disclosure of

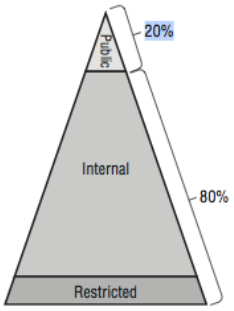
sensitive information.

Information can be generally classified by confidentiality as simply high, medium, or low. However, this is rather vague and not quite as helpful.

If you think about all the information your organization keeps, you'll probably find that it breaks down into three primary categories:

- Public use
- Internal use
- Restricted use

The following figure shows the typical ratios of how this information is broken down. Notice that 80 percent of the information in your organization is primarily for internal or private use. This information would include memos, working papers, financial data, and information records, among other things.



Data Labeling and Handling
A great many users don't consider that there are different types of data and various values associated with it. They don't realize that a misplaced backup copy of the mission statement is not as great a loss from a financial standpoint as a misplaced backup copy of customer contacts.

As a security administrator, you should help users to realize that different types of data unique to your organization have different values and need to be labeled accordingly. Once it has been established and understood that there are significant differences, you can address handling these different types of data. The importance of protecting the data in all forms—online,backups, hard copies, and so on—should be covered as well as reasons why different groups should not access data outside of their permission category.

Compliance with Laws, Best Practices, and Standards
Users need to realize that working with data is the same as driving a car, owning a home, or almost anything else in that there are laws, practices, and standards to which they must adhere. Just as negligence fails tobe an admissible excuse in other areas of the law, the same holds true when working with data. New regulations are passed regularly, and it is your job as an administrator to educate users on those that are applicable in your environment.

User Habits

Password behaviors

Users need to understand that the stronger they make their pass- word, the more difficult they make anyone's attempt to crack it. They should be educated to use long passwords consisting of letters, numbers, and characters and to change them frequently.

They must also be educated that they cannot write their password down on a sticky note right after a change and post it under the keyboard, on the monitor, or anywhere else. The reasons for regularly changing passwords should be explained along with the requirement that you will make them do so at least every three months.

Data Handling
Only those users needing to work with it should access data. It is your job to implement safeguards to keep the data from being seen by those who should not, but the users need to understand why those safeguards are there and abide by them. There are plenty of examples of companies that have suffered great financial loss when their information, trade secrets, and client information was leaked.

Policy on Personally Owned Devices
Empathize with the users who want to bring their gad- gets from home, but make them understand why they cannot. You do not want them plugging in a flash drive, let alone a camera, smartphone, tablet computer, or other device, on which company files could get intermingled with personal files. Allowing this to happen can create situations where data can leave the building that shouldn't as well as introduce malware to the system. There has been a rash of incidents in which data has been smuggled out of an organization through personal devices.

Employees should not sync unauthorized smartphones to their work systems. Some smart- phones use multiple wireless spectrums and unwittingly open up the possibility foran attacker in the parking lot to gain access through the phone to the internal network.

Ban—and make sure the users know that you have done so—all social peer-to-peer (P2P) networking. These are common for sharing files such as movies and music, but you must not allow users to bring in devices and create their own little networks to share files, print- ers, songs, and so on. All networking must be done through administrators and not on a P2P basis. The P2P ports should be listed on the company servers (either whitelisted or blacklisted), and an alert should be sent to you any time someone attempts any P2P activity. Vigilantly look for all such activity, and put a stop to it immediately.

Prevent Tailgating
Tailgating is the term used for someone being so close to you when you enter a building that they are able to come in right behind you without needing to use a key, a card, or any other security device. Many social engineering intruders needing physical access to a site will use this method of gaining entry. Educate users to beware of this and other social engineering ploys and prevent them from happening.

Clean Desk Policy
Information on a desk—in terms of printouts, pads of note paper, sticky notes, and the like—can be easily seen by prying eyes and taken by thieving hands. To protect data and your business, encourage employees to maintain clean desks and to leave out only those papers that are relevant to the project they are working on at that moment. All sensitive information should be put away when the employee is away from their desk.

New Threats and New security trends/alerts
Many users work on data away from the office as well as in the office. They need to understand that the data is only as strong as the weakest place in which it is used, and they must have security measures on their home computers that protect your company's data as well. Although the home systems will never be as secure (most likely) as the business systems, at a minimum the home systems need to be running firewalls and updated virus scanners.

Section 2.7 Compare and contrast physical security and environmental controls

Environmental controls
The location of your computer facility is critical to its security. Computer facilities must be placed in a location that is physically possible to secure. Additionally, the location must have the proper capabilities to manage temperature, humidity, and other environmental factors necessary to the health of your computer systems.

HVAC
If the computer systems for which you're responsible require special environmental considerations, you'll need to establish cooling and humidity control. Ideally, systems are located in the middle of the building, and they're ducted separately from the rest of the HVAC (Heating, Ventilation, and Air Conditioning) system. It's a common practice for modern buildings to use a zone-based air conditioning environment, which allows the environmental plant to be turned off when the building isn't occupied. A computer room will typically require full-time environmental control.

Fire suppression
Fire suppression is a key consideration in computer-center design. Fire suppression is the act of extinguishing a fire versus preventing one. Two primary types of fire-suppression systems are in use: fire extinguishers and fixed systems.

EMI shielding
Shielding refers to the process of preventing electronic emissions from your computer systems from being used to gather intelligence and preventing outside electronic emissions from disrupting your information-processing abilities. In a fixed facility, such as a computer center, surrounding the computer room with a Faraday cage can provide electronic shielding. A Faraday cage usually consists of an electrically conductive wire mesh or other conductor woven into a "cage" that surrounds a room.

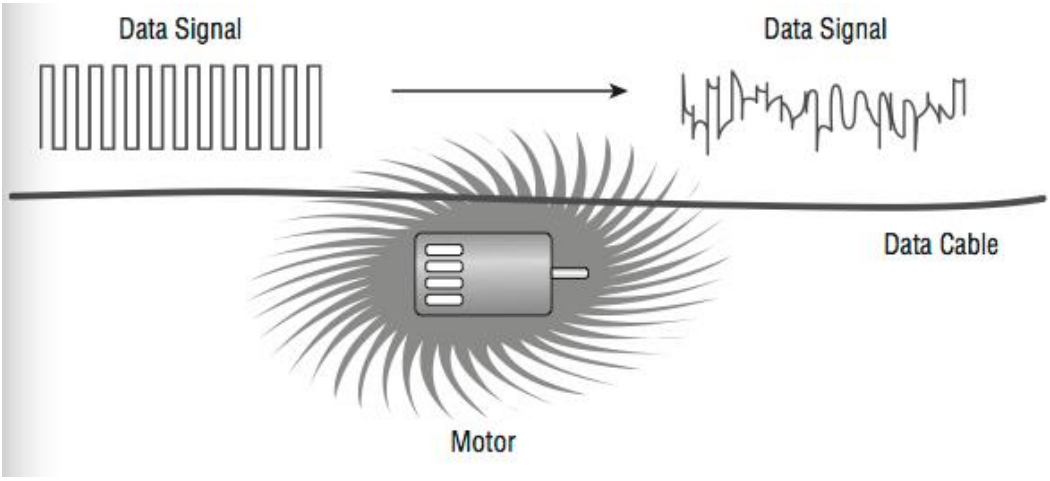
The conductor is then grounded. Because of this cage, few electromagnetic signals can either enter or leave the room, thereby reducing the ability to eavesdrop on a computer conversation. To verify the functionality of the cage, radio frequency (RF) emissions from the room are tested with special measuring devices.

Electromagnetic interference (EMI) and radio frequency interference (RFI) are two additional environmental considerations. Motors, lights, and other types of electromechanical objects cause EMI, which can cause circuit overload, spikes, or electrical component failure. Making sure that all signal lines are properly shielded and grounded can minimize EMI. Devices that

generate EMI should be as physically distant from cabling as is feasible because this type of energy tends to dissipate quickly with distance.

Figure 2.1 shows a motor generating EMI. In this example, the data cable next to the motor is picking up the EMI. This causes the signal to deteriorate, and it might eventually cause the line to be unusable. The gray area in the illustration is representative of the interference generated by the motor.

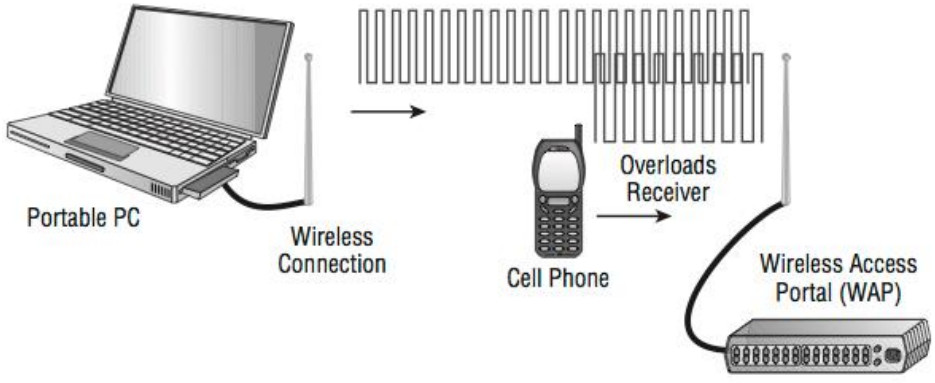
FIGURE 2.2 Electromagnetic interference (EMI) pickup in a data cable



RFI is the byproduct of electrical processes, similar to EMI. The major difference is that RFI is usually projected across a radio spectrum. Motors with defective brushes can generate RFI, as can a number of other devices. If RF levels become too high, it can cause the receivers in wireless units to become deaf. This process is called desensitizing, and it occurs because of the volume of RF energy present. This can occur even if the signals are on different frequencies.

Figure 2.2 demonstrates the desensitizing process occurring with a wireless access portal (WAP). The only solution to this problem is to move the devices farther apart or to turn off the RFI generator.

FIGURE 2.2 RF desensitization occurring as a result of cell phone interference



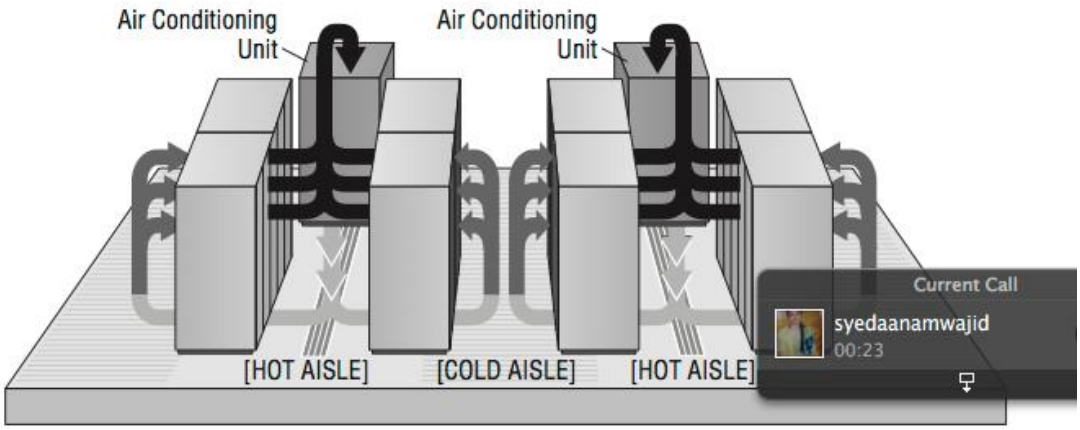
Hot and cold aisles

There are often multiple rows of servers located in racks in server rooms. The rows of servers are known as aisles, and they can be cooled as hot aisles and cold aisles. With a hot aisle, hot

air outlets are used to cool the equipment, whereas with cold aisles, cold air intake is used to cool the equipment. Combining the two, you have cold air intake from below the aisle and hot air outtake above it, providing constant circulation.

It is important that the hot air exhausting from one aisle of racks not be the intake air pulled in by the next row of racks or overheating will occur. Air handlers must move the hot air out, whereas cold air, usually coming from beneath a raised floor, is supplied as the intake air. Figure 2.3 shows an example of a hot and cold aisle design.

FIGURE 2.3 A hot and cold aisle design



Environmental monitoring

Environmental concerns include considerations about water and flood damage as well as fire suppression. Computer rooms should have fire and moisture detectors. Most office buildings have water pipes and other moisture-carrying systems in the ceiling. If a water pipe bursts (which is common in minor earthquakes), the computer room could become flooded. Water and electricity don't mix. Moisture monitors would automatically kill power in a computer room if moisture were detected, so the security professional should know where the water cutoffs are located.

Fire, no matter how small, can cause damage to computer systems. Apart from the high heat, which can melt or warp plastics and metals, the smoke from the fire can permeate the computers. Smoke particles are large enough to lodge under the read/write head of a hard disk, thereby causing data loss. In addition, the fire-suppression systems in most buildings consist of water under pressure, and the water damage from putting out even a small fire could wipe out an entire datacenter.

Temperature and humidity controls

Many computer systems require temperature and humidity control for reliable service. Large servers, communications equipment, and drive arrays generate considerable amounts of heat; this is especially true of mainframe and older minicomputers. An environmental system for this type of equipment is a significant expense beyond the actual computer system costs. Fortunately, newer systems operate in a wider temperature range. Most new systems are designed to operate in an office environment.

Environmental systems should be monitored to prevent the computer center's humidity level from dropping below 50 percent. Electrostatic damage is likely to occur when humidity levels get too low.

Humidity control prevents the buildup of static electricity in the environment. If the humidity drops much below 50 percent, electronic components are extremely vulnerable to damage from electrostatic shock. Most environmental systems also regulate humidity; however, a malfunctioning system can cause the humidity to be almost entirely extracted from a room. Make sure that environmental systems are regularly serviced.

Physical security

Access control is a critical part of physical security, and it can help cut down the possibility of a social engineering or other type of attack from succeeding. Systems must operate in controlled environments in order to be secure. These environments must be, as much as possible, safe from intrusion. Computer system consoles can be a vital point of vulnerability because many administrative functions can be accomplished from the system console. These consoles, as well as the systems themselves, must be protected from physical access.

A key aspect of access control involves physical barriers. The objective of a physical barrier is to prevent access to computers and network systems. The most effective physicalbarrier implementations require that more than one physical barrier be crossed to gain access. This type of approach is called a multiple barrier system or defense in depth.

Ideally, your systems should have a minimum of three physical barriers:

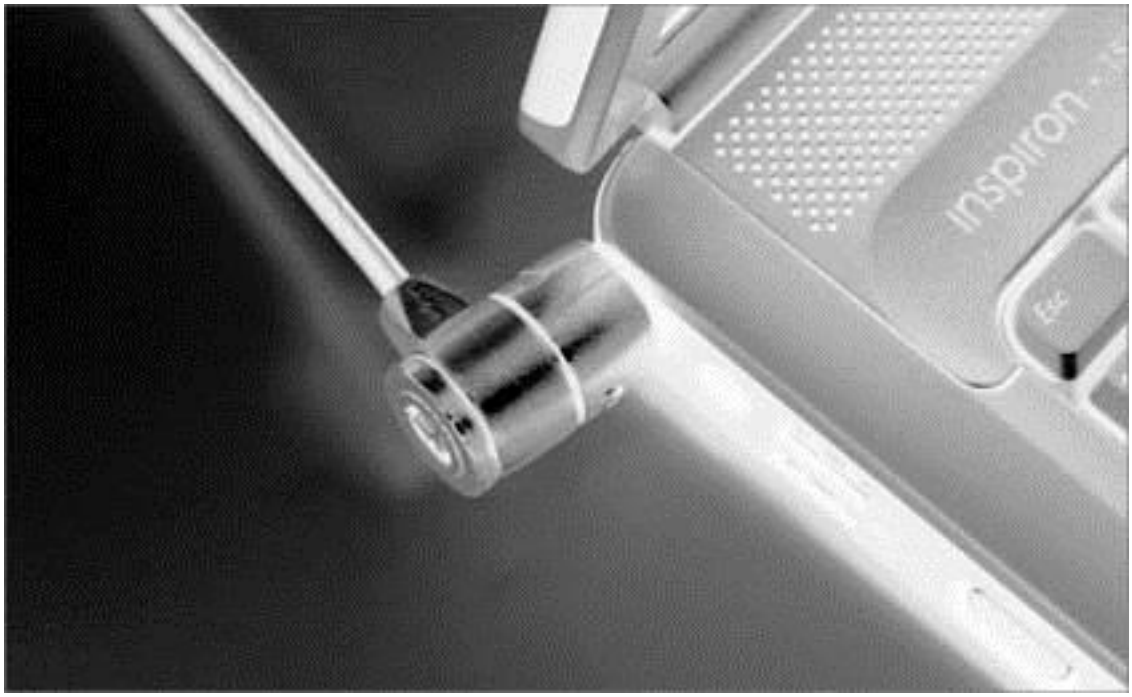
- The external entrance to the building, referred to as a perimeter, which is protected by burglar alarms, external walls, fencing, surveillance, and so on. This should be used with an access list, which identifies who can enter a facility and who can be verified by a guard or someone in authority.
- A locked door protecting the computer center; you should also rely on such items as ID badges, proximity readers, fobs, or keys to gain access.
- The entrance to the computer room itself. This should be another locked door that is carefully monitored. Although you try to keep as many intruders out with the other two barriers, many who enter the building could be posing as someone they are not—heating technicians, representatives of the landlord, and so on. Although these pre-tenses can get them past the first two barriers, the locked computer room door should still stop them.

Hardware locks

Hardware security involves applying physical security modifications to secure the system(s) and preventing them from leaving the facility. Don't spend all of your time worrying about intruders coming through the network wire while overlooking the obvious need for physical security.

Adding a cable lock between a laptop and a desk prevents someone from picking it up and walking away with a copy of your customer database. All laptop cases include a built- in security slot in which a cable lock can be inserted to prevent it from easily being removed from the premises (see Figure 2.4).

FIGURE 2.4 A cable in the security slot keeps the laptop from easily being removed.



When it comes to desktop models, adding a lock to the back cover can prevent an intruder with physical access from grabbing the hard drive or damaging the internal components. The lock that connects through that slot can also go to a cable that then connects to a desk or other solid fixture to keep the entire PC from being carried away. An example of this type of configuration is shown in Figure 2.5.

In addition to running a cable to the desk, you can choose to run an end of it up to the monitor if theft of peripherals is a problem in your company. An example of this type of physical security is shown in Figure 2.6.

You should also consider using a safe and locking cabinets to protect backup media, documentation, and any other physical artifacts that could do harm if they fell into the wrong hands. Server racks should lock the rack-mounted servers into the cabinets to prevent someone from simply pulling one and walking out the front door with it.

FIGURE 2.5 A cable can be used to keep a desktop machine from easily being taken.

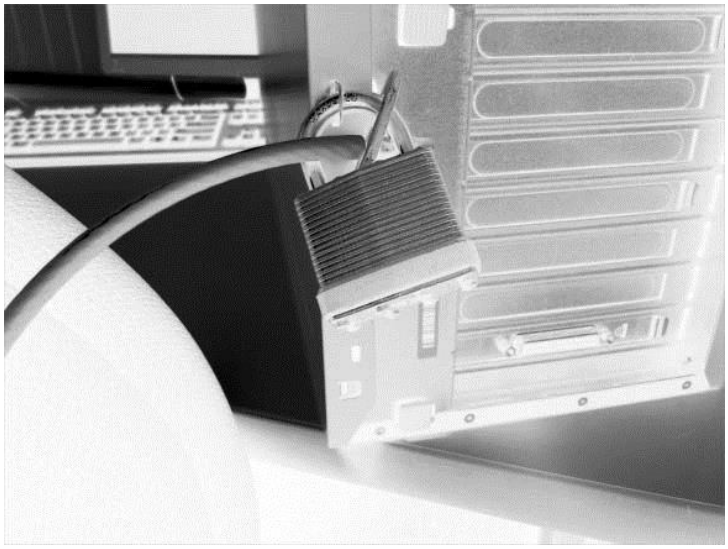
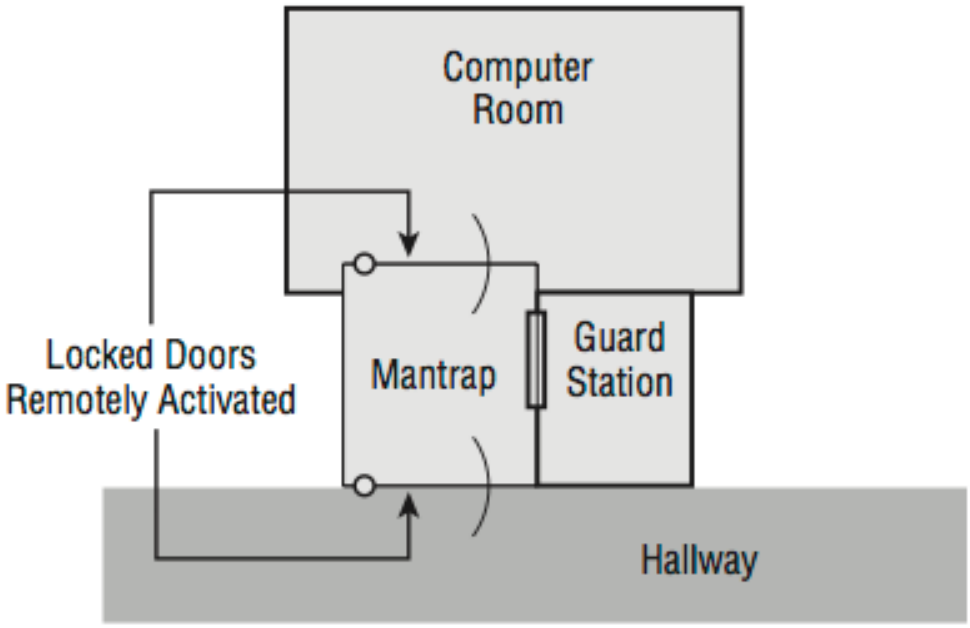


FIGURE 2.6 If theft of equipment is a possibility, run one end of the cable from the monitor to the desktop machine through the hole in the work desk.

Mantraps

High-security installations use a type of intermediate access control mechanism called a mantrap (also occasionally written as man-trap). Mantraps require visual identification, as well as authentication, to gain access. A mantrap makes it difficult for a facility to be accessed by large numbers of individuals at once because it allows only one or two people into a facility at a time. It's usually designed to contain an unauthorized, potentially hostile person physically until authorities arrive. Figure 2.7 illustrates a mantrap. Notice in this case that the visual verification is accomplished using a security guard. A properly developed mantrap includes bulletproof glass, high-strength doors, and locks. After a personis inside the facility, additional security and authentication may be required for further entrance.

FIGURE 2.7 A mantrap in action



Video Surveillance

In high-security and military environments, an armed guard as well as video surveillance would be placed at the mantrap. Beyond mantraps, you can combine guards with cameras (or even the threat of cameras) to create a potent deterrent. The cameras can send signals to a room where they are monitored by a guard capable of responding to a situation when the need arises.

Fencing

Perimeter security, whether physical or technological, is the first line of defense in your security model. In the case of a physical security issue, the intent is to prevent unauthorized access to resources inside a building or facility.

Physical perimeter security is intended to accomplish for a network what perimeter security does for a building. How do you keep intruders from gaining access to systems andinformation in the network through the network? In the physical environment, perimeter security is accomplished through fencing, locks, doors, surveillance systems, and alarm systems. This isn't functionally any different from a network, which uses border routers, intrusion detection systems, and firewalls to prevent unauthorized access.

Few security systems can be implemented that don't have weaknesses or vulnerabilities. A determined intruder can, with patience, overcome most security systems. The task may not be easy, and it may require careful planning and study; however, a determined adversary can usually figure out a way. This is why deterrence is so important.

If you want to deter intruders from breaking into your building, you can install improved door locks, coded alarm systems, and magnetic contacts on doors and windows. Remember

that you can't always keep an intruder out of your building; however, you can make an intrusion riskier and more likely to be discovered if it happens.

Access list

As the name implies, the purpose of an access list is to identify specifically who can enter a facility. Once created, a guard or someone in authority can verify the list. Similar toan access list for physical access, *access control lists* (*ACLs*) enable devices in your network to ignore requests from specified users or systems or to grant them certain network privileges. You may find that a certain IP address is constantly scanning your network, and you can block this IP address. If you block it at the router, the IP address will automatically be rejected any time it attempts to use your network.

Proper Lighting

Lighting can play an important role in the security of any facility. Poor lighting can lead to a variety of unwanted situations: someone sneaking in a door that is not well lit, one individual passing a checkpoint and being mistaken for another person, a biometric reading failure. The latter is particularly true with facial recognition, and proper lighting needs to be in place for both the face and the background.

Signs

One of the least expensive physical security tools that can be implemented is a sign. Signs can be placed around secure areas telling those who venture by that only authorized access is allowed, that trespassers will be prosecuted, and so on. There is a story told of a couple of magicians who drove across country while on tour, and to prevent anyone from breaking into their car, they put a sign on it identifying the car as a transport vehicle for the Centers for Disease Control. Supposedly, it worked and no one ever broke into the vehicle.

Within Microsoft Windows, you have the ability to put signs (in the form of onscreen pop-up banners) that appear before the login telling similar information—authorized access only, violators will be prosecuted, and so forth. Such banners convey warnings or regulatory information to the user that they must "accept" in order to use the machine or network.

In Windows, the banner is turned on in the Registry through an entry beneath HKEY_ LOCAL_MACHINE\SOFTWARE\Microsoft\Windows\CurrentVersion\Policies\System. You can configure legalnoticecaption as the caption of the "sign" that you want to appear and legalnoticetext as the text that will show up and need to be dismissed before the usercan move on. Both are string values accepting any alphanumeric combination.

Guards

As opposed to signs, one of the most expensive physical security tools that can be implemented is a guard. A guard can respond to a situation and be intimidating, but a guard is also fallible and comes at a considerable cost.

Barricades

To stop someone from entering a facility, barricades or gauntlets can be used. These are often used in conjunction with guards, fencing, and other physical security measures, but they can be used as standalones as well.

Biometrics
Biometric systems use some kind of unique biological trait to identify a person, such as fingerprints, patterns on the retina, and handprints. Some methods that are used include hand scanners, retinal scanners, facial recognition applications, and keystroke recognition programs, which can be used as part of the access control mechanisms. These devices should be coupled into security-oriented computer systems that record all access attempts. They should also be under surveillance in order to prevent individuals from bypassing them.

These technologies are becoming more reliable, and they will become widely usedover the next few years. Many laptops sold now have a fingerprint reader built in. The costs associated with these technologies have fallen dramatically in recent years. One of the best independent sources of information on development in the field of biometrics is BiometricNews.net, where you can find links to publications and their blog.

Protected distribution (cabling)
A protected distribution system (PDS) is one in which the network is secure enough to allow for the transmission of classified information in unencrypted format—in other words, where physical network security has been substituted for encryption security. In a small office, for example, you could ban the use of wireless devices and require that all such devices be connected to a bus topology network that is clearly visible as it runs through the space.

Moving forward from this overly simplistic scenario, it is possible to create a much larger network that uses fiber, various topologies, and so on, as long as you still have the ability to monitor and control the span of it. Such networks were once called "approved circuits," and the U.S. government largely uses them.

Alarms
An alarm is used to draw attention to a breach, or suspected breach, when it occurs. This alarm can be sounded in many ways—through the use of a siren, a series of lights (flashing or solid), or an email or voice message—but is always intended to draw attention to the event.

A security zone is an area in a building where access is individually monitored and controlled. A large network, such as the ones foundin a big physical plant, may have many areas that require restricted access. These smaller zones are referred to as security zones. In the physical environment, each floor is broken down into separate zones. An alarm system that identifies a zone of intrusion can inform security personnel about an intruder's location in the building; zone notification tells security where to begin looking when they enter the premises.

The concept of security zones is as old as security itself. Most burglar alarms allow the creation of individual zones within a building or residence; the security staff then treats these zones separately. In a residence, it would be normal for the bedroom to be assigned a zone of its own so that movement here can occur while other parts of the house may be set on a motion detector.

Motion Detection
A motion detection system can monitor a location and signal an alarm if it picks up movement. Systems are commonly used to monitor homes, and the same technology can be used to protect server rooms, office buildings, or any other location. The motion detection

can be accomplished with sensors that are infrared, microwave, or sonic, or that utilize a variety of hybrid sensors.

Control Types
One of the most generic terms in security is control. The word is used so many different ways that its meaning can become blurred. The best thing to do is to equate the word with whatever entity is charged with the task at the moment. That task can be preventing something from happening, logging when something does, responding to it, or any variety of other possibilities. For the exam, CompTIA has categorized controls into six types as follows:

Deterrent A deterrent control is anything intended to warn a would-be attacker that they should not attack. This could be a posted warning notice that they will be prosecuted to the fullest extent of the law, locks on doors, barricades, lighting, or anything can delay or discourage an attack.

Preventive As the name implies, the purpose of preventive controls is to stop something from happening. These can include locked doors that keep intruders out, user training on potential harm (to keep them vigilant and alert), or even biometric devices and guards that deny access until authentication has occurred.

Detective The purpose of a detective control is to uncover a violation. The only time that they would be relevant is when a preventive control has failed and they need to sound an alarm. A detective control can range from a checksum on a downloaded file, an alarm that sounds when a door has been pried open, or an antivirus scanner that actively looks for problems. It could also be a sonic detector, motion sensor, or anything that would detect that an intrusion is under way.

Compensating Compensating controls are backup controls that come into play only when other controls have failed. An office building may have a complex electronic lock on the door (preventive control) and a sign that you will be arrested if you enter (deterrent control), but it is a safe bet they will also have an alarm that sounds (a compensating control) when the door is jimmied as well as a backup generator (another compensating control) to keep that electronic lock active when the power goes out.

Technical Technical controls are those controls implemented through technology. They may be deterrent, preventive, detective, or compensating (but not administrative), and include such things as firewalls, IDS, IPS, and such.

Administrative An administrative control is one that comes down through policies, procedures, and guidelines. An example of an administrative control is the escalation procedure to be used in the event of a break-in: who is notified first, who is called second, and so on. Another example of an administrative control is the list of steps to be followed when a key employee is terminated: disable their account, change the server password, andso forth.

Section 2.8 Summarize risk management best practices.

Business continuity concepts

One of the oldest phrases still in use today is "the show must go on." Nowhere is that more true than in the world of business, where downtime means the loss of significant revenue with each passing minute. Business continuity is primarily concerned with the processes, policies, and methods that an organization follows to minimize the impact of a system failure, network failure, or the failure of any key component needed for operation—that is, essentially whatever it takes to ensure that the business continues and that the show does indeed go on.

Business continuity planning (BCP) is the process of implementing policies, controls, and procedures to counteract the effects of losses, outages, or failures of critical business processes. BCP is primarily a management tool that ensures that critical business functions can be performed when normal business operations are disrupted.
Critical business functions (CBFs) refer to those processes or systems that must be made operational immediately when an outage occurs. The business can't function without them, and many are information-intensive and require access to both technology and data.

Business impact analysis

Two of the key components of BCP are business impact analysis (BIA) and risk assessment. BIA is concerned with evaluating the processes, and risk assessment is concerned with evaluating the risk or likelihood of a loss. Evaluating all of the processes in an organization or enterprise is necessary in order for BCP to be effective.

Identification of critical systems and components

To identify critical functions, a company must ask itself, "What functions are necessary to continue operations until full service can be restored?" This identification process will help you establish which systems must be returned to operation in order for the business to continue. In performing this identification, you may find that a small or overlooked application in a department may be critical for operations. Many organizations have overlooked seemingly insignificant process steps or systems that have prevented business continuity planning (BCP) from being effective. Every department should be evaluated to ensure that no critical processes are overlooked.

Risk assessment

Risk assessment is also known as risk analysis or risk calculation. Risk assessment deals with the threats, vulnerabilities, and impacts of a loss of information-processing capabilities or a loss of information itself. A vulnerability is a weakness that could be exploited by a threat. Each risk that can be identified should be outlined, described, and evaluated for the likelihood of it occurring. The key here is to think outside the box. Conventional threats and risks are often too limited when considering risk assessment.

The key components of a risk-assessment process are outlined here:

Risks to Which the Organization Is Exposed This component allows you to develop scenarios that can help you evaluate how to deal with these risks if they occur. An operating system, server, or application may have known risks in certain environments. You should create a plan for how your organization will best deal with these risks and the best way to respond.

Risks That Need Addressing The risk-assessment component also allows an organization to provide a reality check on which risks are real and which are unlikely. This process helps an organization focus on its resources as well as on the risks that are most likely to occur. For example, industrial espionage and theft are likely, but the risk of a hurricane damaging the server room in Indiana is very low. Therefore, more resources should be allocated to prevent espionage or theft as opposed to the latter possibility.

Coordination with BIA The risk-assessment component, in conjunction with the business impact analysis (BIA), provides an organization with an accurate picture of the situation facing it. It allows an organization to make intelligent decisions about how to respond to various scenarios.

Disaster recovery

Disaster recovery is the ability to recover system operations after a disaster. A key aspect of disaster recovery planning is designing a comprehensive backup plan that includes backup storage, procedures, and maintenance. Many options are available to implement disaster recovery.

Succession planning

Succession planning outlines those internal to the organization who have the ability to step into positions when they open. By identifying key roles that cannot be left unfilled and associating internal employees who can step into these roles, you can groom those employees to make sure that they are up to speed when it comes time for them to fill those positions.

Tabletop Exercises

A tabletop exercise is a simulation of a disaster. It is a way to check to see if your plans are ready to go. There are five levels of testing:

Document Review A review of recovery, operations, resumption plans, and procedures.

Simulation A walkthrough of recovery, operations, resumption plans, and procedures in a scripted "case study" or "scenario."

Parallel Test With this test, you start up all backup systems but leave the main systems functioning.

Cutover Test This test shuts down the main systems and has everything fail over to backup systems.

High Availability

High availability (HA) refers to the measures used to keep services and systems operational during an outage. In short, the goal is to provide all services to all users, where they need them and when they need them. With high availability, the goal is to have key services available 99.999 percent of the time (also known as five nines availability).

Redundancy

Redundancy refers to systems that either are duplicated or fail over to other systems in the event of a malfunction. Failover refers to the process of reconstructing a system or switching

over to other systems when a failure is detected. In the case of a server, the server switches to a redundant server when a fault is detected. This strategy allows service to continue uninterrupted until the primary server can be restored. In the case of a network, this means processing switches to another network path in the event of a network failure in the primary path.

Fault tolerance

Hardware

In addition to software-based encryption, hardware-based encryption can be applied. Within the advanced configuration settings on some BIOS configuration menus, for example, you can choose to enable or disable TPM. A Trusted Platform Module (TPM) can be used to assist with hash key generation. TPM is the name assigned to a chip that can store cryptographic keys, passwords, or certificates. TPM can be used to protect smartphones and devices other than PCs as well.

RAID

The other fundamental aspect of fault tolerance is RAID, or redundant array of independent disks (RAID). RAID allows your servers to have more than one hard drive so that if the main hard drive fails, the system keeps functioning.

Clustering and Load balancing

RAID does a fantastic job of protecting data on systems (which you then protect further with regular backups), but sometimes you need to grow beyond single systems. Anytime you connect multiple computers to work/act together as a single server, it is known as clustering. Clustered systems utilize parallel processing (improving performance and availability) and add redundancy (but also add costs).

High availability can also be obtained through load balancing. This allows you to split the workload across multiple computers. Those computers are often Servers answering HTTP requests (often called a server farm), which may or may not be in the same geo- graphic location. If you split locations, this is usually called a mirror site, and the mirrored copy can add geographic redundancy (allowing requests to be answered quicker) and help prevent downtime.

Disaster recovery concepts

Disaster recovery is the ability to recover system operations after a disaster. A key aspect of disaster recovery planning is designing a comprehensive backup plan that includes backup storage, procedures, and maintenance. Many options are available to implement disaster recovery.

Backup plans/policies

Backups are duplicate copies of key information, ideally stored in a location other than the one where the information is stored currently. Backups include both paper and computer records. Computer records are usually backed up using a backup program, backup systems, and backup procedures.

Cold site

A cold site is a facility that isn't immediately ready to use. The organization using it must bring along its equipment and network. A cold site may provide network capability, but this isn't usually the case; the site provides a place for operations to resume, but it doesn't provide the infrastructure to support those operations. Cold sites work well when an extended outage is anticipated. The major challenge is that the customer must provide all of the capabilities and do all of the work to get back into operation. Cold sites are usually the least expensive to put into place, but they require the most advanced planning, testing, and resources to become operational—occasionally taking up to a month to make operational.

Hot Site

A hot site is a location that can provide operations within hours of a failure. This type of site would have servers, networks, and telecommunications equipment in place to reestablish service in a short time. Hot sites provide network connectivity, systems, and preconfigured software to meet the needs of an organization. Databases can be kept up-to-date using network connections. These types of facilities are expensive, and they're primarily suitable for short-term situations. A hot site may also double as an offsite storage facility, providing immediate access to archives and backup media.

Warm site

A warm site provides some of the capabilities of a hot site, but it requires the customer to do more work to become operational. Warm sites provide computer systems and compatible media capabilities. If a warm site is used, administrators and other staff will need to install and configure systems to resume operations. For most organizations, a warm site could be a remote office, a leased facility, or another organization with which yours has a reciprocal agreement.

Section 2.9 given a scenario; select the appropriate control to meet the goals of security.

Confidentiality

One of the major reasons to implement a cryptographic system is to ensure the confidentiality of the information being used. Confidentiality may be intended to prevent the unauthorized dis- closure of information in a local network or to prevent the unauthorized disclosure of information across a network. A cryptographic system must do this effectively in order to be of value.

The need to keep records secure from internal disclosure may be just as great as the need to keep records secure from outside attacks. The effectiveness of a cryptographic system in pre- venting unauthorized decryption is referred to as its strength: A strong cryptographic system is difficult to crack. Strength is also referred to as the algorithm's work factor: The work factor describes an estimate of the amount of time and effort that would be needed to break a system.

The system may be considered weak if it allows weak keys, has defects in its design, or is easily decrypted. Many systems available today are more than adequate for business and personal use, but they are inadequate for sensitive military or governmental applications.

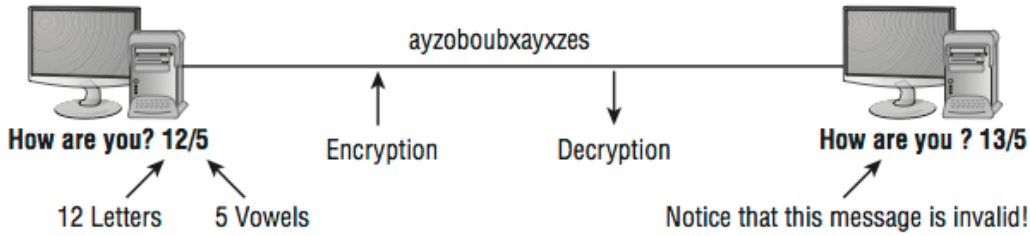
Cipher suites, for example, work with SSL/TLS to combine authentication, encryption, and message authentication. Most vendors allow you to set cipher suite preferences on a server to determine the level of strength required by client connections. With Sybase, for example, you set the cipher suite preference to Weak, Strong, FIPS, or All. If you choose Strong, you are limiting the choices to only encryption algorithms that use keys of 64 bits or more. Choosing Weak adds all the encryption algorithms that are less than 64 bits, while choosing FIPS requires encryptions, hash and key exchange algorithms to be FIPS- compliant (AES, 3DES, DES, and SHA1). Apache offers similar choices but instead of the words Strong and Weak, the names are changed to High, Medium, and Low.

Integrity

The second major reason for implementing a cryptographic system involves providing assurance that a message wasn't modified during transmission. Modification may render a message unintelligible or, even worse, inaccurate. Imagine the consequences if record alterations weren't discovered in medical records involving drug prescriptions. If a message is tampered with, the encryption system should have a mechanism to indicate that the message has been corrupted or altered.

Integrity can be accomplished by adding information such as redundant data that can be used as part of the decryption process. Figure 2.8 provides a simple example of how integrity can be validated in a message. Notice that data about the message's length and the number of vowels in the message are included in the message.

Figure 2.8: A simple integrity-checking process for an encrypted message

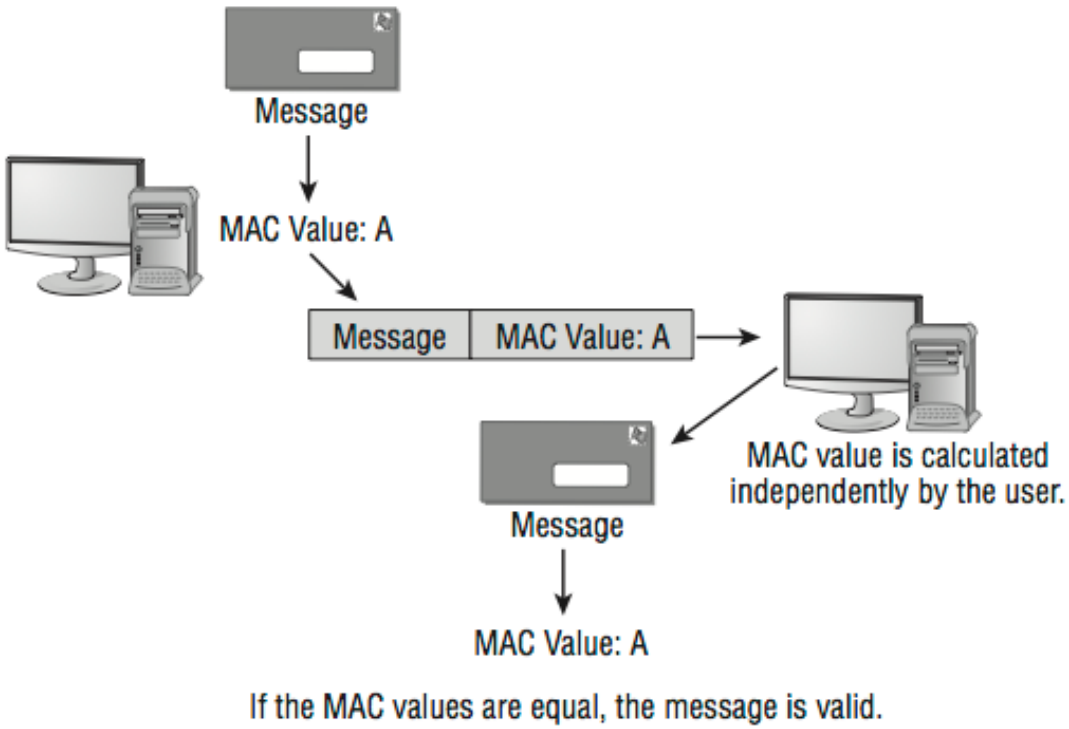


These two additions to the message provide a two-way check on the integrity of the message. In this case, the message has somehow become corrupted or invalidated. The original message had 12 characters; the decrypted message has 13 characters. Of course, the processes used in a real system are much more complicated. The addition of this information could be considered a signature of some sort.

A common method of verifying integrity involves adding a message authentication code (MAC) to the message. The MAC is derived from the message and a shared secret key. This process ensures the integrity of the message. The MAC would be encrypted with the message, adding another layer of integrity checking. From the MAC, you would know that the message came from the originator and that the contents haven't been altered. Figure 2.9 illustrates the MAC value being calculated from the message and included with the message.

The receiver also calculates the MAC value and compares it to the value sent in the message. If the values are equal, the message can be assumed to be intact and genuine.

FIGURE 2.9:The MAC value is calculated by the sender and receiver using the same algorithm.



Hashing
HMAC (Hash-Based Message Authentication Code) uses a hashing algorithm along with a symmetric key.

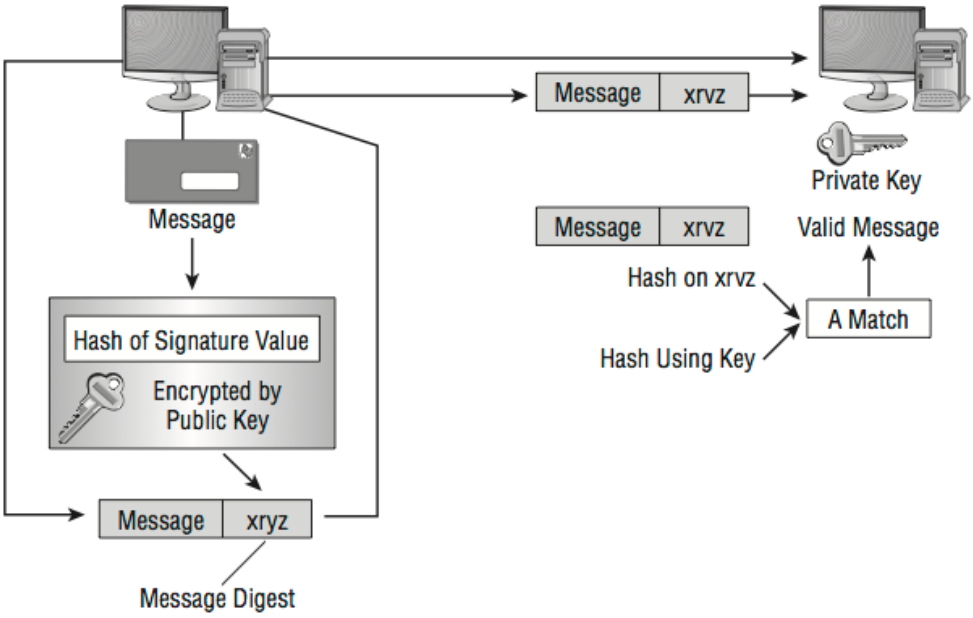
Digital signatures
A digital signature is similar in function to a standard signature on a document. It validates the integrity of the message and the sender. The message is encrypted using the encryption system, and a second piece of information, the digital signature, is added to the message. Figure 2.10 illustrates this concept.

Let's say that the sender in Figure 2.10 wants to send a message to the receiver. It's important that this message not be altered. The sender uses the private key to create a digital signature. The message is, in effect, signed with the private key. The sender then sends the message to the receiver. The receiver uses the public key attached to the message to validate the digital signature. If the values match, the receiver knows the message is authentic.

The receiver uses a key provided by the sender—the public key—to decrypt the message. Most digital signature implementations also use a hash to verify that the message has not been altered, intentionally or accidently, in transit.

The receiver compares the signature area referred to as a message digest in the message with the calculated value. If the values match, the message hasn't been tampered with and the originator is verified as the person they claim to be. This process provides message integrity, nonrepudiation, and authentication.

FIGURE 2.10: Digital signature processing steps



Non-repudiation

Nonrepudiation prevents one party from denying actions they carried out. To use an analogy, imagine coming home to find your house's picture window broken. All three of your kids say they didn't do it, and the babysitter says it must have been broken when she arrived. All the parties who could be guilty are "repudiating" the fact that they did it, and it's their word against common sense. Now, imagine that you had a nanny-cam running and were able to review the video and see who actually broke it. The video cancels out their saying that they knew nothing about the broken window and offers "nonrepudiation" of the facts.

In the electronic world, a similar type of proof can be achieved in a two-key system. The problem is that anyone can claim to be a legitimate sender, and if they have access to this type of system, they can send you a public key. So although you have received the message, you would have no way of verifying that the sender is really who they say they are, and you need nonrepudiation to verify that the sender is who they say they are.

Third-party organizations called certificate authorities (CAs) manage public keys and issue certificates verifying the validity of a sender's message. The verifying aspect serves as nonrepudiation; a respected third party vouches for the individual. The goal of any effective cryptography system must include nonrepudiation. However, the implementation is a little more difficult than the concept.

Safety

Security is a priority for every administrator, but it cannot be the only priority: security cannot forsake all others. One of the other priorities that CompTIA wants you to be familiar

with for this exam is safety. Some of the topics relate to the safety of the data or physical environment, while others are associated with safety of the individuals you work with.

In the order the topics appear in the objectives, these are as follows:

Fencing A fence can keep out unwanted entities—vehicles, people, etc.—and funnel those leaving to an easy to manage exit point where you can manage them easier.

Lighting An area that is not well lit can be more easily compromised than one which is.

Locks Locks are a lot like passwords—they need to be easy enough to work that those who are authorized can effectively navigate them but strong enough to keep those who are not authorized out. As a general rule, the strength of locks and the costs of them are closely related. Be sure to lock up not only the server room but also the wiring closets and physical hardware that could wander off.

CCTV Closed Circuit TV (CCTV) surveillance can help lessen the success of unauthorized access attempts. To be successfully used in prosecution, the recording equipment used with the cameras should be of good quality. To deter attempts, employees—and all others—should be made aware of the presence of the cameras.
Escape Plans With all the fencing, locks, and blinding lighting that is installed in the office, it is highly recommended that escape plans be in place and understood by all. Someone (a designated safety officer) should be responsible for keeping the plan current and making certain all employees are aware of it.

Drills To make certain not only that employees know the escape plan(s) but that it also works, drills should be conducted on a regular basis. The safety office in charge of the escape plans should also be responsible for the drills and make modifications as conditions change or problems arise.

Escape Routes The aforementioned escape plan and drills should direct the employees to safety via an escape route. Alternate routes should be identified in the event that the primary escape route is blocked.

Testing Controls Technical, Management, and Operational. Since there is reliance by at least two of these types on individuals, regularly test to verify that they are working properly and responses are appropriate.

Keep safety as one of your priorities as an administrator. Hope that you never have to respond to an emergency situation, but take comfort in knowing that your employees know how to respond should the need arise.

Topic 3 - Threats and Vulnerabilities

Section 3.1 Explain types of malware.

Malware (for "malicious software") is any program or file that is harmful to a computer user. Thus, malware includes computer viruses, worms, Trojan horses, and also spyware, programming that gathers information about a computer user without permission.

Adware
Generically, adware (spelled all lower case) is any software application in which advertising banners are displayed while the program is running. The authors of these applications include additional code that delivers the ads, which can be viewed through pop-up windows or through a bar that appears on a computer screen. The justification for adware is that it helps recover programming development cost and helps to hold down the cost for the user.

Adware has been criticized because it usually includes code that tracks a user's personal information and passes it on to third parties, without the user's authorization or knowledge. This practice has been dubbed spyware and has prompted an outcry from computer security and privacy advocates, including the Electronic Privacy Information Center.

Software expert Steve Gibson of Gibson Research explains: "Spyware is any software (that) employs a user's Internet connection in the background (the so-called 'backchannel') without their knowledge or explicit permission. Silent background use of an Internet 'backchannel' connection must be preceded by a complete and truthful disclosure of proposed backchannel usage, followed by the receipt of explicit, informed consent for such use. Any software communicating across the Internet absent of these elements is guilty of information theft and is properly and rightfully termed: Spyware."

A number of software applications, including Ad-Aware and Opt Out (by Gibson's company), are available as freeware to help computer users search for and remove suspected spyware programs.

AdWare is also a registered trademark that belongs to AdWare Systems, Inc. AdWare Systems builds accounting and media buying systems for the advertising industry and has no connection to pop-up advertising, spyware, or other invasive forms of online advertising.

Virus
A computer virus is an executable program. Depend on the nature of a virus, it may cause damage of your hard disk contents, and/or interfere normal operation of your computer.

By definition, a virus program is able to replicate itself. This means that the virus multiplies on a computer by making copies of itself. This replication is intentional; it is part of the virus program. In most cases, if a file that contains virus is executed or copied onto another computer, then that computer will also be "infected" by the same virus.

A virus can be introduced to a computer system along with any software program. For Internet users, this threat can come from downloading files through FTP (file transfer protocol), or referencing email attachments.

When a virus is introduced to a computer system, it can attach itself to, or sometimes even replace, an existing program. Thus, when the user runs the program in question, the virus is also executed. This usually happens without the user being aware of it.

A virus program contains instructions to initiate some sort of "event" that affects the infected computer. Each virus has an unique event associated with it. These events and their effects can range from harmless to devastating. For examples:

- An annoying message appearing on the computer screen.
- Reduced memory or disk space.
- Modification of data.
- Files overwritten or damaged.
- Hard drive erased.

Types of Viruses

There are many types of computer viruses:

- **File virus**: Most viruses' fall into this category. A virus attaches itself to a file, usually a program file.
- **Boot sector virus**: These viruses infect floppy and hard drives. The virus program will load first, before the operating system.
- **Macro Virus**: This is a new type of virus that use an application's own macro programming feature to distribute themselves. Unlike other viruses, macro viruses do not infect programs; they infect documents.
- **Virus Hoax:** Although there are thousands of viruses discovered each year, there are still some that only exist in the imaginations of the public and the press - known as virus hoaxes. These viruses' hoaxes do not exist, despite rumor of their creation and distribution.

Worms

Worms are programs that reproduce, execute independently and travel across the network connections. The key difference between a virus and worm is the manner in which it reproduces and spreads. A virus is dependent upon the host file or boot sector, and the transfer of files between computers to spread, whereas a computer worm can execute completely independently and spread on its own accord through network connections.

The security threat from worms is equivalent to that of viruses. Computer worms are skilled of doing an entire series of damage such as destroying crucial files in your system, slowing it down to a large degree, or even causing some critical programs to stop working. Two very prominent examples of worms are the MS-Blaster and Sasser worms.

Computer Worm Examples

The original computer worm was (perhaps accidentally) unleashed on the Internet by Robert Tappan Morris in 1988. The Internet Worm used sendmail, fingerd, and rsh/rexec to spread itself across the Internet.

The SQL Slammer Worm founded in 2003 used vulnerability in Microsoft SQL Server 2000 to spread itself across the Internet. The Blaster Worm also founded in 2003 used vulnerability in Microsoft DCOM RPC to spread itself.

The Melissa worm founded in 1999, the Sobig worms founded in 2003 and the Mydoom worm founded in 2004 all spread through e-mail. These worms shared some features of a Trojan Horse, in that they spread by tempting a user to open an infected e-mail attachment.

Mydoom also attempted to spread itself through the peer-to-peer file sharing application called KaZaA. The Mydoom worms attempted a Denial of Service (DoS) attack against SCO and Microsoft.

Spyware
Spyware is Internet jargon for Advertising Supported software (Adware). It is a way for shareware authors to make money from a product, other than by selling it to the users. There are several large media companies that offer them to place banner ads in their products in exchange for a portion of the revenue from banner sales. This way, you don't have to pay for the software and the developers are still getting paid. If you find the banners annoying, there is usually an option to remove them, by paying the regular licensing fee.

While this may be a great concept, the downside is that the advertising companies also install additional tracking software on your system, which is continuously "calling home", using your Internet connection and reports statistical data to the "mother ship". While according to the privacy policies of the companies, there will be no sensitive or identifying data collected from your system and you shall remain anonymous, it still remains the fact, that you have a "live" server sitting on your PC that is sending information about you and your surfing habits to a remote location

There are also many PC surveillance tools that allow a user to monitor all kinds of activity on a computer, ranging from keystroke capture, snapshots, email logging, chat logging and just about everything else. These tools are often designed for parents, businesses and similar environments, but can be easily abused if they are installed on your computer without your knowledge.

These tools are perfectly legal in most places, but, just like an ordinary tape recorder, if they are abused, they can seriously violate your privacy.

Trojan
Named after the Trojan Horse of ancient Greek history, a Trojan is a network software application designed to remain hidden on an installed computer. Trojans generally serve malicious purposes and are therefore a form of malware, like viruses.

Trojans sometimes, for example, access personal information stored locally on home or business computers then send these data to a remote party via the Internet. Alternatively, Trojans may serve merely as a "backdoor" application, opening network ports to allow other network applications access to that computer. Trojans are also capable of launching Denial of Service (DoS) attacks. A combination of firewalls and antivirus software protect networks against Trojans.

Trojans are similar to worms. In contrast to worms and viruses, however, Trojans do not replicate themselves or seek to infect other systems once installed on a computer.

Rootkits
A rootkit is a collection of tools (programs) that enable administrator-level access to a computer or computer network. Typically, a cracker installs a rootkit on a computer after first obtaining user-level access, either by exploiting a known vulnerability or cracking a password. Once the rootkit is installed, it allows the attacker to mask intrusion and gain root or privileged access to the computer and, possibly, other machines on the network.

A rootkit may consist of spyware and other programs that: monitor traffic and keystrokes; create a "backdoor" into the system for the hacker's use; alter log files; attack other machines on the network; and alter existing system tools to escape detection.

The presence of a rootkit on a network was first documented in the early 1990s. At that time, Sun and Linux operating systems were the primary targets for a hacker looking to install a rootkit. Today, rootkits are available for a number of operating systems, including Windows, and are increasingly difficult to detect on any network.

Rootkits have become more common and their sources more surprising. In late October of 2005, security expert Mark Russinovich of Sys-internals discovered that he had a rootkit on his own computer that had been installed as part of the digital rights management (DRM) component on a Sony audio CD. Experts worry that the practice may be more widespread than the public suspects and that attackers could exploit existing rootkits. "This creates opportunities for virus writers," said Mikko Hypponen, director of AV research for Finnish firm F-Secure Corp. "These rootkits can be exploited by any malware, and when it's used this way, it's harder for firms like ours to distinguish the malicious from the legitimate."

A number of vendors, including Microsoft, F-Secure, and Sys-internals, offer applications that can detect the presence of rootkits. If a rootkit is detected, however, the only sure way to get rid of it is to completely erase the computer's hard drive and reinstall the operating system.

Backdoors
Attackers who have compromised a system to ease their subsequent return to the system often install Backdoors. We consider the problem of identifying a large class of backdoors, namely those providing interactive access on non-standard ports, by passively monitoring a site's Internet access link. We develop a general algorithm for detecting interactive traffic based on packet size and timing characteristics, and a set of protocol-specific algorithms that look for signatures distinctive to particular protocols. We evaluate the algorithms on large Internet access traces and find that they perform quite well. In addition, some of the algorithms are amenable to pre-filtering using a stateless packet filter, which yields a major performance increase at little or no loss of accuracy. However, the success of the algorithms is tempered by the discovery that large sites have many users who routinely access what are in fact benign backdoors, such as servers running on non-standard ports not to hide, but for mundane administrative reasons. Hence, backdoor detection also requires a significant policy component for separating allowable backdoor access from surreptitious access.

Logic bomb

Logic bomb is a program, or portion of a program, which lies dormant until a specific piece of program logic is activated. In this way, a logic bomb is very analagous to a real-world land mine.

The most common activator for a logic bomb is a date. The logic bomb checks the system date and does nothing until a pre-programmed date and time is reached. At that point, the logic bomb activates and executes it's code.

A logic bomb could also be programmed to wait for a certain message from the programmer. The logic bomb could, for example, check a web site once a week for a certain message. When the logic bomb sees that message, or when the logic bomb stops seeing that message, it activates and executes it's code.

A logic bomb can also be programmed to activate on a wide variety of other variables, such as when a database grows past a certain size or a users home directory is deleted.

The most dangerous form of the logic bomb is a logic bomb that activates when something doesn't happen. Imagine a suspicious and unethical system administrator who creates a logic bomb, which deletes all of the data on a server if he doesn't log in for a month. The system administrator programs the logic bomb with this logic because he knows that if he is fired, he won't be able to get back into the system to set his logic bomb. One day on his way to work, a bus hits our suspicious and unethical system administrator. Three weeks later, his logic bomb goes off and the server is wiped clean. The system administrator meant for the logic bomb to explode if he was fired; he did not forsee that a bus would hit him.

Because a logic bomb does not replicate itself, it is very easy to write a logic bomb program. This also means that a logic bomb will not spread to unintended victims. In some ways, a logic bomb is the most civilized programmed threat, because a logic bomb must be targeted against a specific victim.

The classic use for a logic bomb is to ensure payment for software. If payment is not made by a certain date, the logic bomb activates and the software automatically deletes itself. A more malicious form of that logic bomb would also delete other data on the system.

Botnets

Botnet (also known as a zombie army) is a number of Internet computers that, although their owners are unaware of it, have been set up to forward transmissions (including spam or viruses) to other computers on the Internet. Any such computer is referred to as a zombie - in effect, a computer "robot" or "bot" that serves the wishes of some master spam or virus originator. Most computers compromised in this way are home-based. According to a report from Russian-based Kaspersky Labs, botnets -- not spam, viruses, or worms -- currently pose the biggest threat to the Internet. A report from Symantec came to a similar conclusion.

Computers that are coopted to serve in a zombie army are often those whose owners fail to provide effective firewalls and other safeguards. An increasing number of home users have high-speed connections for computers that may be inadequately protected. A zombie or bot is often created through an Internet port that has been left open and through which a small Trojan horse program can be left for future activation. At a certain time, the zombie army

"controller" can unleash the effects of the army by sending a single command, possibly from an Internet Relay Channel (IRC) site.

The computers that form a botnet can be programmed to redirect transmissions to a specific computer, such as a Web site that can be closed down by having to handle too much traffic - a distributed denial-of-service (DDoS) attack - or, in the case of spam distribution, to many computers. The motivation for a zombie master who creates a DDoS attack may be to cripple a competitor. The motivation for a zombie master sending spam is in the money to be made. Both of them rely on unprotected computers that can be turned into zombies.

According to the Symantec Internet Security Threat Report, through the first six months of 2006, there were 4,696,903 active botnet computers.

Ransomware

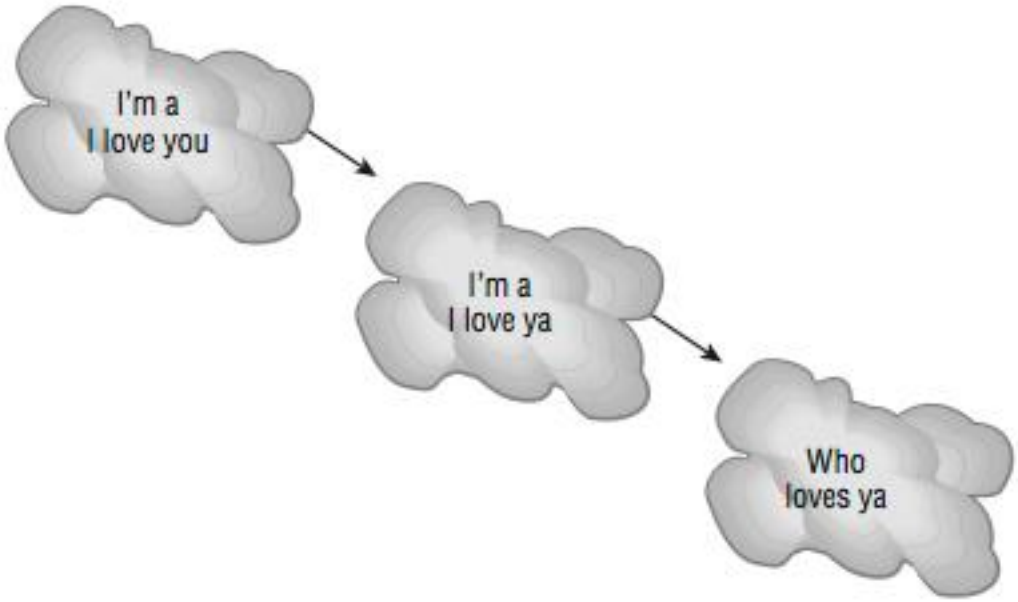
With *ransomware*, software—often delivered through a Trojan—takes control of a system and demands that a third party be paid. The "control" can be accomplished by encrypting the hard drive, by changing user password information, or via any of a number of other creative ways. Users are usually assured that by paying the extortion amount (the ransom) they will be given the code needed to revert their systems to normal operations.

Viruses come in many forms and are far more complicated than the other forms or malware.

Polymorphic malware

Polymorphic Virus Polymorphic viruses and polymorphic malware of any type—though viruses are the only ones truly prevalent—change form in order to avoid detection. These types of viruses attack your system, display a message on your computer, and delete files on your system. The virus will attempt to hide from your antivirus software. Frequently, the virus will encrypt parts of itself to avoid detection. When the virus does this, it's referred to as *mutation*. The mutation process makes it hard for antivirus software to detect common characteristics of the virus. Figure 3.1 uses a phrase to illustrate how the polymorphic virus changes characteristics to avoid detection. Like the phrase, small things within the virus are changed. In this example, the virus changes a signature to fool antivirus software.

FIGURE 3.1 The polymorphic virus changing its characteristics



Armored virus

An *armored virus* is designed to make itself difficult to detect or analyze. Armored viruses cover themselves with protective code that stops debuggers or dis-assemblers from examining critical elements of the virus. The virus may be written in such a way that some aspects of the programming act as a decoy to distract from analysis while the actual code hides in other areas in the program.

From the perspective of the creator, the more time it takes to deconstruct the virus, the longer it can live. The longer it can live, the more time it has to replicate and spread to as many machines as possible. The key to stopping most viruses is to identify them quickly and educate administrators about them—the very things that the armor intensifies the difficulty of accomplishing.

Section 3.2- Summarize various types of attacks.

A computer connected to a computing network is potentially vulnerable to an attack.

An "attack" is the exploitation of a flaw in a computing system (operating system, software program or user system) for purposes that are not known by the system operator and that are generally harmful.

Attacks are always taking place on the Internet, at a rate of several attacks per minute on each connected machine. These attacks are mostly launched automatically from infected machines (by viruses, Trojan horses, worms, etc.) without their owner's knowledge. In rarer cases, they are launched by computer hackers.

In order to block these attacks, it is important to be familiar with the main types of attacks so as to set up preventive measures.

Attacks may be launched for various reasons:

- to obtain access to the system;
- to steal information, such as industrial secrets or intellectual property;
- to gather personal information about a user;
- to retrieve bank account information;
- to get information about the organization (the user's company, etc.);
- to disrupt the proper functioning of a service;
- to use the user's system as a "bounce" for an attack;
- to use the resources of the user's system, particularly when the network on which it is located has a high bandwidth.

Since you need you know about these attacks to counter them, consider the following types of attacks and learn about them.

Man-in-the-middle

A man-in-the-middle attack, as the name implies, generally occurs when attackers are able to place themselves in the middle of two other hosts that are communicating. Ideally, this is done by ensuring that all communication going to or from the target host is routed through the attacker's host (which can be accomplished if the attacker can compromise the router for the target host). The attacker can then observe all traffic before relaying it and can actually modify or block traffic. To the target host, it appears that communication is occurring normally, since all expected replies are received.

The amount of information that can be obtained in a man-in-the-middle attack will obviously be limited if the communication is encrypted. Even in this case, however, sensitive information can still be obtained, since knowing what communication is being conducted, and between which individuals, may in fact provide information that is valuable in certain circumstances.

Man-in-the-Middle Attacks on Encrypted Traffic

The term "man-in-the-middle attack" is sometimes used to refer to a more specific type of attack—one in which the encrypted traffic issue is addressed. Public-key encryption, requires the use of two keys: your public key, which anybody can use to encrypt or "lock" your message, and your private key, which only you know and which is used to "unlock" or decrypt a message locked with your public key.

Denial-of-Service Attacks

Denial-of-service (DoS) attacks can exploit a known vulnerability in a specific application or operating system, or they can attack features (or weaknesses) in specific protocols or services. In a DoS attack, the attackers attempts to deny authorized users access either to specific information or to the computer system or network itself. This can be accomplished by crashing the system—taking it offline—or by sending so many requests that the machine is overwhelmed.

The purpose of a DoS attack can be simply to prevent access to the target system, or the attack can be used in conjunction with other actions to gain unauthorized access to a

computer or network. For example, a SYN flooding attack can be used to prevent service to a system temporarily in order to take advantage of a trusted relationship that exists between that system and another.

SYN flooding is an example of a DoS attack that takes advantage of the way TCP/IP networks were designed to function, and it can be used to illustrate the basic principles of any DoS attack. SYN flooding uses the TCP three-way handshake that establishes a connection between two systems. Under normal circumstances, the first system sends a SYN packet to the system with which it wants to communicate. The second system responds with a SYN/ACK if it is able to accept the request. When the initial system receives the SYN/ACK from the second system, it responds with an ACK packet, and communication can then proceed.

In a SYN flooding attack, the attacker sends fake communication requests to the targeted system. Each of these requests will be answered by the target system, which then waits for the third part of the handshake. Since the requests are fake (a nonexistent IP address is used in the requests, so the target system is responding to a system that doesn't exist), the target will wait for responses that never come. The target system will drop these connections after a specific time-out period, but if the attacker sends requests faster than the time-out period eliminates them, the system will quickly be filled with requests. The number of connections a system can support is finite, so when more requests come in than can be processed, the system will soon be reserving all its connections for fake requests. At this point, any further requests are simply dropped (ignored), and legitimate users who want to connect to the target system will not be able to do so, because use of the system has been denied to them.

DoS attacks are conducted using a single attacking system. A DoS attack employing multiple attacking systems is known as a distributed denial-of-service (DDoS) attack. The goal of a DDoS attack is also to deny the use of or access to a specific service or system. DDoS attacks were made famous in 2000 with the highly publicized attacks on eBay, CNN, Amazon, and Yahoo!

In a DDoS attack, service is denied by overwhelming the target with traffic from many different systems. A network of attack agents (sometimes called zombies) is created by the attacker, and upon receiving the attack command from the attacker, the attack agents commence sending a specific type of traffic against the target. If the attack network is large enough, even ordinary web traffic can quickly overwhelm the largest of sites, such as those targeted in 2000.

Creating a DDoS network is no simple task. The attack agents are not willing agents—they are systems that have been compromised and on which the DDoS attack software has been installed. To compromise these agents, the attacker has to have gained unauthorized access to the system or tricked authorized users to run a program that installed the attack software. The creation of the attack network may in fact be a multistep process in which the attacker first compromises a few systems that are then used as handlers or masters, which in turn compromise other systems. Once the network has been created, the agents wait for an attack message that will include data on the specific target before launching the attack. One important aspect of a DDoS attack is that with just a few messages to the agents, the attacker can have a flood of messages sent against the targeted system.

Replay Attacks

A replay attack occurs when the attacker captures a portion of a communication between two parties and retransmits it at a later time. For example, an attacker might replay a series of commands and codes used in a financial transaction to cause the transaction to be conducted multiple times. Generally replay attacks are associated with attempts to circumvent authentication mechanisms, such as the capturing and reuse of a certificate or ticket.

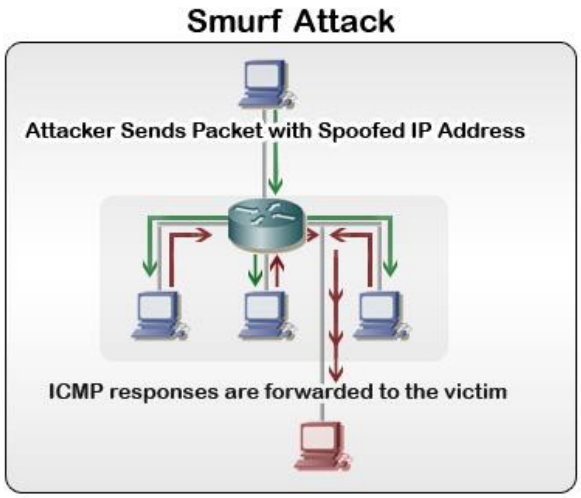
Smurf attack
A SMURF attack (named after the program used to perform the attack) is a method by which an attacker can send a moderate amount of traffic and cause a virtual explosion of traffic at the intended target. The method used is as follows:
The attacker sends ICMP Echo Request packets where the source IP address has been forged to be that of the target of the attack.

The attacker sends these ICMP datagrams to addresses of remote LANs broadcast addresses, using so-called directed broadcast addresses. These datagrams are thus broadcast out on the LANs by the connected router.
All the hosts which are «alive» on the LAN each pick up a copy of the ICMP Echo Request datagram (as they should), and sends an ICMP Echo Reply datagram back to what they think is the source. If many hosts are «alive» on the LAN, the amplification factor can be considerably (100+ is not uncommon).

The attacker can use largish packets (typically up to ethernet maximum) to increase the «effectiveness» of the attack, and the faster network connection the attacker has, the more damage he can inflict on the target and the target's network.

Not only can the attacker cause problems for the target host, the influx of traffic can in fact be so great as to have a seriously negative effect on the upstream network(s) from the target. In fact, those institutions being abused as amplifier networks can also be similarly affected, in that the Echo Reply packets destined for the target can swamp their network connection.

IP Address Spoofing
IP is designed to work so that the originators of any IP packet include their own IP address in the from portion of the packet. While this is the intent, nothing prevents a system from inserting a different address in from portion of the packet. This is known as IP address spoofing. An IP address can be spoofed for several reasons. In a specific DoS attack known as a smurf attack, the attacker sends a spoofed packet to the broadcast address for a network, which distributes the packet to all systems on that network. In the smurf attack, the packet sent by the attacker to the broadcast address is an echo request with from address forged so that it appears that another system (the target system) has made the echo request. The normal response of a system to an echo request is an echo reply, and it is used in the ping utility to let a user know whether a remote system is reachable and is responding. In the smurf attack, the request is sent to all systems on the network, so all will respond with an echo reply to the target system. The attacker has sent one packet and has been able to generate as many as 254 responses aimed at the target. Should the attacker send several of these spoofed requests, or send them to several different networks, the target can quickly become overwhelmed with the volume of echo replies it receives.



Spoofing and Sequence Numbers
How complicated the spoofing is depends heavily on several factors, including whether the traffic is encrypted and where the attacker is located relative to the target. Spoofing attacks from inside a network, for example, are much easier to perform than attacks from outside of the network, because the inside attacker can observe the traffic to and from the target and can do a better job of formulating the necessary packets.

Phishing
Phishing (pronounced "fishing") is a type of social engineering in which an individual attempts to obtain sensitive information from a user by masquerading as a trusted entity in an e-mail or instant message sent to the user. The type of information that the attacker attempts to obtain include usernames, passwords, credit card numbers, or details on the user's bank account. The message sent often encourages the user to go to a web site that appears to be for a reputable entity such as PayPal or eBay, both of which have frequently been used in phishing attempts. The web site the user actually visits will not be owned by the reputable organization, however, and will ask the user to supply information that can be used in a later attack. Often the message sent to the user will tell a story about the user's account having been compromised, and for security purposes they are encouraged to enter their account information to verify the details.

Vishing
Vishing is a variation of phishing that uses voice communication technology to obtain the information the attacker is seeking. Vishing takes advantage of the trust that most people place in the telephone network. Users are unaware that attackers can spoof calls from legitimate entities using voice over IP (VoIP) technology. Voice messaging can also be compromised and used in these attempts. Generally, the attackers are hoping to obtain credit card numbers or other information that can be used in identity theft. The user may receive an e-mail asking him to call a number that is answered by a potentially compromised voice message system. Users may also receive a recorded message that appears to come from a legitimate entity. In both cases, the user will be encouraged to respond quickly and provide the sensitive information so that access to an account is not blocked. If a user ever receives a message that claims to be from a reputable entity and is asking for sensitive information, he should not provide it but instead use theInternet or examine a legitimate account statement to find a phone number that can be used to contact the entity. The user can then verify that the message received was legitimate or report the vishing attempt.

DNS poisoning and ARP poisoning
There has been a long history of attacks on the Domain Name System ranging from brute-force denial-of-service attacks to targeted attacks requiring specialized software. In July 2008 a new DNS cache-poisoning attack was unveiled that is considered especially dangerous because it does not require substantial bandwidth or processor resources nor does it require sophisticated techniques.

With cache poisoning an attacker attempts to insert a fake address record for an Internet domain into the DNS. If the server accepts the fake record, the cache is poisoned and subsequent requests for the address of the domain are answered with the address of a server controlled by the attacker. For as long as the fake entry is cached by the server (entries usually have a time to live -- or TTL -- of a couple of hours) subscriber's browsers or e-mail servers will automatically go to the address provided by the compromised DNS server.

This kind of attack is often categorized as a "pharming" attack and it creates several problems. First, users think they are at a familiar site, but they aren't. Unlike with a "phishing" attack where an alert user can spot a suspicious URL, in this case the URL is legitimate. Remember, the browser resolves the address of the domain automatically so there is no intervention of any kind on the part of the users and, since nothing unusual has happened, they have no reason to be suspicious.

ARP poisoning
ARP Poisoning, also known as ARP Poison Routing, is a network attack that exploits the transition from Layer 3 to Layer 2 addresses.

ARP (address resolution protocol) operates by broadcasting a message across a network, to determine the Layer 2 address (MAC address) of a host with a predefined Layer 3 address (IP address). The host at the destination IP address sends a reply packet containing its MAC address. Once the initial ARP transaction is complete, the originating device then caches the ARP response, which is used within the Layer 2 header of packets that are sent to a specified IP address.

An ARP Spoofing attack is the egression of unsolicited ARP messages. These ARP messages contain the IP address of a network resource, such as the default gateway, or a DNS server, and replace the MAC address for the corresponding network resource with its own MAC address. Network devices, by design, overwrite any existing ARP information in conjunction with the IP address, with the new, counterfeit ARP information. The attacker then takes the role of man in the middle; any traffic destined for the legitimate resource is sent through the attacking system. As this attack occurs on the lower levels of the OSI model, the end-user is oblivious to the attack occurrence.

ARP Poisoning is also capable of executing Denial of Service (DoS) attacks. The attacking system, instead of posing as a gateway and performing a man in the middle attack, can instead simply drop the packets, causing the clients to be denied service to the attacked network resource. The spoofing of ARP messages is the tributary principal of ARP Poisoning.

Password attacks

Password attacks occur when an account is attacked repeatedly. This is accomplished by using applications known as password crackers, which send possible passwords to the account in a systematic manner. The attacks are initially carried out to gain passwords for an access or modification attack. There are several types of password attacks:

Brute-Force Attack
A brute-force attack is an attempt to guess passwords until a successful guess occurs. As an example of this type of attack, imagine starting to guess with "A" and then going through "z"; when no match is found, the next guess series goes from "AA" to "zz" and then adds a third value ("AAA" to "zzz"). Because of the nature of this routine, this type of attack usually occurs over a long period of time. To make passwords more difficult to guess, they should be much longer than two or three characters (six should be the bare minimum), be complex, and have password lockout policies.

Dictionary Attack
A dictionary attack uses a dictionary of common words to attempt to find the user's password. Dictionary attacks can be automated, and several tools exist in the public domain to execute them. As an example of this type of attack, imagine guessing words and word combinations found in a standard English-language dictionary.

Hybrid
A hybrid attack typically uses a combination of dictionary entries and brute force. For example, if you know that there is a good likelihood that the employees of a particular company are using derivatives of the company name in their passwords, then you can seed those values into the values attempted.

Birthday Attack
A birthday attack is built on a simple premise. If 25 people are in a room, there is some probability that two of those people will have the same birthday. The probability increases as additional people enter the room. It's important to remember that probability doesn't mean that something will occur, only that it's more likely to occur. To put it another way, if you ask if anyone has a birthday of March 9th, the odds are 1 in 365 (or 25/365 given the number of people in the room), but if you ask if anyone has the same birthday as any other individual, the odds of there being a match increase significantly.

Although two people may not share a birthday in every gathering, the likelihood is fairly high, and as the number of people increases, so too do the odds that there will be a match. A birthday attack works on the same premise: If your key is hashed, the possibility is that given enough time, another value can be created that will give the same hash value. Even encryption such as that with MD5 has been shown to be vulnerable to a birthday attack.

Rainbow Table
A rainbow table attack focuses on identifying a stored value. By using values in an existing table of hashed phrases or words (think of taking a word and hashing it every way you can imagine) and comparing them to values found, a rainbow table attack can reduce the amount of time needed to crack a password significantly. Salt (random bits added to the password) can greatly reduce the ease which rainbow can use tables.

Some systems will identify whether an account ID is valid and whether the password is wrong. Giving the attacker a clue as to a valid account name isn't a good practice.

Typo squatting/URL hijacking
Typo squatting (also spelled typosquatting) and *URL hijacking* are one and the same. Difficult to describe as an attack, this is the act of registering domains that are similar to those for a known entity but based on a misspelling or typographical error. com in the hopes that the same reader would misspell the word. Instead of arriving at the safe site of the publisher, they would end up at the other site, which could download Trojans, worms, and viruses—oh my.

The best defense against typo squatting is to register those domains around yours for which a user might intentionally type in a value when trying to locate you. This includes top-level domains as well (.com, .biz, .net, and so on) for all reasonable deviations of your site.

Watering hole attack
A watering hole attack can sound a lot more complicated than it really is. The strategy the attacker takes is simply to identify a site that is visited by those they are targeting, poisoning that site, and then waiting for the results.

As an example, suppose an attacker wants to gain unauthorized access to the servers at Spencer Industries, but Spencer's security is really good. The attacker discovers that Spencer does not host its own email, but instead outsources it to a big cloud provider, and so they focus their attention on the weaker security of the cloud provider. On the cloud provider's email site, they install the malware du jour, wait until a Spencer employee gets infected, and they suddenly have the access they coveted.

The best defense against a watering hole attack is to make certain that all of your partners are secure. Identify weak links, and bring them up to the same level of security as the rest of your infrastructure.

Section 3.3- Summarize social engineering attacks and the associated effectiveness with each attack.

A social engineering attack is one in which the intended victim is somehow tricked into doing the attacker's bidding. An example would be responding to a phishing email, following the link and entering your banking credentials on a fraudulent website. The stolen credentials are then used for everything from finance fraud to outright identity theft. An old adage comes to mind here, "it pays to be suspicious". With socially engineered attacks, the opposite is also true - if you aren't suspicious, you likely will end up paying.

In addition to phishing, social engineering attacks can come in many forms - email that masquerades as breaking news alerts, or greeting cards, or announcements of bogus lottery winnings. Pump and dump stock scams are also a form of social engineering, playing on the recipients' natural desire to take advantage of a good deal. It's important to remember that if something sounds too good to be true, it's probably a scam.

There can be a number of social engineering attacks. All you need to do is make right choices. They are depending on you to make a wrong choice and fall victim to their malicious intents.

Shoulder Surfing
Shoulder surfing does not involve direct contact with the user, but instead involves the attacker directly observing the target entering sensitive information on a form, keypad, or keyboard. The attacker may simply look over the shoulder of the user at work or the attacker can set up a camera or use binoculars to view users entering sensitive data. The attacker can attempt to obtain information such as a PIN at an automated teller machine, an access control entry code at a secure gate or door, or calling card or credit card numbers. Some locations now use a small shield to surround a keypad so that it is difficult to observe somebody entering information. More sophisticated systems can actually scramble the location of the numbers so that the top row at one time includes the numbers 1, 2, and 3 and the next time 4, 8, and 0. While this makes it a bit slower for the user to enter information, it does mean that a person attempting to observe what numbers are pressed will not be able to press the same buttons/pattern since the location of the numbers have changed.

Dumpster Diving
Dumpster diving is not a uniquely computer security–related activity. It refers to the activity of sifting through an individual's or organization's trash for things that the dumpster diver might find valuable. In the non-security realm, this can be anything from empty aluminum cans to articles of clothing or discarded household items. From a computer security standpoint, the diver is looking for information that can be obtained from listings or printouts, manuals, receipts, or even yellow sticky notes. The information can include credit card or bank account numbers, user IDs or passwords, details about the type of software or hardware platforms that are being used, or even company sensitive information. In most locations, trash is no longer considered private property after it has been discarded (and even where dumpster diving is illegal, little enforcement occurs). An organization should have policies about discarding materials. Sensitive information should be shredded and the organization should consider securing the trash receptacle so that individuals can't forage through it. People should also consider shredding personal or sensitive information that they wish to discard in their own trash. A reasonable quality shredder is inexpensive and well worth the price when compared with the potential loss that could occur as a result of identity theft.

Hoaxes
At first glance, it might seem that a hoax related to security would be considered a nuisance and not a real security issue. This might be the case for some hoaxes, especially those of the urban legend type, but the reality of the situation is that a hoax can be very damaging if it causes users to take some sort of action that weakens security. One real hoax, for example, told the story of a new, highly destructive piece of malicious software. It instructed users to check for the existence of a certain file and to delete it if the file was found. In reality, the file mentioned was an important file that was used by the operating system, and deleting it caused problems the next time the system was booted. The damage caused by users modifying security settings can be serious. As with other forms of social engineering, training and awareness are the best and first line of defense for users. Users should be trained to be suspicious of unusual e-mails and stories and should know who to contact in the organization to verify the validity if they are received.

Vishing
Vishing is a variation of phishing that uses voice communication technology to obtain the information the attacker is seeking. Vishing takes advantage of the trust that most people place in the telephone network. Users are unaware that attackers can spoof calls from legitimate entities using voice over IP (VoIP) technology. Voice messaging can also be compromised and used in these attempts. Generally, the attackers are hoping to obtain credit card numbers or other information that can be used in identity theft. The user may receive an e-mail asking him to call a number that is answered by a potentially compromised voice message system. Users may also receive a recorded message that appears to come from a legitimate entity. In both cases, the user will be encouraged to respond quickly and provide the sensitive information so that access to an account is not blocked. If a user ever receives a message that claims to be from a reputable entity and is asking for sensitive information, he should not provide it but instead use theInternet or examine a legitimate account statement to find a phone number that can be used to contact the entity. The user can then verify that the message received was legitimate or report the vishing attempt.

Principles (reasons for effectiveness)
A number of principles, or elements, allow social engineering attacks to be effective. Most of these are based on our nature to be help-ful, to trust others in general, and to believe that there is a hierarchy of leadership that should be followed. For the exam, be familiar with the following reasons for its effectiveness:

Authority
If it is possible to convince the person you are attempting to trick that you are in a position of authority, they may be less likely to question your request. That position of authority could be upper management, tech support, HR, or law enforcement.

Intimidation
Although authority can be a source of intimidation, it is possible for intimidation to occur in its absence as well. This can be done with threats, with shouting, or even with guilt.

Consensus/Social Proof
Putting the person being tricked at ease by putting the focuson them—listening intently to what they are saying, validating their thoughts, charming them—is the key to this element. The name comes from a desire that we all have to be told that we are right, attractive, intelligent, and so forth, and we tend to be fond of those who confirm this for us. By being so incredibly nice, the social engineer convinces the other party that there is no way their intentions could possibly be harmful.

Discussions at home with a spouse, or casual conversations with associates where we are bragging or trying to impress others, can lead to sharing more information than we should.

Scarcity
Convincing the person who is being tricked that there is a limited supply of something can often be effective if carefully done. For example, convincing them that there are only one hundred vacation requests that will be honored for the entire year and that they need to go to a fictitious website now and fill out their information (including user- name and password, of course) if they want to take a vacation anytime during the current year, can dupe some susceptible employees.

More than one principle can be used in any given attack. It is not un-common, for example, to see both scarcity and urgency used together.

Urgency
The secret for successfully using the urgency element is for the social engineer to convince the individual they are attempting to trick that time is of the essence. If they don't do something right away, money will be lost, a nonexistent intruder will get away, the company will suffer irreparable harm, or a plethora of other negative possibilities may occur.

Familiarity/Liking
Mental guards are often lowered, many times subconsciously, when we are dealing with other individuals that we like. The "like" part can be gained by some- one having, or pretending to have, the same interests as we do, be engaged in the same activities, or otherwise working to gain positive attention.

Trust
One of the easiest ways to gain trust is through reciprocation. When someone does something for you, there is often a feeling that you owe that person something. For example, to gain your trust someone may help you out of a troublesome situation or buy you lunch.

Section 3.4-Explain types of wireless attacks.

Wireless networking, more commonly termed as Wi-Fi, is the technology that opens your PDA or laptop computer to the world. However this technology is quite vulnerable to many exploits. A malicious intruder can use the most basic software to detect and capture the signal of your wireless device, along with usernames, passwords, emails and other data you would prefer to keep confidential.

An intruder doesn't have to be inside of your home or office building to manipulate a wireless signal. For example, they could be sitting outside in their car sniffing out your data all while enjoying a sandwich. Before they have a chance to complete the meal, the intruder can learn just who you work for, how to access the companynetwork or even transfer money out of your bank account if the right security is not implemented.

Being that wireless technology is so vulnerable, it is important that you take various measures to protect your personal information.

Access control attacks
These attacks attempt to penetrate a network by using wireless or evading WLAN access control measures, like AP MAC filters and 802.1X port access controls.

Type of Attack	Description	Methods and Tools
War Driving	Discovering wireless LANs by listening to beacons or sending probe requests, thereby providing launch point for further attacks.	Airmon-ng, DStumbler, KisMAC, MacStumbler, NetStumbler, Wellenreiter, WiFiFoFum

Rogue Access Points	Installing an unsecured AP inside firewall, creating open backdoor into trusted network.	Any hardware or software AP
Ad Hoc Associations	Connecting directly to an unsecured station to circumvent AP security or to attack station.	Any wireless card or USB adapter
MAC Spoofing	Reconfiguring an attacker's MAC address to pose as an authorized AP or station.	MacChanger, SirMACsAlot, SMAC, Wellenreiter, wicontrol
802.1X RADIUS Cracking	Recovering RADIUS secret by brute force from 802.1X access request, for use by evil twin AP.	Packet capture tool on LAN or network path between AP and RADIUS server

Confidentiality attacks

These attacks attempt to intercept private information sent over wireless associations, whether sent in the clear or encrypted by 802.11 or higher layer protocols.

Type of Attack	Description	Methods and Tools
Eavesdropping	Capturing and decoding unprotected application traffic to obtain potentially sensitive information.	bsd-airtools, Ettercap, Kismet, Wireshark, commercial analyzers
WEP Key Cracking	Capturing data to recover a WEP key using passive or active methods.	Aircrack-ng, airoway, AirSnort, chopchop, dwepcrack, WepAttack, WepDecrypt, WepLab, wesside
Evil Twin AP	Masquerading as an authorized AP by beaconing the WLAN's service set identifier (SSID) to lure users.	cqureAP, D-Link G200, HermesAP, Rogue Squadron, WifiBSD
AP Phishing	Running a phony portal or Web server on an evil twin AP to "phish" for user logins, credit card numbers.	Airpwn, Airsnarf, Hotspotter, Karma, RGlueAP
Man in the Middle	Running traditional man-in-the-middle attack tools on an evil twin AP to intercept TCP sessions or SSL/SSH tunnels.	dsniff, Ettercap-NG, sshmitm

Integrity attacks

These attacks send forged control, management or data frames over wireless to mislead the recipient or facilitate another type of attack (e.g., DoS).

Type of Attack	Description	Methods and Tools
802.11 Frame Injection	Crafting and sending forged 802.11 frames.	Airpwn, File2air, libradiate, void11, WEPWedgie, wnet dinject/reinject
802.11 Data Replay	Capturing 802.11 data frames for later (modified) replay.	Capture + Injection Tools
802.1X EAP Replay	Capturing 802.1X Extensible Authentication Protocols (e.g., EAP	Wireless Capture + Injection Tools between station and AP

	Identity, Success, Failure) for later replay.	
802.1X RADIUS Replay	Capturing RADIUS Access-Accept or Reject messages for later replay.	Ethernet Capture + Injection Tools between AP and authentication server

Authentication attacks
Intruders use these attacks to steal legitimate user identities and credentials to access otherwise private networks and services.

Type of Attack	**Description**	**Methods and Tools**
Shared Key Guessing	Attempting 802.11 Shared Key Authentication with guessed, vendor default or cracked WEP keys.	WEP Cracking Tools
PSK Cracking	Recovering a WPA/WPA2 PSK from captured key handshake frames using a dictionary attack tool.	coWPAtty, genpmk, KisMAC, wpa_crack
Application Login Theft	Capturing user credentials (e.g., e-mail address and password) from cleartext application protocols.	Ace Password Sniffer, Dsniff, PHoss, WinSniffer
Domain Login Cracking	Recovering user credentials (e.g., Windows login and password) by cracking NetBIOS password hashes, using a brute-force or dictionary attack tool.	John the Ripper, L0phtCrack, Cain
VPN Login Cracking	Recovering user credentials (e.g., PPTP password or IPsec Preshared Secret Key) by running brute-force attacks on VPN authentication protocols.	ike_scan and ike_crack (IPsec), anger and THC-pptp-bruter (PPTP)
802.1X Identity Theft	Capturing user identities from cleartext 802.1X Identity Response packets.	Capture Tools
802.1X Password Guessing	Using a captured identity, repeatedly attempting 802.1X authentication to guess the user's password.	Password Dictionary
802.1X LEAP Cracking	Recovering user credentials from captured 802.1X Lightweight EAP (LEAP) packets using a dictionary attack tool to crack the NT password hash.	Anwrap, Asleap, THC-LEAPcracker
802.1X EAP Downgrade	Forcing an 802.1X server to offer a weaker type of authentication using forged EAP-Response/Nak packets.	File2air, libradiate

Availability attacks
These attacks impede delivery of wireless services to legitimate users, either by denying them access to WLAN resources or by crippling those resources.

Type of Attack	**Description**	**Methods and Tools**
AP Theft	Physically removing an AP from a	"Five finger discount"

	public space.	
Queensland DoS	Exploiting the CSMA/CA Clear Channel Assessment (CCA) mechanism to make a channel appear busy.	An adapter that supports CW Tx mode, with a low-level utility to invoke continuous transmit
802.11 Beacon Flood	Generating thousands of counterfeit 802.11 beacons to make it hard for stations to find a legitimate AP.	FakeAP
802.11 Associate / Authenticate Flood	Sending forged Authenticates or Associates from random MACs to fill a target AP's association table.	FATA-Jack, Macfld
802.11 TKIP MIC Exploit	Generating invalid TKIP data to exceed the target AP's MIC error threshold, suspending WLAN service.	File2air, wnet dinject, LORCON
802.11 Deauthenticate Flood	Flooding station(s) with forged Deauthenticates or Disassociates to disconnecting users from an AP.	Aireplay, Airforge, MDK, void11, commercial WIPS
802.1X EAP-Start Flood	Flooding an AP with EAP-Start messages to consume resources or crash the target.	QACafe, File2air, libradiate
802.1X EAP-Failure	Observing a valid 802.1X EAP exchange, and then sending the station a forged EAP-Failure message.	QACafe, File2air, libradiate
802.1X EAP-of-Death	Sending a malformed 802.1X EAP Identity response known to cause some APs to crash.	QACafe, File2air, libradiate
802.1X EAP Length Attacks	Sending EAP type-specific messages with bad length fields to try to crash an AP or RADIUS server.	QACafe, File2air, libradiate

Near field communication

Near field communication (NFC) is a technology that requires a user to bring the client close to the AP in order to verify (often through RFID or Wi-Fi) that the device is present. It can also be used to "bump" phones and send data from one to another.

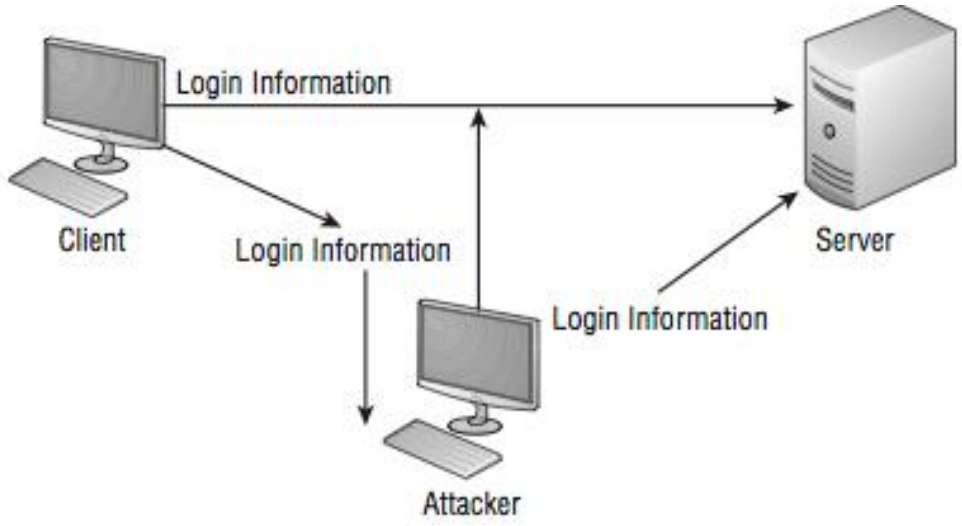
Replay attacks

Replay attacks are becoming quite common. They occur when information is captured over a network. A replay attack is a kind of access or modification attack. In a distributed environment, logon and password information is sent between the client and the authentication system. The attacker can capture the information and replay it later. This can also occur with security certificates from systems such as Kerberos: The attacker resubmits the certificate, hoping to be validated by the authentication system and circumvent any time sensitivity.

Figure 3.2 shows an attacker presenting a previously captured certificate to a Kerberos-enabled system. In this example, the attacker gets legitimate information from the client and

records it. Then the attacker attempts to use the information to enter the system. The attacker later relays information to gain access.

FIGURE 3.2 A replay attack occurring



If this attack is successful, the attacker will have all of the rights and privileges from the original certificate. This is the primary reason that most certificates contain a unique session identifier and a time stamp. If the certificate has expired, it will be rejected and an entry should be made in a security log to notify system administrators.

WEP/WPA attacks

Wireless Application Protocol (WAP) is a technology designed for use with wireless devices. WAP has become a data transmission standard adopted by many manufacturers, including Motorola and Nokia. WAP functions are equivalent to TCP/IP functions in that they're attempting to serve the same purpose for wireless devices. WAP uses a smaller version of HTML called Wireless Markup Language (WML), which is used for Internet displays. WAP-enabled devices can also respond to scripts using an environment called WMLScript. This scripting language is similar to the Java programming language.

The ability to accept web pages and scripts allows malicious code and viruses to be transported to WAP-enabled devices.

The gateway converts information back and forth between HTTP and WAP as well as encodes and decodes between the protocols.

This structure provides reasonable assurance that WAP-enabled devices can be secured. If the interconnection between the WAP server and the Internet isn't encrypted, packets between the devices may be intercepted, referred to as packet sniffing, creating a potential vulnerability. This vulnerability is called a gap in the WAP (the security concern that exists when converting between WAP and SSL/TLS and exposing plain text). It was prevalent in versions of WAP prior to 2.0.

WPS attacks

To simplify network setup, a number of small office and home office (SOHO) routers use a series of EAP messages to allow new hosts to join the network and use WPA/WPA2. Known as Wi-Fi Protected Setup (WPS), this often requires the user to do something in order to complete the enrollment process: press a button on the router within a short time period, enter a PIN, or bring the new device close-by (so that near field communication can take place).

Near field communication (NFC) is a technology that requires a user to bring the client close to the AP in order to verify (often through RFID or Wi-Fi) that the device is present. It can also be used to "bump" phones and send data from one to another.
Unfortunately, WPS attacks have become commonplace, as the technology is susceptible to brute-force attacks used to guess the user's PIN. Once an attacker gains access, they are then on the Wi-Fi network. For that reason, we suggest that you disable WPS in devices that allow it (and update firmware in those where it is a possibility).

Section 3.5- Explain types of application attacks.

Applications such as Content Management Systems (CMS), Wikis, Portals, Bulletin Boards, and discussion forums are being used by small and large organizations. Every week hundreds of vulnerabilities are being reported in these web applications, and are being actively exploited. The number of attempted attacks every day for some of the large web hosting farms range from hundreds of thousands to even millions.

All web frameworks (PHP, .NET, J2EE, Ruby on Rails, ColdFusion, Perl, etc) and all types of web applications are at risk from web application security defects, ranging from insufficient validation through to application logic errors.

The most exploited vulnerabilities are:

PHP Remote File Include:
PHP is the most common web application language and framework in use today. By default, PHP allows file functions to access resources on the Internet using a feature called "allow_url_fopen". When PHP scripts allow user input to influence file names, remote file inclusion can be the result. This attack allows (but is not limited to):

- Remote code execution
- Remote root kit installation
- On Windows, internal system compromise may be possible through the use of PHP's SMB file wrappers

SQL Injection: Injections, particularly SQL injections, are common in web applications. Injections are possible due to intermingling of user supplied data within dynamic queries or within poorly constructed stored procedures. SQL injections allow attackers:

- To create, read, update, or delete any arbitrary data available to the application
- In the worst case scenario, to completely compromise the database system and systems around it

Cross-site scripting
Can occur when a Web application sends user data to a Web browser without first encoding or validating it. Flaws in XSS allow attackers to pass in a script as user data that is then executed in the user's browser. Possible consequences include user session hijack, phishing, the introduction of worms and website defacement.

Cross-site request forgeries (CSRF):
CSRF forces legitimate users to execute commands without their consent. This type of attack is extremely hard to prevent unless the application is free of cross-site scripting vectors, including DOM injections. With the rise of Ajax techniques, and better knowledge of how to properly exploit XSS attacks, CSRF attacks are becoming extremely sophisticated, both as an active individual attack and as automated worms, such as the Samy MySpace Worm.

Directory Traversal:
Directory traversal (file access via ".." or many encoded variants) allows attackers access to controlled resources, such as password files, configuration files, database credentials or other files of the attacker's choosing.

Buffer overflow
The Buffer Overflow attack can be applied in different areas : users entries, parameters
Example:

http://www.test.com/insecurecgi?ABCDEF..ABCDEFcode_excutable Note that the shell code first contains a large number of characters, as well as code in binary and executable form near the end. In this example, the overflow is in the name of the parameter and not in its value, which illustrate how many numerous the overflow possibilities are.

Integer overflow
An integer overflow, like a buffer overflow, involves putting too much information into too small of a space. In this case, the space is that set aside for numbers.
For example, using 8 bits, it is possible to express any number in binary from 0 to 255. If only 8 bits are set aside and the user enters a value of 256 to be converted to binary, it exceeds what can be stored, represented, and so forth, and results in an integer overflow. Depending on how the code is written, it is possible that the program would store only the last eight digits (of what now requires nine—100000000) and thus the value would be accepted, processed, and stored as zero.

Zero day
A zero day attack, also known as a zero hour attack, takes advantage of computer vulnerabilities that do not currently have a solution. Typically, a software company will discover a bug or problem with a piece of software after it has been released and will offer a patch — another piece of software meant to fix the original issue. A zero day attack will take advantage of that problem before a patch has been created. It is named zero day because it occurs before the first day the vulnerability is known.

In most cases, a zero day attack will take advantage of a bug that neither the software's creators nor users are aware of. In fact, this is precisely what malicious programmers hope to find. By finding software vulnerabilities before the software's makers find them, a

programmer can create a virus or worm that exploits that vulnerability and harms computer systems in a variety of ways.

Cookies and attachments

Cookies are text files that a browser maintains on the user's hard disk in order to provide a persistent, customized web experience for each visit. A cookie typically contains information about the user. For example, a cookie can contain a client's history to improve customer service. The next time you return to that store, the server can read your cookie and customize what it presents to you. Cookies can also be used to timestamp a user to limit access. A financial institution may send your browser a cookie once you've authenticated. The server can read the cookie to determine when a session is expired.

Obviously, cookies are considered a risk because they have the potential to contain your personal information, which could get into the wrong hands, and are highly treasured by advertisers today.

If security is your utmost concern, the best protection is to not allow cookies to be accepted. Almost every browser offers the option of enabling or disabling cookies. If you enable them, you can usually choose whether to accept or reject all or only those from an originating server. Know that if you disallow cookies, users will not be able to visit a lot of sites. A compromise is to allow only session cookies.

Locally Shared Objects and Flash Cookies

A Locally Shared Object (LSO) is also commonly known as a Flash Cookie and is nothing more than data stored on a user's computer by Adobe Flash. Often this is used to store data from games that have been played through Flash or user preferences, and it can represent a security/privacy threat.

Malicious Add-Ons

There are any numbers of add-ons that have the potential to harm a system. Some do so unintentionally through poor programming, and some are truly malicious add-ons; the difference between them is intent.

Consider a Java applet, for example. This is a small, self-contained Java script that is downloaded from a server to a client and then run from the browser. The client browser must have the ability to run Java applets in a virtual machine on the client. Java applets are used extensively in web servers today, and they're becoming one of the most popular tools used for website development.

Java-enabled applications can accept programmed instructions (Java scripts) from a server and control certain aspects of the client environment. Java requires you to download a virtual machine in order to run the Java applications or applets. Java scripts run on the client.

The applets run in a restricted area of memory called the sandbox. The sandbox limits the applet's access to user areas and system resources. An applet that runs in the sandboxis considered safe, meaning that it won't attempt to gain access to sensitive system areas. Errors in the Java virtual machine that runs in the applications may allow some applets to run outside the sandbox. When this occurs, the applet is unsafe and may perform malicious operations. Attackers on client systems have exploited this weakness. From a user's stand-

point, the best defense is to make certain that you run only applets from reputable sites with which you're familiar. From an administrator's standpoint, you should make certain that programmers adhere to programming guidelines when creating such applets.

Similarly, ActiveX is a technology that was implemented by Microsoft to customize controls, icons, and other features, which increases the usability of web-enabled systems. ActiveX runs on the client. It uses a method called Authenticode for security. Authenticode is a type of certificate technology that allows ActiveX components to be validated by a server.

ActiveX components are downloaded to the client hard disk, potentially allowing additional security breaches. Web browsers can be configured so that they require confirmation to accept an ActiveX control. However, many users don't understand these confirmation messages when they appear, and they automatically accept the components. Automatically accepting an ActiveX component or control creates the opportunity for security breaches on a client system when the control is used because an ActiveX control contains programming instructions that can contain malicious code or create vulnerabilities in a system.

We highly recommend that you configure browsers so that they do not allow ActiveX to run without prompting the user because of the potential security hole that could be opened.

Session Hijacking
The term session hijacking describes when the item used to validate a user's session, such as a cookie, is stolen and used by another to establish a session with a host that thinks it is still communicating with the first party. To use an overly simplistic analogy, imagine that you just finished a long phone conversation with a family member and then accidentallyleft your smartphone in the room while stepping outside. If Jim were to pick up that phone and press redial, the family member would see the caller ID, know that they had just been talking with you, and falsely assume that you were calling back. If Jim could imitate your voice, he could rattle off numerous nasty comments that would jeopardize your relationship with that family member. This same premise could be true if someone could fool a host into thinking it was still talking to your computer rather than theirs.

Numerous types of attacks use session hijacking, including man-in-the-middle and sidejacking. A weakness in a Firefox extension made news when it became known that an exploit made it possible for public Wi-Fi users to fall prey to this type of attack (Firesheep was an extension created to take advantage of the weakness).
Some of the best ways to prevent session hijacking are to encrypt the sessions, encourage users to log out of sites when finished, and perform secondary checks on the identity of the user.

Header Manipulation
When used with XSRF, the attacker can even change a user's cookie. Internet Explorer 8 and above include InPrivate Filtering to help prevent some of this. By default, your browser sends information to sites, as they need it—think of requesting a map from a site; it needs to

Arbitrary code execution / remote code execution

Though long frowned upon, it is possible for a programmer to create a means by which a program that they write can remotely accept commands and execute them. These commands can be unrelated to the actual program accepting them, and they can run on the host machine

within a shell, command interpreter, and so on. When this is done, it is known as either arbitrary code execution (since it is taking any arbitrary commands fed to it) or remote code execution—both meaning the same thing.

As if this issue is not bad enough in and of itself, the host program can be running with elevated privileges and capable of doing far more harm than what the user might otherwise be limited to.

Section 3.6- Analyze a scenario and select the appropriate type of mitigation and deterrent techniques.

Monitoring system logs
The general goal of monitoring is to detect suspicious behavior by external users or employees, or malfunctions. An organization can do this directly, such as by monitoring for specific events, or indirectly, such as by watching the state of a server over time and investigating anomalous behavior.

Your security organization will have to determine its specific monitoring policy. Within this policy, you will have to determine your organization's specific monitoring goals. Some questions you will have to answer are:

- Are you going to baseline your server's performance? If so, what counters are you going to collect and at what interval? How often are you going to take baselines?
- How are you going to manage your event logs? Are you going to use the tools Microsoft provides, write your own, or purchase a third-party system?
- Which monitoring technologies are you going to use?
- How much are you going to monitor? Monitoring a system has a distinct impact on system performance—the more you query or log a system's state, the more resources the system must expend on these activities.)

Event logs
Event log data is reported to the Event Log service by other parts of the system or by applications running on the system. The Event Log service stores this data in .evt files in the %systemroot%\system32\config directory. The built-in logs in Windows NT 4.0 and Windows 2000 Professional are the system log, the security log, and the application log. Windows 2000 Server installations may add logs for Domain Name System (DNS) and directory services. Any application that needs to log events should register itself with the Event Log service.

Events stored in the event logs have a description field that displays text data describing the event. Typically, this is the text of an error message. Usually, only information unique to a particular instance of an event is stored with the event in the log; the text strings are stored in

an event message file (usually a .dll or .exe file) that is registered with the Event Log service. If, when viewing an event, you see an error in the description field indicating that the event text is not available, you will need to register the event message file with the Event Log service on the computer at which you are trying to view the event. More information on this is available at http://support.microsoft.com/default.aspx?scid=kb;en-us;165959&sd=tech

The Event Log service is automatically started automatically when windows machine starts. All users can view application and system logs. Only administrators can gain access to security logs. Security logging is turned off by default. To ensure that a security log is available the administrator should turn it on.

Windows has several different logs that should be monitored. The most important log being the security log to the security professional as this log tracks the on goings on the network. The different log types are:

Application log these are events logged by applications.

1. Security log this log contains records of valid and invalid logon attempts and events related to resources use, such as creating, opening, or deleting files or other objects. This log is also customizable.
2. System log contains system component event. Driver failures and hardware issues.
3. Domain controllers have two extra logs directory service directory service.
4. File Replication service log containing windows File Replication service events. Sysvol changes are recorded in the file replication log.
5. DNS machines also store DNS events in the logs.

Each log contains different types of logs i.e. Errors, warnings, information, success audit and failure audits. It has become apparent that a third party automation tool is necessary, on any busy machine or on any busy network many hours are logged and megabytes of log files are generated, this makes it logically impossible to monitor all of the logs on all of the networked computers with limited resources.

Below are a few valuable features that prove useful when monitoring logs.

1. Real time monitoring and notification, if events happen that need to come to the security professional's attention. Windows is unable to notify the security official of triggered events.
2. Audit trail is unconsolidated in windows. This means that individual machines hold the isolated event logs making the task of viewing event logs extremely difficult. It is much easier to look at one event log to get a current network status than to look at multiple event logs and miss information because of the vast amount of entries that have not been filtered. So it is ideal to have a central log monitoring system that the security professional can use at a glance.
3. Security logs are also able to be monitored remotely, this means that when intruders attempt to use local accounts to log into the machine the audit trail is limited to the local security logs.
4. Less obvious description of critical event. In normal Microsoft tradition "event 12345%$# means your server was rebooted or something like that." Logs are cryptic and misleading. Consolidation and remote log reading applications have alerts that

can be preprogrammed for specific events to make the administrators life much easier deciphering the misleading logs.

5. Archiving. Institutions such as banks are required in most countries to keep audit logs for over 7 years and even longer in some circumstances. Typical windows default setting is set to overwrite over the logs when certain size is reached. The other issue is that the user has to physically archive and clear the logs. Automation of this process is available and making it central, increasing productivity time on a large network environment as it lessens support calls and lets the administrators see what is happening locally on the user's machine.
6. Log file integrity. Files stored on a user machine have less integrity as the user can clear the logs quickly or an intruder after gaining access can cover the tracks by clearing the event logs. Intruders sometimes produce an excessive
   Amount of events triggering actions to fill up security logs to cover tracks. Using the consolidation and remote log viewing applications, the security professional can be alerted to this phenomenon and can react to it immediately; further more he logs are stored remotely so the user or intruder can not erase them. Applications exist on the Internet that render local machine logs useless as they can create vast amounts of traffic and fill the logs with garbage or delete them completely.

7. Log filtering. Data overload is a huge issue log monitoring applications have the ability to filter out irrelevant noise events that take up time and space and only display the pertinent logs.

1. The ability to monitor access of important files this can be achieved by auditing failed access to these files enables you to find out if someone is attempting to access the files.
2. An application that can alert the security professional by SMS (mobile phone) e-mail and pager prove valuable as the Administrator may not be in the proximity of a computer at all times this should trigger a response. The administrator can then react or have systems in place the can be remotely activated to stop a potential attack.
3. Monitoring of web server log is important and should be mentioned as an isolated point as this is often overlooked by hasty administrators. By using software that monitors your local or remote web server you can add an extra layer of security to your web server. This is where the alerting functionality of log monitoring software is useful because it sometimes is challenging to monitor servers that are on the DMZ.
4. Logging of data in powerful searchable databases like SQL is an advantage and would be preferred in an enterprise environment the most good centralized logging software available does provide this type of functionality.
5. Reporting using well known tools like Crystal is also need in large organizations as trends are easier to see depicted. Log monitoring software should have the capability to link to crystal reports and other well known reporting software.
6. Categorically sorting log events into prioritized sections. Software should be able to let the security administrator view high profile security events at a glimpse, medium profile or low profile security events have taken place this saves time and makes for good managerial reporting.
7. Clearing of logs should also be monitored as only the administrator should be able to clear security logs.
8. The ability to make logging of certain events on certain machines more critical is also useful as machines that need to remain secure should be monitored at a more granular level.

Application Log

This log contains various events logged by applications or programs. Many applications will record their errors in this log. It can be useful particularly if the log is on a server that has database server software like SQL Server installed. Examining this log can provide clues that someone has been attempting to compromise the database.

Security Log
The most important things that you will find in the security log are successful and unsuccessful logon attempts. This log also records events related to resource use, such as creating, opening, or deleting files or other objects. Administrators can specify what events are recorded in the security log. Logon auditing can be turned off, but it never should be.

In Windows a security log is the access log. Linux provides separate logs for successful and failed login attempts. By default, Windows does not log both successes and failures, but for security reasons this should be changed.

Although the Windows operating systems do not create audit logs by name, the logs they create are useful in auditing. If you add Share-point, SQL, or other services, then they will often call the application logs they create audit logs and you will want to carefully monitor them for security-related events.

Hardening
The term hardening is usually applied to operating systems. The idea is to "lock down" the operating system as much as is practical. For example, ensure that all unneeded servicesare turned off, all unneeded software is uninstalled, patches are updated, user accounts are checked for security, and so forth. Hardening is a general process of making certain that the operating system itself is as secure as it can be. In fact, it could be said that if you have not hardened the operating system, then any other security measures are going to be far less effective (and possibly completely ineffective!).

Working with Services
Services are programs that run when the operating system boots, and they are often are running in the background without users interacting directly with them. Many services are quite important—even critical. However, a service can provide an attack vector that someone could exploit against your system, so be sure to enable only those services that are absolutely required. Part of operating system hardening is disabling unnecessary services. To display all the services on your Windows computer (any version—from XP to Windows 8 or Windows Server 2012), you first select the Control Panel and then select Administrative Tools, as shown in Figure 2.2.

In Figure 3.3, the Remote Registry service is shown. This service is used to allow technical support personnel to access that system's Registry remotely. The service can be quite useful in some situations, but it can also function as a means for an attacker to get into your system. If you don't need it, turn it off. The issue is not that a given service is "bad"; it is more of an issue of ensuring that you know what services are running on your system and that you make a conscious decision to allow the service to run (or not). Windowsalso provides a brief summary of what the service does and any services that depend on that service. If you don't know what a service does, then you should probably leave it at its default setting.

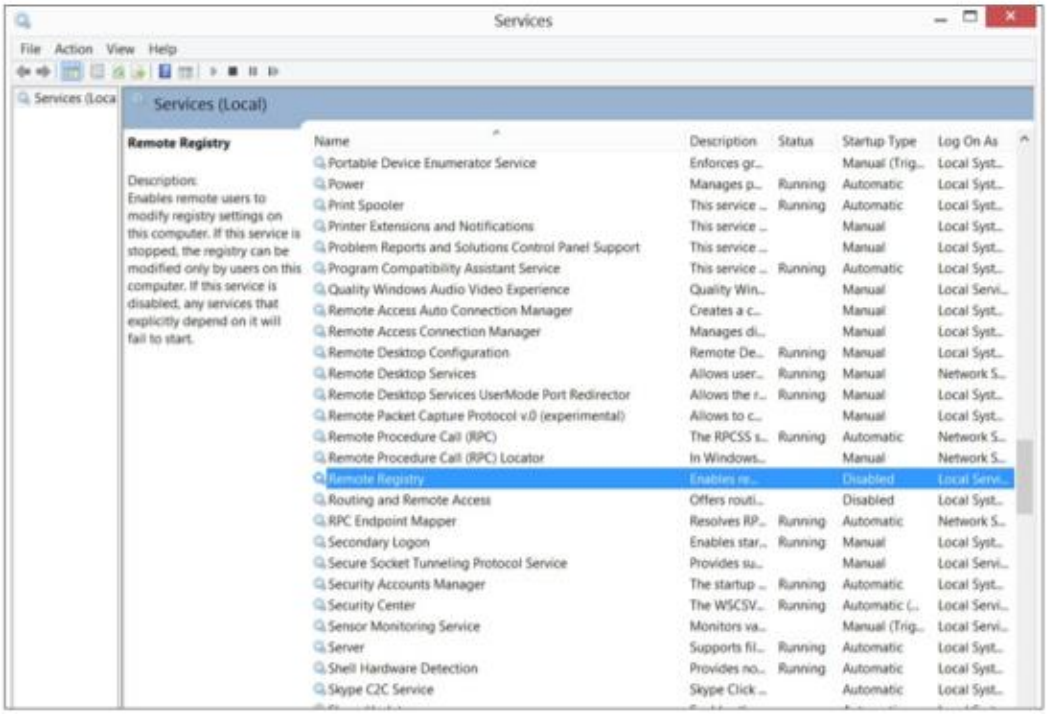
It is critical that you have a good understanding of any service you intend to disable. Some services depend on other services. Turning off one service could render others unusable. Fortunately, the Microsoft Services Console gives you information on dependencies.

As a security administrator, you should regularly check all servers and make certain that only necessary services are running on them. Here are some tips:

File and Print Servers
These are primarily vulnerable to denial-of-service (DoS) and access attacks. DoS attacks can be targeted at specific protocols and overwhelm a port with activity. Make sure that these servers run only the protocols that are needed to support the network.

Networks with PC-Based Systems
In a network that has PC-based systems, make sure that NetBIOS services are disabled on servers or that an effective firewall is in place between the server and the Internet. Many of the popular attacks that are occurring on systems today take place through the NetBIOS services via ports 135, 137, 138, and 139. On Unix systems, make sure that port 111, the Remote Procedure Call (RPC) port, is closed.

FIGURE 3.3 Windows Services



MAC Limiting and Filtering
Limit access to the network to MAC addresses that are known, and filter out those that are not. Even in a home network, you can implement MAC filtering with most routers, and you typically have the option of choosing to allow or deny only those computers with MAC addresses that you list.

If you don't know a workstation's MAC address, use ipconfig /all to find it in the Windows-based world (it is listed as physical address). Use ifconfig or ip a in Unix/Linux.

MAC filtering is not foolproof, and a quick look in a search engine will turn up tools that can be used to change the MAC address and help attackers circumvent this control.

802.1X
The IEEE standard 802.1X defines port-based security for wireless network access control.

The biggest benefit of using 802.1X is that the access points and the switches do not need to do the authentication but instead rely on the authentication server to do the actual work.

Disable Unused Ports
Remember, a port is a connection, like a channel. For example, SMTP uses port 25. For that reason these are sometimes called application ports. All ports not in use should be disabled. Otherwise, they present an open door for an attacker to enter. Essentially, you disable a port by disabling the service and block the port with Windows Firewall (doing one and not the other can result in a single point of failure).

Rogue Machine Detection
On any sizable network it is always possible that someone has added an unauthorized machine. A rogue machine could be an intruder in a neighboring office connecting to your wireless network or an employee adding an unauthorized machine by plugging directly into a network RJ45 jack. Rogue machines pose a serious security risk. Part of your monitoring strategy must be to scan for rogue machines on your network.

Security posture
It is impossible to evaluate your security without having a baseline configuration documented. The baseline must represent a secure state. In other words, it is not simply the state of your network when you decide to start monitoring. It is instead a baseline state you know to be secure. All future security reporting will be relative to this state, so it is rather important.

"Identifies the steps for creation of a baseline configuration, content of the baseline configuration, approval of the initial baseline configuration, maintenance of the baseline configuration (i.e., when it should be updated and by whom), and control of the baseline configuration. If applicable, requirements from higher regulatory bodies are considered and integrated when defining baseline configurations (e.g., requirements from OMB memos, laws such as Health Insurance Portability and Accountability Act (HIPAA), etc.)."

In other words, it is not just the current state of your network, but how it addresses specific compliance issues. Is your network in compliance with HIPAA, PCI, or other relevant regulatory standards? What is the configuration of network security devices (intrusion detection systems, antivirus, and so on)?

It is also a good idea to include network utilization statistics. Being aware of normal traffic flow on your network can be useful when identifying DoS attacks.

Continuous Security Monitoring

Once a baseline security configuration is documented, it is critical to monitor it to see that this baseline is maintained or exceeded. A popular phrase among personal trainers is "that which gets measured gets improved." Well, in network security, "that which gets monitored gets secure."

Continuous monitoring means exactly that: ongoing monitoring. This may involve regular measurements of network traffic levels, routine evaluations for regulatory compliance, and checks of network security device configurations.

Security Audits
Monitoring should take place on several levels. There should be basic, ongoing monitoring that is not labor intensive. Software solutions are available that will accomplish this for you. However, you should also implement scheduled, in-depth checks of security. These are usually called security audits.

A security audit is an integral part of continuous security monitoring. Security audits can be a check of any aspect of your security, including the following:

- Review of security logs
- Review of policies and compliance with policies
- A check of security device configuration
- Review of incident response reports

The scope of the audit and its frequency are determined by the organization. These parameters are determined by security needs and budget. For example, a high school network administrator does not have the budget or the security needs of a defense contractor. Therefore, you could expect the defense contractor to have more frequent and more comprehensive audits. However, every organization needs to have some type of audit policy as a part of continuous monitoring.

Setting a Remediation Policy
The monitoring of your system is very likely to uncover some gaps between the secure baseline that you established and the current state of the network. Those gaps might be quite significant or very minor. For example, you may have a requirement that all RSA cryptography be implemented with 2048-bit keys but discover one service is using 1024-bit keys. This is not a critical gap. Thisdiscrepancy will not render your system wide open to hackers, but it is a gap nonetheless.

Your policies must include a remediation policy. When a gap in the security posture is detected, it should first be classified, and then a remediation plan must be implemented. The specifics of how you classify and respond to a gap will vary from one organization to another. One possible classification system is given here:

Minor
This is a deviation from the security baseline that does not pose any immediate threat to security.

Serious

This is a deviation that could pose an immediate threat, but the threat is either so unlikely or so difficult to exploit as to minimize the danger.

Critical
This is a deviation that poses an immediate threat and that must be addressed as soon as possible.

This is just one possible classification system. An example of a minor threat would be the RSA issue previously mentioned. A serious threat might be the discovery of an obscure vulnerability in a database server that could be exploited but only by someone on the network. A critical threat might be finding out that your web application is vulnerable to SQL injection.

Reporting
Security incidents will occur no matter how well you design your security system. Some of these incidents will be minor, whereas others will be quite serious. Regardless of the severity of the incident, it must be reported. A system must be in place to report all issues.

Alarms
Alarms are indications of an ongoing current problem currently. Think of a siren sounding when someone kicks in the door to a home. These are conditions to which you must respond right now.

Alarm rates can indicate trends that are occurring. Even after you solve the problem, you still need to look for indications that the condition may not be isolated. For example, if your IDS or firewall has an alarm, how is this reported to network security staff? A notification system should be in place that immediately notifies appropriate staff. Once the issue is addressed, those staff members must have a procedure in place to report the specifics of the incident, and how it was addressed, to management.

The point is that your organization needs to have a system for reporting alarms. It cannot be an ad hoc process whereby each individual reports such alarms as they see fit. Incident response cannot occur without some reporting of alarms.

Alerts
Slightly below alarms in terms of security issues are alerts. Alerts are issues to which you need to pay attention but are not about to bring the system down at any moment. (Think of them as storm watches instead of storm warnings.) In Event Viewer, for example, system events are identified either as errors, information, or warnings.
Although errors are the most critical, the others need attention too in order to keep them from eventually becoming errors.

Alerts can also refer to industry alerts. Many antivirus software vendors provide alert services that will email you when a new attack is found or is increasing. Sometimes, other organizations, such as Microsoft, will also send alerts. When a security professional receives such an alert, that information can be communicated both to management and to the staff, as appropriate.

Trends

Trends do not refer to the latest fad in security. Instead they refer to trends in threats. For example, there are more email-based phishing attempts in the last month than in previous months, or waterhole and spear phishing attacks have been increasing recently.

Though not often used in this fashion, the term can also refer to trends in your organizational security profile. Are audits finding an increase in compliance with software policies? Conversely, are you seeing an uptick in the violation of software installation policies?

Detection controls vs. prevention controls
Some security controls are implemented simply to detect potential threats. Others are designed to prevent or at least minimize such threats. For the CompTIA Security+ exam, it is important to know the difference. We will look at security controls here. An intrusion detection system (IDS), as the name implies, is focused on detecting intrusion. One step beyond this, an Intrusion Prevention System (IPS), again as the name implies, is focused on preventing an intrusion from occurring. There are various levels of both IDS and IPS as they can be based on a host (H-IDS, for example) or a network (N-IDS).

Not all approaches are so clear-cut as to include the term "detection" or "prevention" in the title, and many tools fall between the two. One such tool is a honeypot. A honeypot is a computer that has been designated as a target for computer attacks. The best way to visualize a honeypot is to think of Winnie the Pooh and the multiple times the character has become stuck while trying to get the honey out of the jugs in which it is stored. By get- ting stuck, he has incapacitated himself and become an easy target for anyone trying to find him.

The purpose of a honeypot is to allow itself to succumb to an attack. During the process of "dying," the system can be used to gain information about how the attack developed and what methods were used to institute the attack. The benefit of a honeypot system is that it draws attackers away from a higher-value system or allows administrators to gain intelligence about an attack strategy.

Honeypots aren't normally secured or locked down. If they come straight out of the box with an operating system and applications software, they may be configured as is. Elaborate honeypot systems can contain information and software that might entice an attacker to probe deeper and take over the system. If not configured properly, a honeypot system can be used to launch attacks against other systems. There are several initiatives in the area of honeypot technology. One of the more interesting involves the Honeynet Project, which created a synthetic network that can be run on a single computer system and is attached to a network using a normal network interface card (NIC). The system looks like an entire corporate network, complete with applications and data, all of which are fake. As part of the Honeynet Project, the network was routinely scanned, worms were inserted, and attempts were made to contact other systems to infest them—all over the course of a three-day period. At the end of day three, no fewer than three worms had infected the system. This infestation happened without any advertising by the Honeynet Project.

Enticement
Enticement is the process of luring someone into your plan or trap. You might accomplish this by advertising that you have free software, or you might brag that no one can break into your machine. If you invite people to try, you're enticing them to do some- thing that you want them to do.

Entrapment
Entrapment is the process in which a law enforcement officer or a government agent encourages or induces a person to commit a crime when the potential criminal expresses a desire not to go ahead. Entrapment is a valid legal defense in a criminal prosecution.

Although enticement is legally acceptable in the United States, entrapment is not. Your legal liabilities are probably small in either case, but you should seek legal advice before you implement a honeypot on your network. You may also want to contact law enforcement or the prosecutor's office if you want to pursue legal action against attackers.

IDS vs. IPS
This type of system is an IDS that reacts to the intrusion that has been detected, most often by blocking communication from the offending IP address. The problem with this approach is the issue of false positives. No system is perfect—at some point you will have a situation where network activity is anomalous and the IDS indicates an intrusion, butin reality it is not an intrusion. For example, if the IDS is set up to react to traffic outside normal bounds, excessive traffic from a given system could indicate an attack. However, it could also indicate an unusually high workload.

Camera vs. guard
The camera versus guard debate is an old one. You must decide what is best for your own environment. The benefit of a camera (also known as closed-circuit television, or CCTV) is that it is always running and can record everything it sees, creating evidence that can be admissible in court if necessary. On the other hand, it is stationary, lacks any sort of intelligence, is possible to avoid, and needs someone to monitor the feed or review the tape to be effective, which many times does not happen until a problem has been discovered.

The benefit of a guard is that the person can move about, apply intelligence to situations, and collect evidence. The guard, however, is not always recording, can be avoided, and has more downtime.

Section 3.7- Given a scenario, use appropriate tools and techniques to discover security threats and vulnerabilities.

Interpret results of security assessment tools
Similar to packet sniffing, port scanning and other "security tools", vulnerability scanning can help you to secure your own network or it can be used by the bad guys to identify weaknesses in your system to mount an attack against. The idea is for you to use these tools to identify and fix these weaknesses before the bad guys use them against you.

The goal of running a vulnerability scanner is to identify devices on your network that are open to known vulnerabilities. Different scanners accomplish this goal through different means. Some work better than others.

Some may look for signs such as registry entries in Microsoft Windows operating systems to identify that a specific patch or update has been implemented. Others, in particular Nessus,

actually attempt to exploit the vulnerability on each target device rather than relying on registry information.

Kevin Novak did a review of commercial vulnerability scanners for Network Computing Magazine in June of 2003. While one of the products, Tenable Lightning, was reviewed as a front-end for Nessus, Nessus itself was not tested directly against the commercial products. Click here for the complete details and results of the review: VA Scanners Pinpoint Your Weak Spots.

One issue with vulnerability scanners is their impact on the devices they are scanning. On the one hand you want the scan to be able to be performed in the background without affecting the device. On the other hand, you want to be sure that the scan is thorough. Often, in the interest of being thorough and depending on how the scanner gathers its information or verifies that the device is vulnerable, the scan can be intrusive and cause adverse affects and even system crashes on the device being scanned.

There are a number of highly rated commercial vulnerability scanning packages including Found stone Professional, eEye Retina and SAINT. These products also carry a fairly hefty price tag. It is easy to justify the expense given the added network security and peace of mind, but many companies simply don't have the sort of budget needed for these products.

While not a true vulnerability scanner, companies that rely primarily on Microsoft Windows products can use the freely available Microsoft Baseline Security Analyzer (MBSA). MBSA will scan your system and identify if there are any patches missing for products such as the Windows operating systems, Internet Information Server (IIS), SQL Server, Exchange Server, Internet Explorer, Windows Media Player and Microsoft Office products. It has had some issues in the past and there are occasionally errors with the results of MBSA- but the tool is free and is generally helpful for ensuring that these products and applications are patched against known vulnerabilities. MBSA will also identify and alert you to missing or weak passwords and other common security issues.

Nessus is an open-source product and is also freely available. While there is a Windows graphical front-end available, the core Nessus product requires Linux / Unix to run. The up side to that is that Linux can be obtained for free and many versions of Linux have relatively low system requirements so it would not be too difficult to take an old PC and set it up as a Linux server. For administrators used to operating in the Microsoft world there will be a learning curve to get used to Linux conventions and get the Nessus product installed.

Tools

There are various tools that you can use to scan the system and find vulnerabilities. We are listing a few below.

Protocol analyzer

A protocol analyzer (also known as a packet sniffer, network analyzer, or network sniffer) is a piece of software or an integrated software/hardware system that can capture and decode network traffic. Protocol analyzers have been popular with system administrators and security professionals for decades because they are such versatile and useful tools for a network environment. From a security perspective, protocol analyzers can be used for a number of activities, such as the following:

- Detecting intrusions or undesirable traffic (IDS/IPS must have some type of capture and decode ability to be able to look for suspicious/malicious traffic)
- Capturing traffic during incident response or incident handling
- Looking for evidence of botnets, Trojans, and infected systems
- Looking for unusual traffic or traffic exceeding certain thresholds
- Testing encryption between systems or applications

From a network administration perspective, protocol analyzers can be used for activities such as these:

- Analyzing network problems
- Detecting misconfigured applications or misbehaving applications
- Gathering and reporting network usage and traffic statistics
- Debugging client/server communications

Regardless of the intended use, a protocol analyzer must be able to see network traffic in order to capture and decode it. A software-based protocol analyzer must be able to place the NIC it is going to use to monitor network traffic in promiscuous mode (sometimes called promisc mode). Promiscuous mode tells the NIC to process every network packet it sees regardless of the intended destination. Normally, a NIC will process only broadcast packets (that are going to everyone on that subnet) and packets with the NIC's Media Access Control (MAC) address as the destination address inside the packet. As a sniffer, the analyzer must process every packet crossing the wire, so the ability to place a NIC into promiscuous mode is critical.

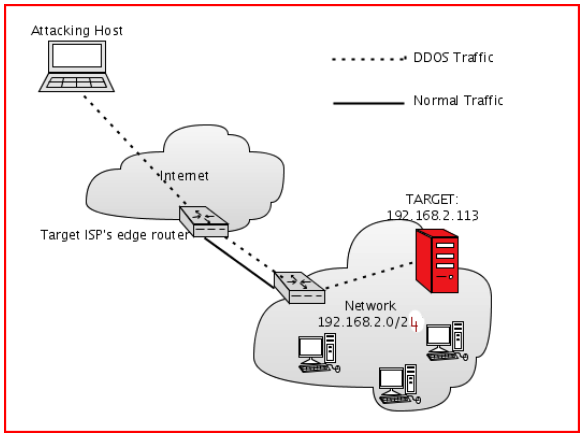
Honeypots and Honeynets

As is often the case, one of the best tools for information security personnel has always been knowledge. To secure and defend a network and the information systems on that network properly, security personnel need to know what they are up against. What types of attacks are being used? What tools and techniques are popular at the moment? How effective is a certain technique? What sort of impact will this tool have on my network? Often this sort of information is passed through white papers, conferences, mailing lists, or even word of mouth. In some cases, the tool developers themselves provide much of the information in the interest of promoting better security for everyone. Information is also gathered through examination and forensic analysis, often after a major incident has already occurred and information systems are already damaged.

One of the most effective techniques for collecting this type of information is to observe activity first-hand—watching an attacker as she probes, navigates, and exploits his way through a network. To accomplish this without exposing critical information systems, security researchers often use something called a honeypot.

A honeypot, sometimes called a digital sandbox, is an artificial environment where attackers can be contained and observed without putting real systems at risk. A good honeypot appears to an attacker to be a real network consisting of application servers, user systems, network traffic, and so on, but in most cases it's actually made up of one or a few systems running specialized software to simulate the user and network traffic common to most targeted

networks. The figure below illustrates a simple honeypot layout in which a single system is placed on the network to deliberately attract attention from potential attackers.



There are many honeypots in use, specializing in everything from wireless to denial-of-service attacks; most are run by research, government, or law enforcement organizations. Why aren't more businesses running honeypots? Quite simply, the time and cost are prohibitive. Honeypots take a lot of time and effort to manage and maintain and even more effort to sort, analyze, and classify the traffic the honeypot collects. Unless they are developing security tools, most companies focus their limited security efforts on preventing attacks, and in many cases, companies aren't even thatconcerned with detecting attacks as long as the attacks are blocked, are unsuccessful, and don't affect business operations. Even though honeypots can serve as a valuable resource by luring attackers away from production systems and allowing defenders to identify and thwart potential attackers before they cause any serious damage, the costs and efforts involved deter many companies from using honeypots.

Port scanner

Port Scanning is one of the most popular reconnaissance techniques attackers use to discover services they can break into. All machines connected to a Local Area Network (LAN) or Internet run many services that listen at well-known and not so well known ports. A port scan helps the attacker find which ports are available (i.e., what service might be listing to a port). Essentially, a port scan consists of sending a message to each port, one at a time. The kind of response received indicates whether the port is used and can therefore be probed further for weakness.

Port Scan – Port Numbers

As you know, public IP addresses are controlled by worldwide registrars, and are unique globally. Port numbers are not so controlled, but over the decades certain ports have become standard for certain services. The port numbers are unique only within a computer system. Port numbers are 16-bit unsigned numbers.The port numbers are divided into three ranges:

- Well Known Ports (0 – 1023)
- Registered Ports (1024 – 49151)
- Dynamic and/or Private Ports (49152 – 65535)
- Assessment technique
- Baseline reporting

- Code review
- Determine attack surface
- Architecture
- Design reviews

Port Scanning Basic Techniques
The simplest port scan tries (i.e., sends a carefully constructed packet with a chosen destination port number) each of the ports from 0 to 65535 on the victim to see which ones are open.

TCP connect ():- The connect() system call provided by an OS is used to open a connection to every interesting port on the machine. If the port is listening, connect() will succeed, otherwise the port isn't reachable.

Strobe -A strobe does a narrower scan; only looking for those services the attacker knows how to exploit. The name comes from one of the original TCP scanning programs, though now virtually all-scanning tools include this feature.

The ident protocol allows for the disclosure of the username of the owner of any process connected via TCP, even if that process didn't initiate the connection. So, e.g., one can connect to port 80 and then use identd to find out whether the HTTP server is running as root.

Passive vs. active tools

Banner grabbing
As the name implies, banner grabbing looks at the banner, or header information messages sent with data to find out about the system(s). Banners often identify the host, the operating system running on it, and other information that can be useful if you are going to attempt to later breach the security of it. Banners can be snagged with Telnet as well as tools like netcat or Nmap.

Risk calculations
For purposes of risk assessment, both in the real world and for the exam, you should familiarize yourself with a number of terms to determine the impact an event could have:

- ALE is the annual loss expectancy value. This is a monetary measure of how much loss you could expect in a year.

- SLE is another monetary value, and it represents how much you expect to lose at any one time: the single loss expectancy. SLE can be divided into two components:

    - AV (asset value)
    - EF (exposure factor)

- ARO is the likelihood, often drawn from historical data, of an event occurring within a year: the annualized rate of occurrence.

When you compute risk assessment, remember this formula:

SLE × ARO = ALE

As an example, if you can reasonably expect that every SLE, which is equal to asset value (AV) times exposure factor (EF), will be the equivalent of $1,000 and that there will be seven such occurrences a year (ARO), then the ALE is $7,000. Conversely, if there is only a 10 percent chance of an event occurring within a year time period (ARO = 0.1), then the ALE drops to $100.

Likelihood
The meaning of the word likelihood is usually self-explanatory; however, there are actual values that can be assigned to likelihood. The National Institute of Standards and Technology (NIST) recommends viewing likelihood as a score representing the possibility of threat initiation. In this way, it can be expressed either in qualitative or quantitative terms.]

Assessment types
From the standpoint of measuring security and vulnerability in the network, you need to focus on three things:

Risk
What is the actual danger under consideration? This is the likelihood of an attack being successful.

Threat
What are the likely dangers associated with the risk? What are the means and source of the potential attack? This needs to be weighed against the likelihood of an attack, which the NIST defines as "a weighted risk factor based on an analysis of the probability that a given threat is capable of exploiting a given vulnerability."

Vulnerability
Where is the system weak? Identify the flaws, holes, areas of exposure, and perils.

Baseline Reporting
The term baseline reporting became popular with legislation such as Sarbanes–Oxley, which requires IT to provide internal controls that reduce the risk of unauthorized transactions. As the name implies, baseline reporting checks to make sure that things are operating status quo, and change detection is used to alert administrators when modificationsare made. A changes-from-baseline report can be run to pinpoint security rule breaches quickly.

This is often combined with gap analysis to measure the controls at a particular company against industry standards. One popular tool for baseline reporting is CA Policy and Configuration Manager (www.ca.com).

Code Review
The purpose of code review is to look at all custom written code for holes that may exist. The review needs also to examine changes that the code—most likely in the form of a finished application—may make: configuration files, libraries, and the like. During this examination, look for threats such as opportunities for injection to occur (SQL, LDAP, code, and so on), cross-site request forgery, and authentication.

Code review is often conducted as a part of gray box testing. Looking at source code can often be one of the easiest ways to find weaknesses within the application. Simply reading the code is known as manual assessment, whereas using tools to scan the code is known as automated assessment.

Determine Attack Surface
The attack surface of an application is the area of that application that is available to users—those who are authenticated and, more importantly, those who are not. As such, it can include the services, protocols, interfaces, and code. The smaller the attack surface, the less visible the application is to attack; the larger the attack surface, the more likely it is to become a target. The goal of attack surface reduction (ASR) is to minimize the possibility of exploitation by reducing the amount of code and limiting potential damage. The potential damage can be limited by turning off unnecessary functions, reducing privileges, limiting entry points, and adding authentication requirements.

Section 3.8 Explain the proper use of penetration testing versus vulnerability scanning.

Penetration testing
It is becoming more common for companies to hire penetration testers to test their system's defenses. Essentially, a penetration tester will use the same techniques a hacker would use to find any flaws in your system's security.

Hacking and penetration testing are areas that seem quite exciting to many people. Unfortunately, this has led to a number of unqualified (or at least underqualified) people calling themselves penetration testers. It is imperative when hiring a penetration tester that you ensure the person in question has the requisite skill set. Check their references and verify their training and skills. It is also important to do a thorough background check on the person in question, as you are giving this person permission to try hacking techniques on your network. You will want to be certain that they conduct themselves in an ethical manner.

Vulnerability scanning
Many security experts view vulnerability scanning as separate from penetration testing. However, it should be either part of the penetration test or done alongside it. Vulnerability scanning allows you to identify specific vulnerabilities in your network, and most penetration testers will start with this procedure so that they can identify likely targets to attack. A penetration test is essentially an attempt to exploit these vulnerabilities.

Once you have identified the vulnerabilities, it is time to attempt to exploit them. Of course the most egregious vulnerability is any aspect of your system where vulnerability scanning reveals a lack of security controls. Some of the more common vulnerabilities involve misconfiguration.

Passively Testing Security Controls
The vulnerability scanner can test the security controls without doing any actual harm. It looks only for the openings that are there and reports them back to you. As such, its testing is considered to be passive as opposed to active.

Interpreting Results
Most of the vulnerability scanning programs, and the commercial ones in particular, interpret the results of their findings and deliver a report that can be shared with management.

Identifying Vulnerability
Just knowing that the port is open means little unless you can associate it with the vulnerability tied to it. For example, port 23 being open is a problem since it is commonly associated with Telnet.

Identifying Lack of Security Controls
Looking for weaknesses in security controls is well and good, but just as important is identifying areas where there are no controls in place. You want to know not just what is weak, but also what is missing altogether.

Identifying Common Misconfigurations
All too often, problems are introduced when perfectly good applications and services are improperly configured. Those misconfigurations can allow more users than should be permitted to access an application, cause the application to crash, or introduce any of a number of other security concerns.

Credentialed vs. non-credentialed
Vulnerability scanning can be done either in a credentialed or non-credentialed manner. The difference is that a credentialed vulnerability scan uses actual network credentials toconnect to systems and scan for vulnerabilities. Tenable Security, the creators of the Nessus vulnerability scanner, have this to say about credentialed scanning:
This type of scan has several benefits:

- **Not disrupting operations or consuming too many resources** Because the scan is performed with credentials, operations are executed on the host itself rather than across the network. Running commands on the host, then sending the results of those commands back to the Nessus server do everything from operating system identification to port scanning. This allows Nessus to consume far less system and network resources than per- forming a traditional network scan that probes ports and services remotely.
- **Definitive list of missing patches** Rather than probe a service remotely and attempt to find vulnerability, Nessus will query the local host to see if a patch for a given vulnerability has been applied. This type of query is far more accurate (and safer) than running a remote check.
- **Client-side software vulnerabilities are uncovered** By looking at the soft- ware installed and its version, Nessus will find client-side software vulnerabilities that are otherwise missed in a traditional network-based audit.
- **Several other "vulnerabilities"** Nessus can read password policies, obtain a list of USB devices, check anti-virus software configurations and even enumerate Bluetooth devices attached to scanned hosts.

Whether you use credentialed or non-credentialed vulnerability scanning be prepared for false positives. A false positive occurs when the scan mistakenly identifies something as a vulnerability when it is not. No software program is perfect, and this means that any vulnerability scanner will yield some occasional false positives.

False positive
False positives are events that aren't really incidents. Event flagging is often based on established rules of acceptance (deviations from which are known as anomalies) and things such as attack signatures. If the rules aren't set up properly, normal traffic may set off an analyzer and generate an event. You don't want to declare an emergency unless you're sure that you have one. The opposite of a false positive is a false negative. With a false negative, you are not alerted to a situation when you should be alerted. In this case, you miss something crucial and it slips right by.

Black Box
The tester has absolutely no knowledge of the system and is functioning in the same manner as an outside attacker.

White Box
The tester has significant knowledge of your system. This simulates an attack from an insider—a rogue employee.

Gray Box
This is a middle ground between the first two types of testing. In gray box testing, the tester has some limited knowledge of the target system.

In addition to classifying a penetration test based on the amount of informationgiven to the tester, it is also possible to classify the test as intrusive versus nonintrusive. Nonintrusive tests involve passively testing security controls—performing vulnerability scans, probing for weaknesses, but not exploiting them. Intrusive tests involve actually trying to break into the network. In the strictest sense, passive tests are really just vulnerability scans and not penetration tests, while active tests provide more meaningful results. With active tests, it is possible that they may disrupt business operations in the same way as a real attack.

Topic 4.0- Application, Data and Host Security

Section 4.1-Explain the importance of application security controls and techniques.

Fuzzing
Fuzz testing or Fuzzing is a Black Box software testing technique, which basically consists in finding implementation bugs using malformed/semi-malformed data injection in an

automated fashion.

Let's consider an integer in a program, which stores the result of a user's choice between 3 questions. When the user picks one, the choice will be 0, 1 or 2. Which makes three practical cases. But what if we transmit 3, or 255? We can, because integers are stored a static size variable. If the default switch case hasn't been implemented securely, the program may crash and lead to "classical" security issues: (un) exploitable buffer overflows and DoS.

Fuzzing is the art of automatic bug finding, and its role is to find software implementation faults, and identify them if possible.

Security coding concepts

Error and exception handling
Correct handling of errors and exceptions is important for correct broker operation. You should be aware of this, and understand how and when your user-defined extension needs to handle errors and exceptions.

The message broker generates C++ exceptions to handle error conditions. These exceptions are caught in the relevant software layers in the broker and handled accordingly. However, programs written in C cannot catch C++ exceptions, and any exceptions thrown, by default, bypass any C user-defined extension code and be caught in a higher layer of the message broker.

Utility functions, by convention, normally use the return value to pass back requested data; for example, the address or handle of a broker object. The return value sometimes indicates that a failure has occurred. For example, if the address or handle of a broker object could not be retrieved, then zero (CCI_NULL_ADDR) is returned. Additionally, the reason for an error condition is stored in the return code output parameter, which is, by convention, part of the function prototype of all utility functions. If the utility function completed successfully and returnCode was not null, returnCode contains CCI_SUCCESS. Otherwise, it contains one of the return codes described below. The value of returnCode can always be tested safely to determine whether a utility function was successful.

Types of exception and broker behavior

The broker generates a set of exceptions that can be passed to a user-defined extension. These exceptions can also be generated by a user-defined extension when an error condition is encountered. The exception classes are:

Fatal
Fatal exceptions are generated when a condition occurs that prevents the broker process from continuing execution safely, or where it is broker policy to terminate the process. Examples of fatal exceptions are a failure to acquire a critical system resource, or an internally-caught severe software error. The broker process terminates following the throwing of a fatal exception.

Recoverable
These are generated for errors, which, although not terminal in nature, mean that the processing of the current message flow has to be ended. Examples of recoverable exceptions

are invalid data in the content of a message, or a failure to write a message to an output node. When a recoverable exception is thrown, the processing of the current message is aborted on that thread, but the thread recommences execution at its input node.

Configuration
Configuration exceptions are generated when a configuration request fails. This can be because of an error in the format of the configuration request, or an error in the data. When a configuration exception is thrown, the request is rejected and an error response message is returned.

Parser
These are generated by message parsers for errors,which prevent the parsing of the message content or creating a bit stream. The broker treats a parser exception as a recoverable exception.

Conversion
These are generated by the broker character conversion functions if invalid data is found when trying to convert to another data type. The broker treats a conversion exception as a recoverable exception.

User
These are generated when a Throw node throws a user-defined exception.

Database
These are generated when a database management system reports an error during broker operation. A database exception is treated as a recoverable exception by the broker.

Input validation
Web applications are notorious for taking practically any type of input, assuming that it's valid, and processing it further. Not validating input is one of the greatest mistakes that Web-application developers can make. This can lead to system crashes, malicious database manipulation, and even database corruption.

Input attacks
Several attacks can be run against a Web application that insert malformed data — often, too much at once — which can confuse, crash, or make the Web application divulge too much information to the attacker.

Cross-site scripting prevention

Using a client-side scripting language, it is possible for an attacker to trick a user who visits the site into having code execute locally. When this is done, it is known as cross-site scripting (XSS). Let’s look at an example. UserA gets a message telling him that he needs to make changes to his XYZ account, but the link in the message is not really to the XYZ site (aphishing ploy). When he visits the site, a script routine begins to run on his machine with his permissions and can begin doing such things as running malevolent routines to send, delete, or alter data.

Cross-Site Request Forgery—also known as XSRF, session riding, and one-click attack—involves unauthorized commands coming from a trusted user to the website. This is often done without the user's knowledge, and it employs some type of social networking to pull it off.

For example, assume that Evan and Spencer are chatting through Facebook. Spencer sends Evan a link to what he purports is a funny video that will crack him up. Evan clicks the link, but it actually brings up Evan's bank account information in another browser tab, takes a screenshot of it, closes the tab, and sends the information to Spencer. The reason the attack is possible is because Evan is a trusted user with his own bank. In order for itto work, Evan would need to have recently accessed that bank's website and have a cookie that had yet to expire.

The best protection against cross-site scripting is to disable the running of scripts (and browser profiles).

Application Configuration Baseline (proper settings)
Base lining always involves comparing performance to a metric. That metric is a historical measurement that you can point to and identify as coming before a configuration change, before the site became busy, before you added new services, and so on. Baselining can be done with any metric, such as network performance or CPU usage, as well as with applications.

Application Hardening
A good way to begin securing a network is to make sure that every system in the network is up-to-date and to verify that only those protocols you need are enabled. Unfortunately, these steps aren't enough. Your servers and workstations also run applications and services. Server services (especially web, email, and media servers) are particularly vulnerable to exploitation and attack. These applications must also be hardened to make them as difficult as possible to exploit.

Application Patch Management
Just as you need to keep operating system patches current, as they often fix security problems discovered within the OS, you need to do the same with application patches. Once an exploit in an application becomes known, an attacker can take advantage of it to enter or harm a system. Most vendors post patches on a regular basis, and you should routinely scan for any available ones.

A large number of attacks today are targeted at client systems for the simple reason that clients do not always manage application patching well. When you couple that with the fact that most clients have many applications running, the odds of being able to find a weakness to exploit are increased dramatically.

Server-side vs. Client-side validation
Some attacks, such as SQL injection, depend entirely on unfiltered input being sent through a web application. OWASP recommends that all data input by a user be validated before it is processed. There are two primary ways to do input validation: client-side validation and server-side validation.

Client-side validation usually works by taking the input that a user enters into a text field and,

on the client side, checking for invalid characters or input. This process can be as simple as verifying that the input does not exceed the required length, or it can be a complete check for SQL injection characters. In either case, the validation is accomplished on the client web page before any data is sent to the server.

Server-side validation involves validating data after the server has received it. This process can include checking business logic to see if the data sent conforms to expected parameters. It is unusual to have just server-side validation. You may have systems with only client-side validation, but server-side validation is normally done in conjunction with client-side validation.

Section 4.2- Summarize mobile security concepts and technologies.

Device Security
Mobile devices, such as laptops, tablet computers, and smartphones, provide security challenges above those of desktop workstations, servers, and such in that they leave the office and this increases the odds of their theft. In 2010, AvMed Health Plans, a Florida-based company, had two laptop computers stolen. Together, over one million personal customer records were on those computers, and this is but one of many similar stories that happen on a regular basis.

At a bare minimum, the following security measures should be in place on mobile devices:

Screen Lock The display should be configured to time out after a short period of inactivity and the screen locked with a password. To be able to access the system again, the user must provide the password. After a certain number of attempts, the user should not be allowed to attempt any additional logons; this is called *lockout*.

Strong Password Passwords are always important, but even more so when you consider that the device could be stolen and in the possession of someone who has unlimited access and time to try various values.

Device Encryption Data should be encrypted on the device so that if it does fall into the wrong hands, it cannot be accessed in a usable form without the correct passwords.

Remote Wipe/Sanitation Many programs, such as Microsoft Exchange Server 2010 or Google Apps, allow you to send a command to a phone that will remotely clear the data on that phone. This process is known as a *remote wipe*, and it is intended to be used if the phone is stolen or going to another user.

Voice Encryption Voice encryption can be used with mobile phones and similar devices to encrypt transmissions. This is intended to keep the conversation secure and works by adding cryptography to the digitized conversation.

GPS Tracking Should a device be stolen, GPS (Global Positioning System) tracking can be used to identify its location and allow authorities to find it. Note that removable storage can circumvent GPS. For example, if a device has GPS tracking but it also has removable storage, a thief can simply remove the data they want and leave the device.

Application Control Application control is primarily concerned with controlling what applications are installed on the mobile device. Most viruses that are found on Android phones stem from bad applications being installed. Related to application control is disabling unused services. If you do not need a service, turn it off.

Storage Segmentation By segmenting a mobile device's storage you can keep work data separate from personal or operating system data. You can even implement whole device encryption or just encrypt the confidential data.

Asset Tracking You must have a method of asset tracking. It can be as simple as a serial number etched in the device or as complex as a GPS locator. Related to this is inventory control. A complete and accurate list of all devices is an integral part of mobile device management.

Device Access Control Device access control, in this context, refers to controlling who in the organization has a mobile device. Not every employee should have one. Limiting access to such devices reduces risk.

Application Security

There are a number of issues to be cognizant of when it t comes to application security.

Key Management Key management is an area of importance that is continuing to grow as PKI services increase and expand to mobile.

Credential Management Credentials allow usernames and passwords to be stored in one location and then used to access websites and other computers. Newer versions of Windows include Credential Manager (beneath the Control Panel) to simplify management.
Authentication Authentication has always been an issue, but now that mobile is expanding and the need for authentication with applications associate with it has grown, the issue has become even more important. Users should be taught best practices and should never configure any application to automatically log them in.

Geo-Tagging Geo-tagging (usually written as GeoTagging) allows GPS coordinates (latitude, longitude, etc.) to accompany a file such as an image. This is a common practice with pictures taken using a smartphone or digital camera. While it can be useful if you are trying to remember details of a family vacation, it can also raise security concerns in a business environment. As an example, suppose a picture is taken of your server room and posted—the geotagged information accompanying it would allow anyone to know the precise location of your server room and that could easily be something you would rather protect.

Encryption Encryptionopens up a lot of possibilities for increasing security, but brings it with issues that company policies should be created to address: for example, what is the procedure when a user forgets their password to an application/data?

Application White-Listing White lists are lists of those items that are allowed (as opposed to a black list—things that are prohibited). A white list of applications should exist to identify what applications are approved and accepted on your network.

Transitive Trust/Authentication Anytime one entity accepts a user without requiring additional authentication on the behalf of another entity, the possibility is introduced for problems to occur. As much of a pain as it is for users, the more stops that you have requiring them to authenticate before passing through, the safer you make you make your environment.

BYOD Concerns
BYOD (Bring Your Own Device) refers to employees bringing their personal devices into the corporate network environment. This is a common issue in the modern workplace, and it can pose substantial security risks.

The first risk involves those devices connecting to the company network. If an employee has personal smartphone, for example, and they bring it to work and connect it to the company's Wi-Fi network, then any virus, spyware, or other malware that may have infected that phone can spread to the company network. One way to address this is to have a second Wi-Fi network—not connected to the main corporate network, but simply a guest network—and only allow personal devices to connect to that Wi-Fi and not to the main network.

Another risk involves compromising confidential data. Modern mobile devices are complex computer systems. An employee could use a cell phone to photograph sensitive documents, record conversations, and acquire a great deal of sensitive data. Some Department of Defense contractors do not allow phones in certain sensitive areas of their buildings. This may be more restrictive than at most civilian companies, but at least you should be aware of this potential issue and have a policy to address it. That policy could be as simple as allemployees agreeing that if they bring a mobile device onto company property, it is subject to random search.

Data ownership becomes an issue with BYOD. If the device is personally owned but used for company business, who owns the data on the device? The company or the individual? Related to that is the issue of support ownership. Is the individual responsible for support or the company? Patch management is closely related to support ownership. Who will be responsible for ensuring the personal device has patches updated? Antivirus management is another related issue. What antivirus software will be used? How will it be updated? These are all important questions that will need to be answered.

Adherence to corporate policies is an obvious issue. If individuals own their own devices, which they have purchased with their own funds, ensuring the user and the device adheres to corporate policies will be a challenge. Related to that issue are legal concerns. When a device is owned by the individual but used for company business, a number of legal issues arise. As just one example, what if the device is used to send spam? Is the company responsible? Another example would involve the employee leaving the company. How does the organization verify the device does not have any proprietary data on it? Forensics is another legal issue. If there is, for example, litigation against the company, usually computer records is subpoenaed, but the data that might reside on a personal device is a legal gray area.

Then there are purely technical concerns. Architecture and infrastructure considerations are critical. Will the personal device be compatible with the organizational infrastructure? On-board cameras and video also pose a challenge. Some organizations forbid the use of cameras within the company, or at least within secure areas. And finally there is the issue of acceptable use policies. Companies generally have acceptable use policies regarding how

computers can be used within the organization. How will that be implemented with devices that don't belong to the company?

Some organizations simply opt to forbid such devices, but in our modern world of ubiquitous devices, that approach may not be feasible in your organization.

Section 4.3 Given a scenario, select the appropriate solution to establish host security.

Operating system security and settings
The ability to run the administrative interfaces within the operating system, and the applications associated with them, is often the difference between a standard user and an administrative user. The person running the administrative interfaces can make configuration changesto the system(s) and modify settings in ways that can have wide-ranging consequences. For example, a user who is able to gain access to the administrative tools could delete other users, set their own ID equal to the root user, change passwords, or delete key files.

To protect against this, access to management and administrative interfaces should be restricted to only those administrators who need it. Not only should you protect server utilities, but also you should also even go so far as to remove users' access to workstation utilities such as regedit and regedit32 that have administrative depth.
The System And Security applet beneath the Control Panel (known just as Security in operating systems earlier than Windows 7) is the main interface for security features inWindows operating systems. From here, you can configure Windows Firewall; automatic scans of your computer, and Windows Defender.

One of the best tools to use when looking for possible illicit activity on a workstation is Performance Monitor (known as System Monitor in early versions of Windows). This utility can be used to examine activity on any counter. Excessive processor usage is one counter worth paying attention to if you suspect the workstation is affected or being illegitimately accessed.

It is important that you use password protection to protect the management functionality and consoles on a workstation or server. Just because users are authorized to use that machine does not mean they should be authorized to access all management functions.

OS Hardening
The space of hardening an OS is vast. It includes new ideas yet to be implemented in widely used OS (Windows, Linux). It includes re-designing and re-implementing existing ideas such as "change-roots". It includes analyzing the source code of an OS extremely carefully by experts as well as via software tools based on mathematical proof techniques.

Least Privilege
On Unix/Linux systems, the user called root or superuser, user id 0, can bypass all security restrictions. Windows systems have the "System" and "Administrator" accounts. The super

user privilege should be split into many smaller privileges. E.g., a backup process should be able to read any file, but it should not be able to shut down the system, modify files, or send network packets. Processes, not users, should be given privileges. The backup program should be permitted to read all files, even though the administrator user who invokes the program should not be allowed such access. The backup process should not pass down its privileges to processes that it starts. The use of such finely divided abilities instead of sweeping powerful mechanisms is called the least privilege principle.

The traditional Unix model allows for access control of only files. So, a number of resources become "files": processes, hard disks, network interfaces, etc. In order to apply the principle of least privilege, we also need to divide the resources into finer units (often called objects, but unrelated to OOP). The users and processes are called subjects.

Capabilities
A capability is a word used with different meanings in the context of OS design. In OS research literature, processes hold tokens, called capabilities, denoting resources to be accessed and what can be done with them. Capabilities can be explicitly passed among processes. Linux, Windows, ... are not capability based in this sense. This usage of the word is unrelated to "POSIX capabilities" which are implemented in Linux.

Mandatory Access Control
Newer OS designs add additional security attributes, called sensitivity labels (SLs), to files, processes, network interfaces, host addresses, and other resources based on how we wish to compartmentalize them. Access control using SLs is called mandatory access control (MAC). It is called mandatory because no inheritance of privileges is assumed. E.g., MAC can be applied at the network interface level. Incoming packets are assigned SLs, based on the source IP address or the network interface. Outgoing packets will have the label of the process that created them. A filtering rule can be formulated so that an incoming or outgoing packet can be dropped if the SL does not satisfy some conditions. MAC, in its academic form, is not about access control but about information flow.

There are Linux kernel modifications that accomplish the following. Secure policy enforcement; Supports read, write, append, execute, view, and read-only ptrace object permissions; Supports hide, protect, and override subject flags; Supports the PaX flags; Shared memory protection feature; Integrated local attack response on all alerts; Subject flag that ensures a process can never execute trojaned code; Intelligent learning mode that produces least-privilege ACLs with no configuration; Full-featured fine-grained auditing; Resource ACLs; Socket ACLs; File/process ACLs; Capabilities; Protection against exploit bruteforcing; /proc/pid filedescriptor/memory protection; ACLs can be placed on non-existent files/processes; ACL regeneration on subjects and objects; Administrative mode to use for regular sysadmin tasks; ACL system is resealed up admin logout; Globbing support on ACL objects; Configurable log suppression; Configurable process accounting; Human-readable configuration; Not filesystem dependent; Not architecture dependent; Scales well: supports as many ACLs as memory can handle; No runtime memory allocation; SMP safe; Include directive for specifying additional ACLs; Enable, disable, reload capabilities; Userspace option to test permissions on an ACL; Option to hide kernel processes;

Role-Based Access Control

We can divide the permissions/privileges based on the function ("role") they have, such as backup, file system integrity check, filtration of incoming packets. Each user is permitted a collection of roles. RBAC can implement MAC. There is a considerable amount of discrete mathematics developed for RBAC and MAC.

Read-Only File System
Attackers with root privileges can have access to any file. He can also access raw devices and corrupt the file system on it. We should mount important file volumes as read-only. But the ordinary mount cannot securely accomplish that because of access to raw devices. A read-only file system must disable file-write system calls and this would also prevent modifying file system through raw devices.

Anti-Malware
It is important to stop malware before it ever gets hold of a system. Although tools that identify malware when they find it on a system are useful, real-time tools that stop it from ever making it to the system are much better. One of the available tools for Windows is Microsoft Security Essentials, and it runs on Windows 7 as well as Windows Vista and Windows XP SP2. A beefed-up Windows Defender replaced Microsoft Security Essentials in Windows 8.

Also note that another free tool from Microsoft is the Malicious Software Removal Tool, which helps remove any infection found but is not intended to be a full anti-malware suite.

An updated version of this tool is released on the second Tuesday of each month, and once installed, it is included, by default, in Microsoft Update and Windows Update.

Patch management
Patch management is an area of systems management that involves acquiring, testing, and installing multiple patches (code changes) to an administered computer system. Patch management tasks include: maintaining current knowledge of available patches, deciding what patches are appropriate for particular systems, ensuring that patches are installed properly, testing systems after installation, and documenting all associated procedures, such as specific configurations required. A number of products are available to automate patch management tasks, including RingMaster's Automated Patch Management, PatchLink Update, and Gibraltar's Everguard.

Like its real world counterpart, a patch is a "make-do" fix rather than an elegant solution. Patches are sometimes ineffective, and can sometimes cause more problems than they fix. Patch management experts, such as Mark Allen, CTO of Gibraltar Software, suggest that system administrators take simple steps to avoid problems, such as performing backups and testing patches on non-critical systems prior to installations.

Patch management is a circular process and must be ongoing. The unfortunate reality about software vulnerabilities is that, after you apply a patch today, a new vulnerability must be addressed tomorrow.

Develop and automate a patch management process that includes each of the following:

- **Detect**. Use tools to scan your systems for missing security patches. The detection should be automated and will trigger the patch management process.
- **Assess**. If necessary updates are not installed, determine the severity of the issue(s) addressed by the patch and the mitigating factors that may influence your decision. By balancing the severity of the issue and mitigating factors, you can determine if the vulnerabilities are a threat to your current environment.
- **Acquire**. If the vulnerability is not addressed by the security measures already in place, download the patch for testing.
- **Test**. Install the patch on a test system to verify the ramifications of the update against your production configuration.
- **Deploy**. Deploy the patch to production computers. Make sure your applications are not affected. Employ your rollback or backup restore plan if needed.
- **Maintain**. Subscribe to notifications that alert you to vulnerabilities as they are reported. Begin the patch management process again

Trusted OS

A trusted operating system (TOS) is any operating system that meets the government's requirements for security. The most common set of standards for security is Common Criteria (CC). This document is a joint effort among Canada, France, Germany, the Netherlands, the United Kingdom, and the United States. The standard outlines a comprehensive set of evaluation criteria, broken down into seven Evaluation Assurance Levels (EALs).

EAL 1 EAL 1 is primarily used when the user wants assurance that the system will operate correctly but threats to security aren't viewed as serious.

EAL 2 EAL 2 requires product developers to use good design practices. Security isn't considered a high priority in EAL 2 certification.

EAL 3 EAL 3 requires conscientious development efforts to provide moderate levels of security.

EAL 4 EAL 4 requires positive security engineering based on good commercial development practices. It is anticipated that EAL 4 will be the common benchmark for commercial systems.

EAL 5 EAL 5 is intended to ensure that security engineering has been implemented in a product from the early design phases. It's intended for high levels of security assurance. The EAL documentation indicates that special design considerations will most likely be required to achieve this level of certification.

EAL 6 EAL 6 provides high levels of assurance of specialized security engineering. This certification indicates high levels of protection against significant risks. Systems with EAL 6 certification will be highly secure from penetration attackers.

EAL 7 EAL 7 is intended for extremely high levels of security. The certification requires extensive testing, measurement, and complete independent testing of every component.

EAL certification has replaced the Trusted Computer Systems Evaluation Criteria (TCSEC) system for certification, which was popular in the United States. It has also replaced the

Information Technology Security Evaluation Criteria (ITSEC), which was popular in Europe. The recommended level of certification for commercial systems is EAL 4.

Currently, only a few operating systems have been approved at the EAL 4 level, and even though an operating system straight out of the box may be, that doesn't mean your own individual implementation of it is functioning at that level. If your implementation doesn't use the available security measures, then you're operating below that level.

As an administrator, you should know and thoroughly understand that just because the operating system you have is capable of being certified at a high level of security doesn't mean that your implementation is at that level.

Host-based firewall

Host-based firewalls for servers typically use rule sets similar to those of network firewalls. Some host-based firewalls for desktops and laptops also use similar rulesets, but most allow or deny activity based on lists of applications. Activity involving any application not on the lists is either denied automatically, or permitted or denied on the basis of the user's response to a prompt asking for a decision about the activity. To prevent malware incidents, organizations should configure host-based firewalls with deny-by-default rulesets for incoming traffic.

In addition to restricting network activity based on rules, many host-based firewalls for workstations incorporate antivirus software and intrusion prevention software capabilities, as well as suppressing Web browser pop-up windows, restricting mobile code, blocking cookies, and identifying potential privacy issues within Web pages and e-mails. Host-based firewalls that integrate these functions can be very effective not only at preventing most types of malware incidents, but also at stopping the spread of malware infections.

For example, a host-based firewall with antivirus capabilities can monitor inbound and outbound e-mails for signs of mass mailing viruses or worms and temporarily shut off e-mail services if such activity is detected. Accordingly, host-based firewalls for workstations that offer several types of malware prevention capabilities typically offer the best single, host-based technical control for malware threat mitigation, as long as they are configured properly and kept up-to-date at all times with the latest signatures and software updates.

Host-based firewalls are particularly important for systems that are network-connected but are not protected by network firewalls and other network-based security controls. Systems that are directly accessible from the Internet should be protected whenever possible through host-based firewalls to prevent network service worms and other threats from connecting to and attacking them.

Host-based Intrusion detection

A host-based IDS (HIDS) is designed to run as software on a host computer system. These systems typically run as a service or as a background process. An HIDS examines the machine logs, system events, and applications interactions; it normally doesn't monitor incoming network traffic to the host. An HIDS is popular on servers that use encrypted channels or channels to other servers.

Two major problems with HIDS aren't easily overcome. The first problem involves a compromise of the system. If the system is compromised, the log files to which the IDS reports may become corrupt or inaccurate. This may make fault determination difficult or it may make the system unreliable. The second major problem with an HIDS is that it must be deployed on each system that needs it. This can create a headache for administrative and support staff.

One of the major benefits of HIDSs is the potential to keep checksums on files. These checksums can be used to inform system administrators that files have been altered by an attack. Recovery is simplified because it's easier to determine where tampering has occurred. The other advantage is that HIDS can read memory when NIDS cannot.
HIDSs typically respond in a passive manner to an incident. An active response would theoretically be similar to those provided by network-based IDS.

Hardware Security
Hardware security involves applying physical security modifications to secure the system(s) and preventing them from leaving the facility. Don't spend all of your time worrying about intruders coming through the network wire while overlooking the obvious need for physical security.

Adding a cable lock between a laptop and a desk prevents someone from picking it up and walking away with a copy of your customer database. All laptop cases include a built- in security slot in which a cable lock can be inserted to prevent it from easily being removed from the premises.

When it comes to desktop models, adding a lock to the back cover can prevent an intruder with physical access from grabbing the hard drive or damaging the internal components. The lock that connects through that slot can also go to a cable that then connects to a desk or other solid fixture to keep the entire PC from being carried away.

In addition to running a cable to the desk, you can choose to run an end of it up to the monitor if theft of peripherals is a problem in your company.

You should also consider using a safe and locking cabinets to protect backup media, documentation, and any other physical artifacts that could do harm if they fell into the wrong hands. Server racks should lock the rack-mounted servers into the cabinets to pre- vent someone from simply pulling one and walking out the front door with it.

Virtualization
An equally popular—and complementary—buzzword to "the cloud" is virtualization. The cost savings promised by virtualization are often offset by the threats to security should the hypervisor be compromised. One reason for the popularity of virtualization is that in order to have cloud computing, you must have virtualization—it is the foundation on which cloud computing is built. At the core of virtualization is the hypervisor, which is the software/hard-ware combination that makes it possible. There are two methods of implementation: Type I and Type II. The Type I hypervisor model, also known as bare metal, is independent of the operating system and boots before the OS. The Type II hypervisor model, also known as hosted, is dependent on the operating system and cannot boot until the OS is up and running. It needs the OS to stay up so that it can boot.

Snapshots
Snapshots allow you to take an image of a system at a particular point in time. With most virtual machine implementations, you can take as many snapshots as you want (provided you have enough storage space) in order to be able to revert a machine to a "saved" state. Snapshots contain a copy of the virtual machine settings (hardware configuration), information on all virtual disks attached, and the memory state of the machine at the time of the snapshot.

Snapshots can also be used for virtual machine cloning, allowing the machine to be copied once—or multiple times—for testing.

Patch Compatibility
As with any server implementation, patch compatibility needs to be factored in before systems are updated. With VMware, patch releases are based on the code from the immediately preceding update, and compatibility for patches is assumed to have the same compatibility as the preceding update release. Although this approach differs for each vendor, most follow similar guidelines.

Host Availability/Elasticity
Host availability is a topic that needs to be addressed in the Service Level Agreement (SLA) with any vendor with whom you contract for cloud services. The goal is always to have minimal downtime: five 9s, or 99.999 percent uptime, is the industry standard. According to NIST, one of the five essential characteristics of the cloud is not just elasticity, but rapid elasticity.

Security Control Testing
Security control testing (SCT) often includes interviews, examinations, and testing of systems to look for weaknesses. It should also include contract reviews of SLAs, a look at the history of prior breaches that a provider has had, a focus on shared resources as well as dedicated servers, and so on.

If this sounds a lot like penetration testing, that is because it is a subset of it. In some organizations, security can be pushed aside in favor of design, and there is a great opportunity for this to happen when transitioning to virtualization and/or cloud computing. As a security professional, it is your responsibility to see that design does not overwhelm security.

Sandboxing
Sandboxing involves running apps in restricted memory areas. By doing so, it is possible to limit the possibility of an app's crash, allowing a user to access another app or the data associated with it. Without sandboxing, the possibility exists that a crash in anothercustomer's implementation could expose a path by which a user might hop ("server hop") to your data. It is important to know that though this possibility exists—and you should test extensively to keep it from happening—the possibility of it occurring has been greatly exaggerated by some in the media.

Section 4.4- Implement the appropriate controls to ensure data security.

Cloud Storage

The first couple of PCs that this author owned booted from media (tape with one and floppies with another) and did not include hard drives. After saving up for quite a while, I bought and installed my first hard drive—costing more than $600. It had a capacity of 20 MB, and I could not fathom what I would possibly do with all of that space.

Today that number is so small, it's laughable. The trend for both individuals and enterprises has been to collect and store as much data as possible. This has led to large local hard drives—DAS (direct attached storage), NAS (network area storage), SANs (storage area networks), and now the cloud.

Just as the cloud holds such promise for running applications, balancing loads, and a plethora of other options, it also offers the ability to store more and more data on it and to let a provider worry about scaling issues instead of local administrators.

SAN
A storage area network (SAN) is a separate network set up to appear as a server to the main organizational network. For example, multiple servers, network storage devices, and switches might be configured to store several terabytes of data. This mini-network has one purpose: to store data. It is then connected to the main organizational network. Users can access the data in the SAN without being concerned about the complexities involved in the SAN. SANs usually have redundant servers, and they are connected via high-speed fiber- optic connections or iSCSI running on copper.
Security for a SAN is similar to that for any server, with the exception of network isolation. There needs to be a firewall, perhaps an intrusion detection system (IDS), user access control, and all of the other security features that you would expect on many networks. SANS are primarily used when there is a large amount of data to store that must be accessible to users on the network.

Big Data
Increasingly, organizations have to store extremely large amounts of data, often many terabytes. This is sometimes referred to simply as Big Data. This data normally cannot fit ona single server, and it is instead stored on a storage area network (SAN). One of the issues with Big Data is that it reaches a size where it becomes difficult to search, to store, to share, to back up, and to truly manage.

Data encryption
Data encryption refers to mathematical calculations and algorithmic schemes that transform plaintext into cypher text, a form that is non-readable to unauthorized parties. The recipient of an encrypted message uses a key, which triggers the algorithm mechanism to decrypt the data, transforming it to the original plaintext version.

Before the Internet, the public seldom used data encryption, as it was more of a military security tool. With the prevalence of online shopping, banking and other services, even basic home users are now aware of data encryption.

Today's web browsers automatically encrypt text when making a connection to a secure server. This prevents intruders from listening in on private communications. Even if they are able to capture the message, encryption allows them to only view scrambled text or what many callUn-readable gibberish. Upon arrival, the data is decrypted, allowing the intended recipient to view the message in its original form.

Types of Data Encryption
There are many different types of data encryption, but not all are reliable. In the beginning, 64-bit encryption was thought to be strong, but was proven wrong with the introduction of 128-bit solutions. AES (Advanced Encryption Standard) is the new standard and permits a maximum of 256-bits. In general, the stronger the computer, the better chance it has at breaking a data encryption scheme.

Data encryption schemes generally fall in two categories: symmetric and asymmetric. AES, DES and Blowfish use symmetric key algorithms. Each system uses a key, which is shared between the sender and the recipient. This key has the ability to encrypt and decrypt the data. With asymmetric encryption such as Diffie-Hellman and RSA, a pair of keys is created and assigned: a private key and a public key. The public key can be known by anyone and used to encrypt data that will be sent to the owner. Once the message is encrypted, the owner of the private key can only decrypt it. Asymmetric encryption is said to be somewhat more secure than symmetric encryption as the private key is not to be shared.

Strong encryption like SSL (Secure Sockets Layer) and TLS (Transport Layer Security) will keep data private, but cannot always ensure security. Websites using this type of data encryption can be verified by checking the digital signature on their certificate, which should be validated by an approved CA (Certificate Authority).

Hardware based encryption devices
In addition to software-based encryption, hardware-based encryption can be applied. Within the advanced configuration settings on some BIOS configuration menus, for exam- ple, you can choose to enable or disable TPM. A *Trusted Platform Module (TPM)* can be used to assist with hash key generation. TPM is the name assigned to a chip that can store cryptographic keys, passwords, or certificates. TPM can be used to protect smart phones and devices other than PCs as well. It can also be used to generate values used with whole disk encryption such as BitLocker. BitLocker can be used with or without TPM. It is much more secure when coupled with TPM (and is preferable) but does not require it.

The TPM chip may be installed on the motherboard; when it is, in many cases it is set to off in the BIOS by default.

In addition to TPM, *HSM (Hardware Security Module)* is also a cryptoprocessor that can be used to enhance security. HSM is commonly used with PKI systems to augment security with CAs. As opposed to being mounted on the motherboard like TPMs, HSMs are traditionally PCI adapters.

Data in-transit/Data at-rest, Data in-use

Protecting Data at Rest:

You must either (a) encrypt the entire contents of the storage media or (b) you must have complete knowledge of how any system or user organizes data when writing to the storage media so that you can encrypt the data that needs to be protected. (a) is FDE. (b) can accomplished by any one of a number of other solutions, but is very difficult because even if you know how the system stores everything, you don’t know (or have to enforce through restriction) how the user may store something (you must disable his/her ability to store

anything "sensitive" on the media in a location that is not encrypted). Furthermore (c) you must enforce strong keys/passwords and (d) you must prevent the user from storing the password on the media.

Finally, remember, (e) for detachable media, including laptop hard drives, the USER is considered the "node associated with the media", so really, your data can't be considered secure, because the user is the node, and the user has the key. (Unless, I suppose, you have the ability to revoke the key remotely, preventing Disgruntled Joe from taking a laptop out and then quitting with a copy of your code base already in his possession).

By far, (c)/(d)/(e) are going to be the hardest. A suitably strong password that prevents a dictionary attack is going to be burdensome to the user to retain, so they're either going to forget it, or write it down and stickynote it to the monitor, etc. The only way to mitigate this risk effectively is to *limit access to the data in the first place* – people look at FDE as a "silver bullet" to allow them to say, "We can now allow our vice president to take a copy of the financial database home on his laptop, because it is encrypted, so we don't have to worry if the laptop is stolen", but that assumes that (c)/(d)/(e) aren't problems, which is screwy. Sensitive data shouldn't leave the house, people. If the VP wants access to the data because it makes his life easier, say "No, you need to be in the office to get access to that," or make sure ahead of time that everyone at the CEO/Board of Directors level knows that you have *no real data protection* – your data is only as secure as everyone is trustworthy. And while I may trust a particular worker to not read data to a corporate rival over the phone, I simply don't trust any number of workers > 2 to *not put their password on a sticky note on the screen of their laptop*.

Protecting Data in Use:
This is basically impossible in today's OS market… anyone who claims that they have "secure data is use" is full of baloney. The best you can do here is mitigate the attack vectors. If you use FDE, you solve some of the problems because the swap space is encrypted, which prevents one attack vector, or you can get rid of swap altogether (and make sure you're not using NVRAM). However, if you look at the various ways that Data in Use can be mishandled, virtually all of the major vulnerabilities are exploitable at the OS level, which is something that you've more or less outsourced to your OS vendor. Your only mitigation here is to lock down the OS as much as you possibly can (including using FDE to protect the OS files at rest!), and this is more often way more trouble than it is worth, given that even if you could cover all of your bases, it doesn't protect from Kevin Mitnick. From a cost/benefit analysis, aside from taking basic steps to secure an operating system, you're wasting money – locking down Windows to the point of near un-usability isn't going to protect you from a zero-day IE exploit.

The number one way to prevent OS level exploits is to use a web proxy at your border and disallow all attachments via email. Anybody who can successfully sell #2, please let me know how you did it. If you can't do those two things, though, spending more than a minimal effort locking down the host OS is largely a waste of time.

Protecting Data in Transit:
Here's where S/MIME and SSL and IPSec and all that good stuff comes in. Actually, next to protecting Data at Rest, protecting Data in Transit is probably one of the easier tasks to accomplish at the present time, except for the fact that both hosts have to be able to protect the Data in Use, and we illustrated in the previous paragraph how hard that is. Yes, you can

man-in-the-middle data in transit in many, many instances in today's networked world, but we already have many of the technologies to mitigate this; we just don't deploy them properly.

Permissions/ACL
User permissions may be the most basic aspect of security. Remember the concept of least privileges, which means that any given user will be granted only the privileges necessary to perform their job function.

Microsoft describes five file permissions and one additional folder permission:

Full Control This means the user cannot only read, executes, and write, but they can also assign permissions to other uses.
Modify This is the same as read and write, with delete added.
Read and Execute Not all files are documents. For example, programs are files, and theRead and Execute privilege is needed to run the program.
Read This permission allows the user to read the file but not to modify it.
Write This permission allows the user to modify the file.

Folders have the same permissions, with one added permission: list folder contents. This permission allows the user to see what is in a folder but not to read the files.
Access Control Lists
Related to permissions is the concept of the access control list (ACL). An ACL is literally a list of who can access what resource and at what level. It can be an internal part of an operating system or application. For example, a custom application might have an ACL that lists which users have what permissions (access levels) in that system.

An ACL can also be a physical list of who is allowed to enter a room or a building. This is a much less common definition for ACL, but it is relevant to physical security.
Related to ACLs are white lists and black lists. In fact, you could consider these to be special types of access control lists. Essentially, a white list is a list of items that are allowed. It could be a list of websites that are okay to visit with company computers, or it could be a list of third-party software that is authorized to be installed on company computers. Black lists are the opposite. They are lists of things that are prohibited. It could be specific websites that employees should not visit or software that is forbidden to be installed on client computers.

Section 4.5- Compare and contrast alternative methods to mitigate security risks in static environments.

Environments

SCADA (supervisory control and data acquisition) refers to equipment often used to manage automated factory equipment, dams, power generators, and similar equipment. The Security+ exam does not heavily emphasize this, because the security measures will depend on the device. However, the infamous Stuxnet virus targeted specific SCADA equipment, so the need for SCADA security is not simply hypothetical.

Embedded systems (such as printers, smart TVs, and HVAC controls) have their own security needs. Most modern printers, even midrange printers, have hard drives, RAM, and an operating system. That means they have specific vulnerabilities. Some advanced HVAC control systems and smart TVs also have sophisticated operations that are vulnerable to attack. Even game consoles can be vulnerable to viruses. Like SCADA, the specifics of mitigating risk will depend on the device, but the Security+ exam will expect you to be aware that these devices have security risks.

Smartphones are probably a more obvious security risk. All of those issues obviously apply to smart phones. But specific phones, such as Android and IOS, will have their own security issue that have to be addressed.

Mainframes usually do not present significant security risks; they tend to be more stable and less susceptible to attacks. However, that does not mean they are invulnerable. You should examine the mainframe your organization uses and see what steps are appropriate for that system.

A new and emerging issue is that of in-vehicle computing systems. Automobiles tend to have sophisticated systems, such as computers complete with hard drives and GPS. There have already been preliminary security tests showing that these systems can be breached. Much like SCADA, the specifics will depend on the implementation. The Security+ test will ask you about the concept in a general way.

Methods

Some generalized methods can be used to mitigate the security risks to any network. One of the most basic is the combination of network segmentation and security layers. These are very closely related subjects. Network segmentation means dividing your network into segments. Ideally the connection points between each segment (routers) will also implement security features such as a firewall and intrusion detection system. This means that a breach of one segment of your network does not jeopardize the entire network. It is only logical to segment your network based on security layers, or zones based on security needs. The most obvious example is an external zone (called a demilitarized zone [DMZ]) for publicly accessible resources like a web server, and an internal zone for your actual corporate network. You can use as many zones as are needed, each with a different (but appropriate) level of security.

Network protection can be enhanced with some simple techniques. Application firewalls are usually better protection for database servers or web servers than are other types of firewalls. Application firewalls, in addition to packet filtering, filter specific application- related content. For example, a web server might use an application firewall to filter common SQL injection attacks.

It is just as important to make sure firmware updates are applied. Firmware version control is closely related to updating the firmware. You need to be sure that each device is using the appropriate version of firmware. You may even need to manually update devices with critical updates. Certain viruses specifically target the firmware in routers and switches. This risk is mitigated by firmware version control.

One very important technique is controlling redundancy and diversity. Although this may sound complex, it simply means two things. The first is implementing more than one of each security control. If you have an intrusion detection system in your DMZ, you may want to have another in your network. Diversity means using different controls of the same type. For example, if you use the Cisco IDS on your perimeter, you may wish to use SNORT IDS inside your network. The reasoning is that if an attack thwarts one of your IDSs, it may not evade both. This concept applies to all security controls. Wrappers are a related topic. This technique involves wrapping sensitive systems with a specific control, such as having your sensitive data servers in their own network segment with their own firewall, IDS, and antivirus protection.
A variety of specialized systems have security issues specific to those systems. You must mitigate the risk on each of these systems.

Topic 5.0 Access Control and Identity Management

Section 5.1 Compare and contrast the function and purpose of authentication services.

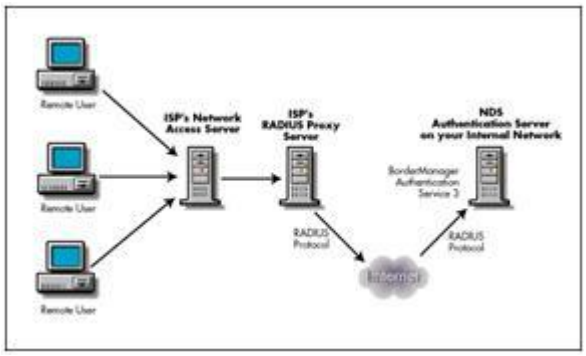
RADIUS
RADIUS (Remote Authentication Dial In User Service), defined in RFC 2865, is a protocol for remote user authentication and accounting.

RADIUS enables centralized management of authentication data, such as usernames and passwords.

When a user attempts to login to a RADIUS client, such as a router, the router send the authentication request to the RADIUS server. The communication between the RADIUS client and the RADIUS server are authenticated and encrypted through the use of a shared secret, which is not transmitted over the network.

The RADIUS server may store the authentication data locally, but it can also store authentication data in an external SQL database or an external Unix /etc/passwd file. The RADIUS server can also plug into a PAM (Pluggable Authentication Service) architecture to retrieve authentication data.

The role of the RADIUS server as the centralized authentication server makes is an excellent choice for also performing accounting



RADIUS can significantly increase security by enabling the centralization of password management. Of course, the other side of that argument is that once you take over the RADIUS server, you have everything.

RADIUS servers are available from many vendors. In addition, GNU RADIUS is an excellent non-commercial option.

RADIUS utilizes the MD5 algorithm for secure password hashing.

RADIUS is the de facto authentication provider in 802.11i wireless networks.

TACACS and TACACS+

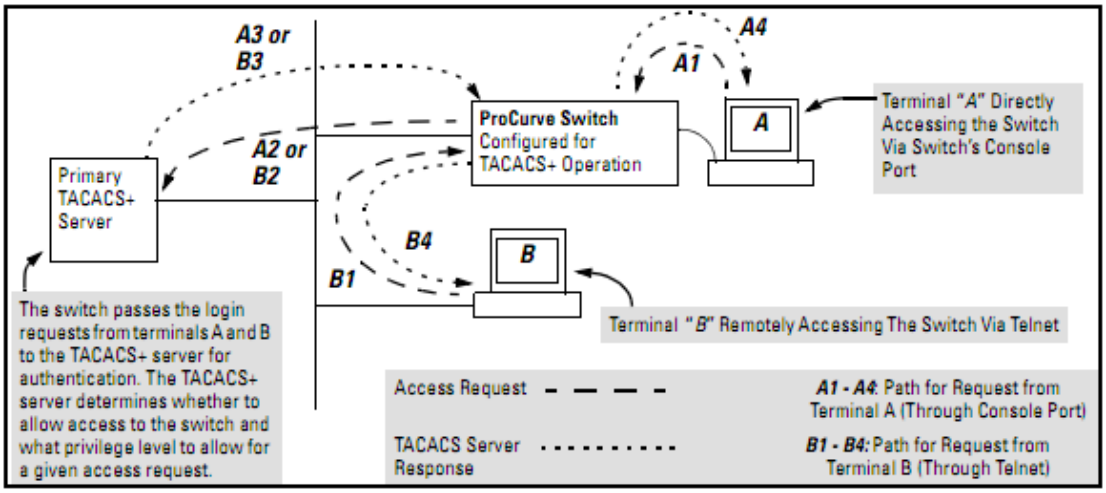
Terminal Access Controller Access-Control System (TACACS) is a remote authentication protocol that is used to communicate with an authentication server commonly used in UNIX networks. TACACS allows a remote access server to communicate with an authentication server in order to determine if the user has access to the network. TACACS is defined in RFC 1492, and uses (either TCP or UDP) port 49 by default. A later version of TACACS introduced by Cisco in 1990 was called Extended TACACS (XTACACS). The XTACACS protocol was developed by and is proprietary of Cisco Systems.

TACACS allows a client to accept a username and password and send a query to a TACACS authentication server, sometimes called a TACACS daemon or simply TACACSD. This server was normally a program running on a host. The host would determine whether to accept or deny the request and send a response back. The TIP (routing node accepting dial-up line connections, which the user would normally want to log in into) would then allow access or not, based upon the response. In this way, the process of making the decision is "opened up" and the algorithms and data used to make the decision are under the complete control of whoever is running the TACACS daemon.

TACACS+ and RADIUS have generally replaced TACACS and XTACACS in more recently built or updated networks. TACACS+ is an entirely new protocol and not compatible with TACACS or XTACACS. TACACS+ uses the Transmission Control Protocol (TCP) and RADIUS uses the User Datagram Protocol (UDP). Some administrators recommend using TACACS+ because TCP is seen as a more reliable protocol. Whereas RADIUS combines authentication and authorization in a user profile, TACACS+ separates the two operations.

TACACS+ in the switch manages authentication of logon attempts through either the Console port or Telnet. TACACS+ uses an authentication hierarchy consisting of (1) remote

passwords assigned in a TACACS+ server and (2) local passwords configured on the switch. That is, with TACACS+ configured, the switch first tries to contact a designated TACACS+ server for authentication services. If the switch fails to connect to any TACACS+ server, it defaults to its own locally assigned passwords for authentication control if it has been configured to do so. For both Console and Telnet access you can configure a login (read-only) and an enable (read/write) privilege level access.



Kerberos

Kerberos is a distributed authentication service that allows a process (a client) running on behalf of a principal (a user) to prove its identity to a verifier (an application server, or just server) without sending data across the network that might allow an attacker or the verifier to subsequently impersonate the principal. Kerberos optionally provides integrity and confidentiality for data sent between the client and server. Kerberos was developed in the mid-'80s as part of MIT's Project Athena. As use of Kerberos spread to other environments, changes were needed to support new policies and patterns of use. To address these needs, design of Version 5 of Kerberos (V5) began in 1989. Though V4 still runs at many sites, V5 is considered to be standard Kerberos.

Limitations of Kerberos

Limitations of Kerberos have been described in the literature. In particular, Kerberos is not effective against password guessing attacks; if a user chooses a poor password, then an attacker guessing that password can impersonate the user. Similarly, Kerberos requires a trusted path through which passwords are entered. If the user enters a password to a program that has already been modified by an attacker (a Trojan horse), or if the path between the user and the initial authentication program can be monitored, then an attacker may obtain sufficient information to impersonate the user. Kerberos can be combined with other techniques, to address these limitations.

To be useful, Kerberos must be integrated with other parts of the system. It does not protect all messages sent between two computers; it only protects the messages from software that has been written or modified to use it. While it may be used to exchange encryption keys when establishing link encryption and network level security services, this would require changes to the network software of the hosts involved.

Kerberos does not itself provide authorization, but V5 Kerberos passes authorization information generated by other services. In this manner, Kerberos can be used as a base for building separate distributed authorization services.

How Kerberos works
The Kerberos Authentication System uses a series of encrypted messages to prove to a verifier that a client is running on behalf of a particular user. The Kerberos protocol is based in part on the Needham and Schroeder authentication protocol, but with changes to support the needs of the environment for which it was developed. Among these changes are the use of timestamps to reduce the number of messages needed for basic authentication, the addition of a ``ticket-granting'' service to support subsequent authentication without re-entry of a principal's password, and different approach to cross-realm authentication (authentication of a principal registered with a different authentication server than the verifier).

The description is simplified for clarity; additional fields are present in the actual protocol. Readers should consult RFC 151 for a more thorough description of the Kerberos protocol.

Kerberos Encryption
Though conceptually, Kerberos authentication proves that a client is running on behalf of a particular user, a more precise statement is that the client has knowledge of an encryption key that is known by only the user and the authentication server. In Kerberos, the user's encryption key is derived from and should be thought of as a password; we will refer to it as such in this article. Similarly, each application server shares an encryption key with the authentication server; we will call this key the server key.

Encryption in the present implementation of Kerberos uses the data encryption standard (DES). It is a property of DES that if ciphertext (encrypted data) is decrypted with the same key used to encrypt it, the plaintext (original data) appears. If different encryption keys are used for encryption and decryption, or if the ciphertext is modified, the result will be unintelligible, and the checksum in the Kerberos message will not match the data. This combination of encryption and the checksum provides integrity and confidentiality for encrypted Kerberos messages.

LDAP
The Lightweight Directory Access Protocol, better known as LDAP, is based on the X.500 standard, but significantly simpler and more readily adapted to meet custom needs. Unlike X.500, LDAP supports TCP/IP, which is necessary for Internet access. The core LDAP specifications are all defined in RFCs.

The advantages of LDAP directories
Now that we've straightened that out, what are the advantages of LDAP directories? The current popularity of LDAP is the culmination of a number of factors. I'll give you a few basic reasons, provided you keep in mind that it's just part of the story.

Perhaps the biggest plus for LDAP is that your company can access the LDAP directory from almost any computing platform, from any one of the increasing number of readily available, LDAP-aware applications. It's also easy to customize your company's internal applications to add LDAP support.

The LDAP protocol is both cross-platform and standards-based, so applications needn't worry

about the type of server hosting the directory. In fact, LDAP is finding much wider industry acceptance because of its status as an Internet standard. Vendors are more willing to write LDAP integration into their products because they don't have to worry about what's at the other end. Your LDAP server could be any one of a number of open-source or commercial LDAP directory servers (or perhaps even a DBMS server with an LDAP interface), since interacting with any true LDAP server involves the same protocol, client connection package, and query commands. By contrast, vendors looking to integrate directly with a DBMS usually must tailor their product to work with each database server vendor individually.

Unlike many relational databases, you do not have to pay for either client connection software or for licensing.

Most LDAP servers are simple to install, easily maintained, and easily optimized.

LDAP servers can replicate either some or all of their data via push or pull methods, allowing you to push data to remote offices, to increase security, and so on. The replication technology is built-in and easy to configure. By contrast, many of the big DBMS vendors charge extra for this feature, and it's far more difficult to manage.

LDAP allows you to securely delegate read and modification authority based on your specific needs using ACIs (collectively, an ACL, or Access Control List). For example, your facilities group might be given access to change an employee's location, cube, or office number, but not be allowed to modify entries for any other fields. ACIs can control access depending on who is asking for the data, what data is being asked for, where the data is stored, and other aspects of the record being modified. This is all done through the LDAP directory directly, so you needn't worry about making security checks at the user application level.

LDAP is particularly useful for storing information that you wish to read from many locations, but update infrequently. For example, your company could store all of the following very efficiently in an LDAP directory:

- The company employee phone book and organizational chart
- External customer contact information
- Infrastructure services information, including NIS maps, email aliases, and so on
- Configuration information for distributed software packages
- Public certificates and security keys

SAML

Security Assertion Markup Language (SAML) is an open standard based on XML that is used for authentication and authorization data. Service providers often use SAML to prove the identity of someone connecting to the service provider. The current version is SAML v2.0.

Section 5.2 Given a scenario, select the appropriate authentication, authorization or access control.

Identification vs. authentication vs. authorization

Identification is the process whereby a network element recognizes a valid user's identity.

Authentication is the process of verifying the claimed identity of a user. A user may be a person, a process, or a system (e.g., an operations system or another network element) that accesses a network element to perform tasks or process a call. A user identification code is a non-confidential auditable representation of a user. Information used to verify the claimed identity of a user can be based on a password, Personal Identification Number (PIN), smart card, biometrics, token, exchange of keys, etc. Authentication information should be kept confidential.

If users are not properly identified then the network element is potentially vulnerable to access by unauthorized users. Because of the open nature of ONA, ONA greatly increases the potential for unauthorized access. If strong identification and authorization mechanisms are used, then the risk that unauthorized users will gain access to a system is significantly decreased.

The exploitation of the following vulnerabilities, as well as other identification and authentication vulnerabilities, will result in the threat of impersonating a user.

- Weak authentication methods are used;
- The potential exists for users to bypass the authentication mechanism;
- The confidentiality and integrity of stored authentication information is not preserved, and
- Authentication information, which is transmitted over the network, is not encrypted.

Computer intruders have been known to compromise PSN assets by gaining unauthorized access to network elements.

The severity of the threat of impersonating a user depends on the level of privilege that is granted to the unauthorized user.

Authorization

The Least-Privileged User Account Approach
A defense-in-depth strategy, with overlapping layers of security, is the best way to counter these threats, and the least-privileged user account (LUA) approach is an important part of that defensive strategy. The LUA approach ensures that users follow the principle of least privilege and always log on with limited user accounts. This strategy also aims to limit the use of administrative credentials to administrators, and then only for administrative tasks.

The LUA approach can significantly mitigate the risks from malicious software and accidental incorrect configuration. However, because the LUA approach requires organizations to plan, test, and support limited access configurations, this approach can generate significant costs and challenges. These costs can include redevelopment of custom programs, changes to operational procedures, and deployment of additional tools.

It is difficult to find utilities and guidance on using limited user accounts, so this white paper refers to third-party tools and guidance from Web logs and other unofficial sources. Microsoft makes no warranty about the suitability of the tools or guidance for your environment. You should test any of these instructions or programs before you deploy them. As with all security issues, there is no perfect answer, and this software and guidance is no

exception.

Separation of Duties

Another fundamental approach to security is separation of duties. This concept is applicable to physical environments as well as network and host security. Separation of duty ensures that for any given task, more than one individual needs to be involved. The task is broken into different duties, each of which is accomplished by a separate individual. By implementing a task in this manner, no single individual can abuse the system for his or her own gain. This principle has been implemented in the business world, especially financial institutions, for many years. A simple example is a system in which one individual is required to place an order and a separate person is needed to authorize the purchase.

While separation of duties provides a certain level of checks and balances, it is not without its own drawbacks. Chief among these is the cost required to accomplish the task. This cost is manifested in both time and money. More than one individual is required when a single person could accomplish the task, thus potentially increasing the cost of the task. In addition, with more than one individual involved, a certain delay can be expected, as the task must proceed through its various steps.

Discretionary Access Control

Both discretionary access control and mandatory access control are terms originally used by the military to describe two different approaches to controlling an individual's access to a system. As defined by the "Orange Book," a Department of Defense document that at one time was the standard for describing what constituted a trusted computing system, DACs are "a means of restricting access to objects based on the identity of subjects and/or groups to which they belong. The controls are discretionary in the sense that a subject with a certain access permission is capable of passing that permission (perhaps indirectly) on to any other subject." While this might appear to be confusing "government-speak," the principle is rather simple. In systems that employ DACs, the owner of an object can decide which other subjects can have access to the object and what specific access they can have. One common method to accomplish this is the permission bits used in UNIX-based systems. The owner of a file can specify what permissions (read/write/execute) members in the same group can have and also what permissions all others can have. ACLs are also a common mechanism used to implement DAC.

Mandatory Access Control

A less frequently employed system for restricting access is mandatory access control. This system, generally used only in environments in which different levels of security classifications exist, is much more restrictive regarding what a user is allowed to do. Referring to the "Orange Book," a mandatory access control is "a means of restricting access to objects based on the sensitivity (as represented by a label) of the information contained in the objects and the formal authorization (i.e., clearance) of subjects to access information of such sensitivity." In this case, the owner or subject can't determine whether access is to be granted to another subject; it is the job of the operating system to decide.

Role-Based Access Control

ACLs can be cumbersome and can take time to administer properly. Another access control mechanism that has been attracting increased attention is the role-based access control (RBAC). In this scheme, instead of each user being assigned specific access permissions for the objects associated with the computer system or network, each user is assigned a set of roles that he or she may perform. The roles are in turn assigned the access permissions

necessary to perform the tasks associated with the role. Users will thus be granted permissions to objects in terms of the specific duties they must perform—not according to a security classification associated with individual objects.

Rule-Based Access Control
The first thing that you might notice is the ambiguity that is introduced with this access control method also using the acronym RBAC. Rule-based access control again uses objects such as ACLs to help determine whether access should be granted or not. In this case, a series of rules are contained in the ACL and the determination of whether to grant access will be made based on these rules. An example of such a rule is one that states that no employee may have access to the payroll file after hours or on weekends.

As with MAC, users are not allowed to change the access rules, and administrators are relied on for this. Rule-based access control can actually be used in addition to or as a method of implementing other access control methods. For example, MAC methods can utilize a rule-based approach for implementation.

Time of day restrictions
Some systems allow for the specification of time of day restrictions in their access control policies. This means that a user's access to the system or specific resources can be restricted to certain times of the day and days of the week. If a user normally accesses certain resources during normal business hours, an attempt to access these resources outside this time period (either at night or on the weekend) might indicate an attacker has gained access to the account. Specifying time of day restrictions can also serve as a mechanism to enforce internal controls of critical or sensitive resources. Obviously, a drawback to enforcing time of day restrictions is that it means that a user can't go to work outside of normal hours in order to "catch up" with work tasks. As with all security policies, usability and security must be balanced in this policy decision.

Authentication

Smart card
A smart card is a plastic card about the size of a credit card, with an embedded microchip that can be loaded with data, used for telephone calling, electronic cash payments, and other applications, and then periodically refreshed for additional use. Currently or soon, you may be able to use a smart card to:

- Dial a connection on a mobile telephone and be charged on a per-call basis
- Establish your identity when logging on to an Internet access provider or to an online bank
- Pay for parking at parking meters or to get on subways, trains, or buses
- Give hospitals or doctors personal data without filling out a form
- Make small purchases at electronic stores on the Web (a kind of cybercash)
- Buy gasoline at a gasoline station

Over a billion smart cards are already in use. Currently, Europe is the region where they are most used. Ovum, a research firm, predicts that 2.7 billion smart cards will be shipped annually by 2003. Another study forecasts a $26.5 billion market for recharging smart cards

by 2005. Compaq and Hewlett-Packard are reportedly working on keyboards that include smart card slots that can be read like bank credit cards. The hardware for making the cards and the devices that can read them is currently made principally by Bull, Gemplus and Schlumberger.

Implicit Deny

What has become the Internet was originally designed as a friendly environment where everybody agreed to abide by the rules implemented in the various protocols. Today, the Internet is no longer the friendly playground of researchers that it once was. This has resulted in different approaches that might at first seem less than friendly but that are required for security purposes. One of these approaches is implicit deny. Frequently in the network world, decisions concerning access must be made. Often a series of rules will be used to determine whether or not to allow access. If a particular situation is not covered by any of the other rules, the implicit deny approach states that access should not be granted. In other words, if no rule would allow access, then access should not be granted. Implicit deny applies to situations involving both authorization and access.

Access Control

The term access control describes a variety of protection schemes. It sometimes refers to all security features used to prevent unauthorized access to a computer system or network. In this sense, it may be confused with authentication. More properly, access is the ability of a subject (such as an individual or a process running on a computer system) to interact with an object (such as a file or hardware device). Authentication, on the other hand, deals with verifying the identity of a subject.

To understand the difference, consider the example of an individual attempting to log in to a computer system or network. Authentication is the process used to verify to the computer system or network that the individual is who he claims to be. The most common method to do this is through the use of a user ID and password. Once the individual has verified his identity, access controls regulate what the individual can actually do on the system—just because a person is granted entry to the system does not mean that he should have access to all data the system contains.

Consider another example. When you go to your bank to make a withdrawal, the teller at the window will verify that you are indeed who you claim to be by asking you to provide some form of identification with your picture on it, such as your driver's license. You might also have to provide your bank account number. Once the teller verifies your identity, you will have proved that you are a valid (authorized) customer of this bank. This does not, however, mean that you have the ability to view all information that the bank protects—such as your neighbor's account balance. The teller will control what information, and funds, you can access and will grant you access only to information for which you are authorized to see. In this example, your identification and bank account number serve as your method of authentication and the teller serves as the access control mechanism.

In computer systems and networks, access controls can be implemented in several ways. An access control matrix provides the simplest framework for illustrating the process. In this matrix, the system is keeping track of two processes, two files, and one hardware device. Process 1 can read both File 1 and File 2 but can write only to File 1. Process 1 cannot access Process 2, but Process 2 can execute Process 1. Both processes have the ability to write to the printer. While simple to understand, the access control matrix is seldom used in computer

systems because it is extremely costly in terms of storage space and processing. Imagine the size of an access control matrix for a large network with hundreds of users and thousands of files. The actual mechanics of how access controls are implemented in a system varies, though access control lists (ACLs) are common. An ACL is nothing more than a list that contains the subjects that have access rights to a particular object. The list identifies not only the subject but the specific access granted to the subject for the object. Typical types of access include read, write, and execute as indicated in the example access control matrix.

No matter what specific mechanism is used to implement access controls in a computer system or network, the controls should be based on a specific model of access.

Single sign on
Single sign-on (SSO) is mechanism whereby a single action of user authentication and authorization can permit a user to access all computers and systems where he has access permission, without the need to enter multiple passwords. Single sign-on reduces human error, a major component of systems failure and is therefore highly desirable but difficult to implement.

The SSO feature maintains a mapping between a user, or group of users, and the credentials (username and password) needed to access a particular data source. This mapping is referred to as an Application Definition (or "app def" for short). Only server administrators can create and modify app defs, using the browser-based Central Admin UI.

When the DataFormWebPart needs to access a remote data source using SSO, it calls the SSO API to retrieve the necessary credentials for the given app def. If they happen to be Windows credentials, the web part temporarily "logs in" with those credentials, and then attempts to connect to the data source. This means the Windows credentials are only making one hop – from the web server to the database server – and not two.

SPD can create Data Views using both kinds of SSO app defs:

- **Group**
- **Individual**

A **Group** app def is used to let everyone in a domain group access a database using a single account. For example, you might have a special account for database use that only has read-only permissions on a few tables; SSO lets you force everyone in your workgroup to connect to the database with the limited permission account.

An **Individual** app def lets users provide their own account information (username and password). The first time a connection is attempted and the end user's credentials are not already in SSO, the Data View will redirect to a web page to collect and store them. Subsequent attempts will reuse stored credentials without prompting.

Either type of app def can be used to store Windows credentials (from a network domain account, in the format "domain\account"), or basic (non-Windows) credentials. In the case of SQL Server, you can use SSO to establish either Windows auth connections or SQL auth connections. You can also use SSO access web services that require a specific username and password, or Windows authentication.

Authentication factors
When two or more access methods are included as part of the authentication process, you're implementing a multifactor authentication system. A system that uses smart cards and passwords is referred to as a two-factor authentication system. This example requires both a smart card and a logon password process.

A multifactor system can consist of a two-factor system, three-factor system, and so on. As long as more than one factor is involved in the authentication process, it is considered a multifactor system.

For obvious reasons, the two or more factors employed should not be from the same category. Although you do increase difficulty in gaining system access by requiring the user to enter two sets of username/password combinations, it is much preferred to pair a single username/password combination with a biometric identifier or other security check.

The most basic form of authentication is known as single-factor authentication (SFA), because only one type of authentication is checked. SFA is most often implemented as the traditional username/password combination. A username and password are unique identifiers for a logon process. Here's a synopsis for how SFA works: When users sit down in front of a computer system, the first thing a security system requires is that they establish who they are. Identification is typically confirmed through a logon process. Most operating systems use a user ID (username) and password to accomplish this. These values can be sent across the connection as plain text or they can be encrypted.

The logon process identifies that you are who you say you are to the operating sys- tem and possibly the network. Figure 4.1 illustrates the logon and password process. Notice that the operating system compares this information to the stored information from the security processor, and it either accepts or denies the logon attempt. The operating system might establish privileges or permissions based on stored data about that particular user ID.

Whenever two or more parties authenticate each other, it is known as mutual authentication. A client may authenticate to a server and a server may authenticate to a client when there is a need to establish a secure session between the two and employ encryption. Mutual authentication ensures that the client is not unwittingly connecting and providing its credentials to a rogue server, which can then turn around and steal the data from the real server.

Identification
Understanding the difference between identification and authentication is critical to correctly answering access control questions on the Security+ exam. Identification means finding out who someone is. Authentication is a mechanism of verifying that identification. Put another way, identification is claiming an identity; authentication is proving it.

In the physical world, the best analogy would be that any person can claim to be anyone (identification). To prove it (authentication), however, that person needs to provide some evidence, such as a driver's license, passport, and so forth.

- Authentication systems or methods are based on one or more of these five factors:
- Something you know, such as a password or PIN

- Something you have, such as a smart card, token, or identification device
- Something you are, such as your fingerprints or retinal pattern (often called biometrics) Something you do, such as an action you must take to complete authentication Somewhere you are (this is based on geo-location)

Federation

A federation is a collection of computer networks that agree on standards of operation, such as security standards. Normally, these are networks that are related in some way. In some cases, it could be an industry association that establishes such standards.

Another example of a federation would be an instant messaging (IM) federation. In this scenario, multiple IM providers form common communication standards, thus allowing users on different platforms with different clients to communicate freely.
In other situations, a group of partners might elect to establish common security and communication standards, thus forming a federation. This would facilitate communication between employees in each of the various partners.

A federated identity is a means of linking a user's identity with their privileges in a manner that can be used across business boundaries (for example, Microsoft Passport or Google checkout). This allows a user to have a single identity that they can use across different business units and perhaps even entirely different businesses.

Transitive trust/authentication

The word transitive means involving transition—keep this in mind as you learn how transitive access problems occur. With transitive access, one party (A) trusts another party (B).

If the second party (B) trusts another party (C), then a relationship can exist where the first party (A) also may trust the third party (C).

In early operating systems, this process was often exploited. In current operating systems, such as Windows Server 2012, the problems with transitive access are solved by creating transitive trusts, which are a type of relationship that can exist between domains (the opposite is non-transitive trusts). When the trust relationship is transitive, the relation- ship between party (A) and party (B) flows through as described earlier (for instance, A now trusts C). In all versions of Active Directory, the default is that all domains in a forest trust each other with two-way, transitive trust relationships.

Section 5.3 Install and configure security controls when performing account management, based on best practices.

Mitigate issues associated with users with multiple account/roles and/or shared accounts

Account policy enforcement
Policy enforcement is the manner in which the Server allows or disallows accounts that violate provisioning policies.

When policy, person, or account data is changed, an account that was originally compliant with a provisioning policy can become noncompliant.

Policy enforcement can be configured globally or for a specific service.

Policy enforcement occurs whenever it is necessary in the server business process to ensure system integrity and perform appropriate actions. A provisioning policy governs the access rights for users for specific services, thus provisioning enforcement is incorporated into all server business processes that manage a user's identity and access rights on a managed resource.

All services except a few like IBM's DSML Identity Feed services have policy enforcement available.

A service's policy enforcement can be modified at any time. When the change is made, policy enforcement will be scheduled. Note that changes made to policy enforcement action might trigger a chain of events as result of the change.

Password complexity
Most users log on to their local computer and to remote computers by using a combination of their user name and a password typed at the keyboard. Although alternative technologies for authentication, such as biometrics, smartcards, and one-time passwords, are available for all popular operating systems, most organizations still rely on traditional passwords and will continue to do so for years to come. Therefore it is very important that organizations define and enforce password policies for their computers that include mandating the use of strong passwords. Strong passwords meet a number of requirements for complexity - including length and character categories - that make passwords more difficult for attackers to determine. Establishing strong password policies for your organization can help prevent attackers from impersonating users and can thereby help prevent the loss, exposure, or corruption of sensitive information.

Depending on whether the computers in your organization are members of an Active Directory domain, stand-alone computers, or both, to implement strong password policies you will need to perform one or both of the following tasks:

- Configure password policy settings in an Active Directory Domain.
- Configure password policy settings on stand-alone computers.

Once you have configured the appropriate password policy settings, users in your organization will be able to create new passwords only if the passwords meet the length and complexity requirements for strong passwords, and users will not be able to immediately change their new passwords.

For Windows 2000, Windows XP, and Windows Server 2003 and Windows Server 2008 there are five settings you can configure that relate to password characteristics: **Enforce**

password history, **Maximum password age**, **Minimum password age**, **Minimum password length**, and **Passwords must meet complexity requirements**.

- **Enforce password history** determines the number of unique new passwords a user must use before an old password can be reused. The value of this setting can be between 0 and 24; if this value is set to 0, enforce password history is disabled. For most organizations, set this value to 24 passwords.
- **Maximum password age** determines how many days a password can be used before the user is required to change it. The value of this between 0 and 999; if it is set to 0, passwords never expire. Setting this value too low can cause a frustration for your users; setting it too high or disabling it gives potential attackers more time to determine passwords. For most organizations, set this value to 42 days.
- **Minimum password age** determines how many days a user must keep new passwords before they can change them. This setting is designed to work with the **Enforce password history** setting so that users cannot quickly reset their passwords the required number of times and then change back to their old passwords. The value of this setting can be between 0 and 999; if it is set to 0, users can immediately change new passwords. It is recommended that you set this value to 2 days.
- **Minimum password length** determines how short passwords can be. Although Windows 2000, Windows XP, and Windows Server 2003 support passwords up to 28 characters, the value of this setting can be only between 0 and 4 characters. If it is set to 0, users are allowed to have blank passwords, so you should not use a value of 0. It is recommended that you set this value to 8 characters.
- **Passwords must meet complexity requirements** determines whether password complexity is enforced. If this setting is enabled, user passwords meet the following requirements:
  - The password is at least six characters long.
  - The password contains characters from at least three of the following five categories:
    - English uppercase characters (A - Z)
    - English lowercase characters (a - z)
    - Base 10 digits (0 - 9)
    - Non-alphanumeric (For example: !, $, #, or %)
    - Unicode characters
  - The password does not contain three or more characters from the user's account name.

    If the account name is less than three characters long, this check is not performed because the rate at which passwords would be rejected is too high. When checking against the user's full name, several characters are treated as delimiters that separate the name into individual tokens: commas, periods, dashes/hyphens, underscores, spaces, pound-signs and tabs. For each token that is three or more characters long, that token is searched for in the password; if it is present the

password change is rejected. For example, the name "Erin M. Hagens" would be split into three tokens: "Erin," "M," and "Hagens." Because the second token is only one character long, it would be ignored. Therefore, this user could not have a password that included either "erin" or "hagens" as a substring anywhere in the password. All of these checks are case insensitive.

These complexity requirements are enforced upon password change or creation of new passwords. It is recommended that you enable this setting.

Password Account Lockout
Given enough time and potential to try multiple username and password combinations an attacker might eventually succeed in compromising the security of a server or other computer. Account lockout policies allow you to set thresholds to automatically shut down an account if too many incorrect username and password combinations are attempted in order to protect the machine.

Sometimes you, or other users of a server or workstation, have a hard time remembering the correct username and password. It may be from a simple typo while entering the information or it may be a result of having too many different usernames and passwords to remember. Whatever the reason, there are times when incorrect authentication information will be entered when someone is trying to log in. You don't need to be alarmed by a single failed attempt. You probably don't even need to be concerned about two or three attempts.

At some point though you have to figure that it is no longer an honest mistake and is either a program or individual systematically trying to guess different username or password combinations to gain unauthorized access to the machine. Windows offers a way to protect the machine from such attempts through the Account Lockout Policies. By configuring the operating system to lock the account and bar access after a certain number of failed login attempts you allow the system to proactively block such attempts.

Privileges
To ease the task of user account administration, you should assign privileges primarily to group accounts, rather than to individual user accounts. When you assign privileges to a group account, users are automatically assigned those privileges when they become a member of that group. This method of administering privileges is far easier than assigning individual privileges to each user account when the account is created.

The following table lists and describes the privileges that can be granted to a user.

Privilege	Description
Act as part of the operating system	This user right allows a process to impersonate any user without authentication. The process can therefore gain access to the same local resources as that user. Processes that require this privilege should use the *LocalSystem* account, which already includes this privilege, rather than using a separate user account with this privilege specially assigned. If your organization only uses servers that are members of the Windows Server 2003 family, you do not need to assign this privilege to your users. However, if your organization uses servers running Windows 2000 or Windows NT 4.0, you might need to assign this privilege to

	use applications that exchange passwords in plaintext. **Default:** Local System.
Add workstations to a domain	This security setting determines which groups or users can add workstations to a domain. This security setting is valid only on domain controllers. By default, any authenticated user has this right and can create up to 10 computer accounts in the domain. Adding a computer account to the domain allows the computer to participate in Active Directory-based networking. For example, adding a workstation to a domain enables that workstation to recognize accounts and groups that exist in Active Directory. **Default:** Authenticated Users on domain controllers.
Adjust memory quotas for a process	This privilege determines who can change the maximum memory that can be consumed by a process. This user right is defined in the Default Domain Controller Group Policy object (GPO) and in the local security policy of workstations and servers. **Default:** Administrators.
Back up files and directories	This user right determines which users can bypass file and directory, registry, and other persistent object permissions for the purposes of backing up the system. **Default:** Administrators and Backup Operators.

Bypass traverse checking	This user right determines which users can traverse directory trees even though the user may not have permissions on the traversed directory. This privilege does not allow the user to list the contents of a directory, only to traverse directories.
	This user right is defined in the Default Domain Controller Group Policy object (GPO) and in the local security policy of workstations and servers.
	Default:
	• On workstations and servers:
	  • Administrators
	  • Backup Operators
	  • Power Users
	  • Users
	  • Everyone
	• On domain controllers:
	  • Administrators

	• Authenticated Users

Change the system time	This user right determines which users and groups can change the time and date on the internal clock of the computer. Users that are assigned this user right can affect the appearance of event logs. If the system time is changed, events that are logged will reflect this new time, not the actual time that the events occurred.
	This user right is defined in the Default Domain Controller Group Policy object (GPO) and in the local security policy of workstations and servers.
	Default:
	• On workstations and servers:
	  • Administrators
	  • Power Users
	• On domain controllers:
	  • Administrators
	  • Server Operators

Create a pagefile	Allows the user to create and change the size of a pagefile. This is done by specifying a paging file size for a particular drive under**Performance Options** on the **Advanced** tab of **System properties**. **Default setting**: Administrators
Create a token object	Allows a process to create a token which it can then use to get access to any local resources when the process uses NtCreateToken() or other token-creation APIs. It is recommended that processes requiring this privilege use the LocalSystem account, which already includes this privilege, rather than using a separate user account with this privilege specially assigned **Default setting**: No one
Create global objects	This security setting determines which accounts are allowed to create global objects in a terminal services session. **Default:** Administrators and Local System.
Create permanent shared objects	Allows a process to create a directory object in the Windows Server 2003 family and Windows XP Professional object manager. This privilege is useful to kernel-mode components that extend the object namespace. Components that are running in kernel mode already have this privilege inherently; it is not necessary to assign them the privilege. **Default setting**: No one
Debug programs	This user right determines which users can attach a debugger to any process or

	to the kernel. Developers who are debugging their own applications to not need to be assigned this user right. Developers who are debugging new system components will need this user right to be able to do so. This user right provides complete access to sensitive and critical operating system components. **Default setting**: • Administrators • Local System
Enable computer and user accounts to be trusted for delegation	This security setting determines which users can set the **Trusted for Delegation** setting on a user or computer object. The user or object that is granted this privilege must have write access to the account control flags on the user or computer object. A server process running on a computer (or under a user context) that is trusted for delegation can access resources on another computer using the delegated credentials of a client, as long as the account of the client does not have the **Account cannot be delegated** account control flag set. This user right is defined in the Default Domain Controller Group Policy object (GPO) and in the local security policy of workstations and servers. **Default setting**: On domain controllers: • Administrators
Force shutdown from a remote system	This security setting determines which users are allowed to shut down a computer from a remote location on the network. Misuse of this user right can result in a denial of service. This user right is defined in the Default Domain Controller Group Policy object (GPO) and in the local security policy of workstations and servers. **Default:** • On workstations and servers: Administrators. • On domain controllers: Administrators, Server Operators.
Generate security audits	This security setting determines which accounts can be used by a process to add entries to the security log. The security log is used to trace unauthorized system access. Misuse of this user right can result in the generation of many auditing events, potentially hiding evidence of an attack or causing a denial of service if the **Audit: Shut down system immediately if unable to log security audits** security policy setting is enabled. **Default:** Local System.
Impersonate a client after authentication	This security setting determines which accounts are allowed to impersonate other accounts. **Default:** Administrators and Service.
Increase	This security setting determines which accounts can use a process with Write

scheduling priority	property access to another process to increase the execution priority assigned to the other process. A user with this privilege can change the scheduling priority of a process through the Task Manager user interface. **Default:** Administrators.
Load and unload device drivers	This user right determines which users can dynamically load and unload device drivers or other code in to kernel mode. This user right does not apply to Plug and Play device drivers. It is recommended that you do not assign this privilege to other users. Instead, use the StartService() API. **Default setting**: Administrators. It is recommended that you not assign this privilege to any other user. Device drivers run as trusted (or highly privileged) programs.
Lock pages in memory	This security setting determines which accounts can use a process to keep data in physical memory, which prevents the system from paging the data to virtual memory on disk. Exercising this privilege could significantly affect system performance by decreasing the amount of available random access memory (RAM). **Default:** None. Certain system processes have the privilege inherently.
Manage auditing and security log	This security setting determines which users can specify object access auditing options for individual resources, such as files, Active Directory objects, and registry keys. This security setting does not allow a user to enable file and object access auditing in general. For such auditing to be enabled, theAudit object access setting in Computer Configuration\Windows Settings\Security Settings\Local Policies\Audit Policies must be configured. You can view audited events in the security log of the Event Viewer. A user with this privilege can also view and clear the security log. **Default:** Administrators.
Modify firmware environment values	This security setting determines who can modify firmware environment values. Firmware environment variables are settings stored in the nonvolatile RAM of non-x86-based computers. The effect of the setting depends on the processor. • On x86-based computers, the only firmware environment value that can be modified by assigning this user right is the Last Known Good Configuration setting, which should only be modified by the system. • On Itanium-based computers, boot information is stored in nonvolatile RAM. Users must be assigned this user right to run**bootcfg.exe** and to change the Default Operating System setting on **Startup and Recovery** in **System properties**. • On all computers, this user right is required to install or upgrade Windows. **Default setting**: • Administrators

	• Local System
profile a single process	This security setting determines which users can use performance monitoring tools to monitor the performance of nonsystem processes. **Default:** Administrators, Power users, Local System.
profile system performance	This security setting determines which users can use performance monitoring tools to monitor the performance of system processes. **Default:** Administrators, Local System.
Remove computer from docking station	This security setting determines whether a user can undock a portable computer from its docking station without logging on. If this policy is enabled, the user must log on before removing the portable computer from its docking station. If this policy is disabled, the user may remove the portable computer from its docking station without logging on. **Default:** Disabled.
Replace a process level token	Determines which user accounts can initiate a process to replace the default token associated with a started subprocess. This user right is defined in the Default Domain Controller Group Policy object and in the local security policy of workstations and servers. **Default setting**: Local Service and Network Service.
Restore files and directories	This security setting determines which users can bypass file, directory, registry, and other persistent objects permissions when restoring backed up files and directories and determines which users can set any valid security principal as the owner of an object. Specifically, this user right is similar to granting the following permissions to the user or group in question on all files and folders on the system: • Traverse Folder/Execute File • Write **Default:** • Workstations and servers: Administrators, Backup Operators. • Domain controllers: Administrators, Backup Operators, Server Operators.
Shut down the system	This security setting determines which users who are logged on locally to the computer can shut down the operating system using the **Shut Down** command. Misuse of this user right can result in a denial of service. **Default:** • Workstations: Administrators, Backup Operators, Power Users, Users. • Servers: Administrators, Backup Operators, Power Users. • Domain controllers: Account Operators, Administrators, Backup

	Operators, Server Operators, Print Operators.
Synchronize directory service data	This security setting determines which users and groups have the authority to synchronize all directory service data. This is also known as Active Directory synchronization. **Defaults:** None.
Take ownership of files or other objects	This security setting determines which users can take ownership of any securable object in the system, including Active Directory objects, files and folders, printers, registry keys, processes, and threads. **Default setting**: Administrators

User assigned privileges
Two methods of privilege assignment prevalent today are group-based and user-assigned. As the name implies, group-based privileges are those acquired as a result of belonging to a group.

User-assigned privileges are those that can be assigned by the user. For example, when you create a file in most operating systems, you can change the permissions associated with that file. It is possible, for example, to give others the privilege of only being able to read it, or to read and write to it.

User access reviews
In addition to assigning user access properly, it is important to review that access periodically. Access review is a process to determine whether a user's access level is still appropriate. People's roles within an organization can change over time. It is important to review user accounts periodically and determine if they still require the access they currently have. An example of such a scenario would be a network administrator who was responsible for the domain controller but then moved over to administer the remote access servers. The administrator's access to the domain controller should now be terminated. This concept of access review is closely related to the concept of least privileges. It is important that users do not have "leftover" privileges from previous job roles.

Continuous monitoring
Another closely related topic is continuous monitoring. Continuous monitoring implies an ongoing audit of what resources a user actually accesses. This can be critical for stop- ping insider threats. For example, a human resources employee would need access to employee files. However, if that employee is accessing a given employee's file without a valid work-related reason, this is a security breach. Only by continuously monitoring access can you detect such breaches.

Topic 6.0 Cryptography

Section 6.1- Given a scenario, utilize general cryptography concepts.

Cryptography is the science of encrypting, or hiding, information—something people have sought to do since they began using language. Although language allowed them to communicate with one another, people in power attempted to hide information by controlling who was taught to read and write. Eventually, more complicated methods of concealing information by shifting letters around to make the text unreadable were developed.

The Romans typically used a different method known as a shift cipher. In this case, one letter of the alphabet is shifted a set number of places in the alphabet for another letter. A common modern-day example of this is the ROT13 cipher, in which every letter is rotated 13 positions in the alphabet: n is written instead of a, o instead of b, and so on. These ciphers were simple to use and also simple to break. Because hiding information was still important, more advanced transposition and substitution ciphers were required. As systems and technology became more complex, some mechanical or electromechanical device frequently automated ciphers. A famous example of a modern encryption machine is the German Enigma machine from World War II. This machine used a complex series of substitutions to perform encryption, and interestingly enough it gave rise to extensive research in computers.

Cryptanalysis, the process of analyzing available information in an attempt to return the encrypted message to its original form, required advances in computer technology for complex encryption methods. The birth of the computer made it possible to easily execute the calculations required by more complex encryption algorithms. Today, the computer almost exclusively powers how encryption is performed. Computer technology has also aided cryptanalysis, allowing new methods to be developed, such as linear and differential cryptanalysis. Comparing the input plaintext to the output ciphertext to try and determine the key used to encrypt the information does differential cryptanalysis. Linear cryptanalysis is similar in that it uses both plaintext and ciphertext, but it puts the plaintext through a simplified cipher to try and deduce what the key is likely to be in the full version of the cipher.

Symmetric Encryption Summary

Symmetric algorithms are important because they are comparatively fast and have few computational requirements. Their main weakness is that two geographically distant parties both need to have a key that matches exactly. In the past, keys could be much simpler and still be secure, but with today's computational power, simple keys can be brute-forced very quickly. This means that larger and more complex keys must be used and exchanged. This key exchange is difficult because the key cannot be simple, such as a word, but must be shared in a secure manner. It might be easy to exchange a 4-bit key such as b in hex, but exchanging the 128-bit key 4b36402c5727472d5571373d22675b4b is far more difficult to do securely. This exchange of keys is greatly facilitated by our next subject, asymmetric, or public key, cryptography.

Asymmetric Encryption

Asymmetric cryptography is in many ways completely different than symmetric cryptography. While both are used to keep data from being seen by unauthorized users,

asymmetric cryptography uses two keys instead of one. Whitfield Diffie and Martin Hellman invented it in 1975. Asymmetric cryptography is more commonly known as public key cryptography. The system uses a pair of keys: a private key that is kept secret and a public key that can be sent to anyone. The system's security relies upon resistance to deducing one key, given the other, and thus retrieving the plaintext from the ciphertext.

Fundamental differences and encryption methods

Encryption can be a relatively simple process, or as difficult as the user wants to make it. The degree of difficulty does not necessarily relate to the security of the encryption method.

It is of the utmost importance that the user understands, at a bare minimum, the principals of encryption. Otherwise, he places both himself and those to whom he sends messages at risk.

Beware of the snake oil salesmen. It seems that every week, there's someone hawking a "new, revolutionary, military grade, unbreakable" encryption algorithm or method. "Caveat Emptor" with such claims. A knowledgeable user will be much more likely to weed through all of the snake oil pitches.

The defacto standard for encryption is PGP. There are those who will argue this point for a variety of reasons, but the fact remains that PGP is the most widely used and supported, and most readily available encryption method.

PGP is available for almost every operating system, with a variety of versions for each. The features and functionality of each version should help determine which is best for you. The newer versions of PGP include plugins for popular email clients, and some include desktop security features as well.

Block vs. stream

Stream Ciphers vs. Block Ciphers

Unlike what we've seen, private-key (aka symmetric) encryption schemes used in practice generally

- are not be based on nice computational problems
- are not proven secure via reductions
- are designed for a particular input length (so can only be treated with concrete security)
- but are extremely e-cient

Stream Ciphers

- Essentially meant to be pseudorandom generators, used for stateful encryption.
- Examples: linear feedback shift registers (not secure, but used as component in better
- stream ciphers), RC4, SEAL, ...
- Extremely simple and fast
- Practical issues: can generate pseudorandom bits oine and decrypt very quickly without

- buffering, but requires synchronization

Block ciphers

- For every key k 2 f0; 1g n, Ek : f0; 1g` ! f0; 1g is a permutation, and both Ek and E ¡1 k can be computed quickly given k. (n=key length, ` = block length)
- Examples: DES, AES/Rijndael, IDEA, ...
- Main tools for private-key encryption in practice.
- Have both stateless modes and stateful/stream-like mode**s**

Non-repudiation
An item of some confusion, the concept of nonrepudiation is actually fairly simple. Nonrepudiation means that the message sender cannot later deny that she sent the message. This is important in electronic exchanges of data, because of the lack of face to face meetings. Nonrepudiation is based upon public key cryptography and the principle of only you knowing your private key. The presence of a message signed by you, using your private key, which nobody else should know, is an example of nonrepudiation.

When a third party can check your signature using your public key, that disproves any claim that you were not the one who actually sent the message. Nonrepudiation is tied to asymmetric cryptography and cannot be implemented with symmetric algorithms.

Hashing
Hashing functions are commonly used encryption methods. A hashing function is a special mathematical function that performs one-way encryption, which means that once the algorithm is processed, there is no feasible way to use the ciphertext to retrieve the plaintext that was used to generate it. Also, ideally, there is no feasible way to generate two different plaintexts that compute to the same hash value.

Common uses of hashing functions are storing computer passwords and ensuring message integrity. The idea is that hashing can produce a unique value that corresponds to the data entered, but the hash value is also reproducible by anyone else running the developed in 1993, was designed as the algorithm to be used for secure hashing in the U.S. Digital Signature Standard (DSS). It is modeled on the MD4 algorithm and implements fixes in that algorithm discovered by the NSA. It creates message digests 160 bits long that can be used by the Digital Signature Algorithm (DSA), which can then compute the signature of the message. This is computationally simpler, as the message digest is typically much smaller than the actual message—smaller message, less work.

SHA-1 works, as do all hashing functions, by applying a compression function to the data input. It accepts an input of up to 264 bits or less and then compresses down to a hash of 160 bits. SHA-1 works in block mode, separating the data into words first, and then grouping the words into blocks. The words are 32-bit strings converted to hex; grouped together as 16 words, they make up a 512-bit block. If the data that is input to SHA-1 is not a multiple of 512, the message is padded with zeros and an integer describing the original length of the message.

At one time, SHA-1 was one of the more secure hash functions, but it has been found

vulnerable to a collision attack. Thus, most people are suggesting that implementations of SHA-1 be moved to one of the other SHA versions. These longer versions, SHA-256, SHA-384, and SHA-512, all have longer hash results, making them more difficult to attack successfully. The added security and resistance to attack in SHA-1 does require more processing power to compute the hash.

Key escrow

The impressive growth of the use of encryption technology has led to new methods for handling keys. Encryption is adept at hiding secrets, and with computer technology being affordable to everyone, criminals and other ill-willed people began using it to conceal communications and business dealings from law enforcement agencies. Because they could not break the encryption, government agencies began asking for key escrow. Key escrow is a system by which your private key is kept both by you and by the government. This allows people with a court order to retrieve your private key to gain access to anything encrypted with your public key. Your key and the government key, giving the government access to your plaintext data, essentially encrypt the data.

Steganography

Steganography, an offshoot of cryptography technology, gets its meaning from the Greek steganos meaning covered. Invisible ink placed on a document hidden by innocuous text is an example of a steganographic message. Another example is a tattoo placed on the top of a person's head, visible only when the person's hair is shaved off. Hidden writing in the computer age relies on a program to hide data inside other data. The most common application is the concealing of a text message in a picture file. The Internet contains multiple billions of image files, allowing a hidden message to be located almost anywhere without being discovered. The nature of the image files also makes a hidden message difficult to detect. While it is most common to hide messages inside images, they can also be hidden in video and audio files.

Steganographic encoding can be used in many ways and through many different media. LSB, Least Significant Bit, is a method of encoding information into an image while altering the actual visual image as little as possible. A computer image is made up of thousands or millions of pixels, all defined by 1s and 0s. If an image is composed of Red Green Blue (RGB) values, each pixel has an RGB value represented numerically from 0 to 255. For example, 0,0,0 is black, and 255,255,255 is white, which can also be represented as 00000000, 00000000, 00000000 for black and 11111111, 11111111, 11111111 for white. Given a white pixel, editing the least significant bit of the pixel to 11111110, 11111110, 11111110 changes the color. The change in color is undetectable to the human eye, but in a image with a million pixels, this creates a 125KB area in which to store a message.

Digital signatures

Digital signatures have been touted as the key to truly paperless document flow, and they do have promise for improving the system. Digital signatures are based on both hashing functions and asymmetric cryptography. Both encryption methods play an important role in signing digital documents.

Unprotected digital documents are very easy for anyone to change. If a document is edited after an individual signs it, it is important that any modification can be detected. To protect

against document editing, hashing both parties use functions to create a digest of the message that is unique and easily reproducible. This ensures that the message integrity is complete.

Elliptic curve and quantum cryptography
Elliptic curve cryptography (ECC) works on the basis of elliptic curves. An elliptic curve is a simple function that is drawn as a gently looping curve on the X,Y plane. They are defined by this equation:

$y2 = x3 + ax2 + b$

Elliptic curves work because they have a special property—you can add two points on the curve together and get a third point on the curve. For cryptography, the elliptic curve works as a public key algorithm. Users agree on an elliptic curve and a fixed curve point. This information is not a shared secret, and these points can be made public without compromising the security of the system. User 1 then chooses a secret random number, K1, and computes a public key based upon a point on the curve:

$P1 = K1 * F$

User 2 performs the same function and generates P2. Now user 1 can send user 2 a message by generating a shared secret:

$S = K1 * P2$

User 2 can generate the same shared secret independently:

$S = K2 * P1$

This is true because

$K1 * P2 = K1 * (K2 * F) = (K1 * K2) * F = K2 * (K1 * F) = K2 * P1$

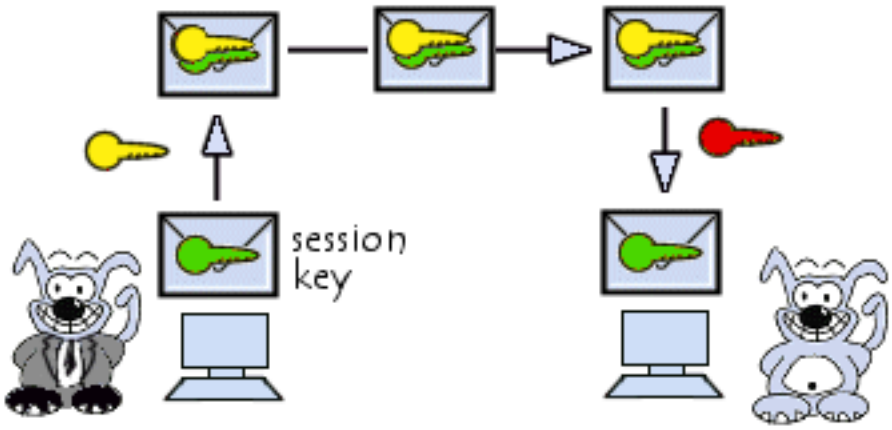
The security of elliptic curve systems has been questioned, mostly because of lack of analysis. However, all public key systems rely on the difficulty of certain math problems. It would take a breakthrough in math for any of the mentioned systems to be weakened dramatically, but research has been done about the problems and has shown that the elliptic curve problem has been more resistant to incremental advances. Again, as with all cryptography algorithms, only time will tell how secure they really are.

Cryptography Algorithm Use
The use of cryptographic algorithms grows every day. The best way to do that with current technology is to use encryption. Security is typically defined as a product of five components: confidentiality, integrity, availability, authentication, and nonrepudiation. Encryption addresses four of these five components: confidentiality, integrity, nonrepudiation, and authentication.

Session keys

The concept of session keys is a compromise between symmetric and asymmetric encryption that makes it possible to combine the two techniques.

The principle of session keys is simple: it involves randomly generating a reasonably sized session key and encrypting this key using a public-key encryption algorithm (more precisely, using the recipient's public key).



The recipient is capable of decrypting the session key with his private key. The sender and recipient are in possession of a shared key that only they know. They can therefore send each other encrypted documents using a symmetric encryption algorithm.

In-band vs. out-of-band key exchange

Key exchange is an important topic in relation to symmetric cryptography. There are two primary approaches to key exchange: in-band key exchange and out-of-band key exchange. In-band key exchange essentially means that the key is exchanged within the same communications channel that is going to be encrypted. Out-of-band key exchange means that some other channel, other than the one that is going to be secured, is used to exchange the key.

Forward secrecy is a property of any key exchange system, which ensures that if one key is compromised, subsequent keys will not also be compromised. Perfect forward secrecy occurs when this process is unbreakable.

Ephemeral key
Taher Elgamal developed ElGamal in 1984. It is an asymmetric algorithm, and several variations of ElGamal have been created, including Elliptic Curve ElGamal. ElGamal and related algorithms use what is called an ephemeral key. An ephemeral key is simply a key that exists only for that session. Essentially, the algorithm creates a key to use for that single communication session and it is not used again.

Adding an ephemeral key to Diffie-Hellman turns it into DHE (which, despite the order of the acronym, stands for Ephemeral Diffie-Hellman). Adding an ephemeral key to Elliptic Curve Diffie-Hellman turns it into ECDHE (again, overlook the order of the acronym letters, it is called Ephem- eral Elliptic Curve Diffie-Hellman). It is the ephemeral component of each of these that provides the perfect forward secrecy.

Perfect forward secrecy
Adding an ephemeral key to Diffie-Hellman turns it into *DHE* (which, despite the order of the acronym, stands for Ephemeral Diffie-Hellman). Adding an ephemeral key to Elliptic Curve Diffie-Hellman turns it into *ECDHE* (again, overlook the order of the acronym letters, it is called Ephem- eral Elliptic Curve Diffie-Hellman). It is the ephemeral component of each of these that provides the perfect forward secrecy.

Section 6.2- Given a scenario, use appropriate cryptographic methods.

WEP vs. WPA/WPA2 and pre-shared key
Wired Equivalent Privacy (WEP) was intended to provide basic security for wireless networks, whereas wireless systems frequently use the Wireless Application Protocol (WAP) for network communications. Over time, WPA and WPA2 have replaced WEP in most implementations.

Wired Equivalent Privacy
Wired Equivalent Privacy (WEP) is a wireless protocol designed to provide a privacy equivalent to that of a wired network. WEP was implemented in a number of wireless devices, including smartphones and other mobile devices. WEP was vulnerable because of weak- nesses in the way its encryption algorithms (RC4) are employed. These weaknesses allowed the algorithm to be cracked potentially in as little as five minutes using available PC soft- ware. This made WEP one of the more vulnerable security protocols.

As an example, the initialization vector (IV) that WEP uses for encryption is 24-bit, which is quite weak and means that IVs are reused with the same key. By examining the repeating result, it was easy for attackers to crack the WEP secret key. This is known as an IV attack.

Since the IV is shorter than the key, it must be repeated when used. To put it in perspective, the attack happened because the algorithm used is RC4, the IV is too small, the IV is static, and the IV is part of the RC4 encryption key.

MD5
MD5 was developed in 1991 and is structured after MD4 but with additional security to overcome the problems in MD4. Therefore, it is very similar to the MD4 algorithm, only slightly slower and more secure. MD5 creates a 128-bit hash of a message of any length. Like MD4, it segments the message into 512-bit blocks and then into sixteen 32-bit words. First, the original message is padded to be 64 bits short of a multiple of 512 bits. Then a 64-bit representation of the original length of the message is added to the padded value to bring the entire message up to a 512-bit multiple.
SHA-256
SHA-256 is similar to SHA-1, in that it will also accept input of less than 264 bits and reduces that input to a hash. This algorithm reduces to 256 bits instead of SHA-1's 160. Defined in FIPS 180-2 in 2002, SHA-256 is listed as an update to the original FIPS 180 that defined SHA. Similar to SHA-1, SHA-256 will accept 264 bits of input and uses 32-bit words and 512-bit blocks. Padding is added until the entire message is a multiple of 512. SHA-256 uses sixty-four 32-bit words, eight working variables, and results in a hash value of eight 32-

bit words, hence 256 bits. SHA-256 is more secure than SHA-1, but the attack basis for SHA-1 can produce collisions in SHA-256 as well since they are similar algorithms. The SHA standard does have two longer versions, however.

SHA-384
SHA-384 is also similar to SHA-1, but it handles larger sets of data. SHA-384 will accept 2128 bits of input, which it pads until it has several blocks of data at 1024-bit blocks. SHA-384 also used 64-bit words instead of SHA-1's 32-bit words. It uses six 64-bit words to produce the 284-bit hash value.

SHA-512
SHA-512 is structurally similar to SHA-384. It will accept the same 2128 input and uses the same 64-bit word size and 1024-bit block size. SHA-512 does differ from SHA-384 in that it uses eight 64-bit words for the final hash, resulting in 512 bits.

RIPEMD
RIPEMD is a cryptographic hash based upon MD4. It's been shown to have weaknesses and has been replaced by RIPEMD-128 and RIPMD-160. These are cryptographic hash functions designed by Hans Dobbertin, Antoon Bosselaers, and Bart Preneel.
The name comes from the project they were designed for: EU project RIPE (RACE Integrity Primitives Evaluation, 1988-1992).
RIPEMD-160 produces a hash of the same length as SHA1 but is slightly slower. RIPEMD-128 has been designed as a drop-in replacement for MD4/MD5 whilst avoiding some of the weaknesses shown for these two algorithms. (These are possible hash collisions.) It is about half the speed of MD5.
Implementations for both these hashing functions are present in Trf, and a tcllib module provides a pure-Tcl version if required. The tcllib version will use Trf if it is available.
AES
Because of the advancement of technology and the progress being made in quickly retrieving DES keys, NIST put out a request for proposals for a new Advanced Encryption Standard (AES). It called for a block cipher using symmetric key cryptography and supporting key sizes of 128, 192, and 256 bits. After evaluation, the NIST had five finalists:

- **MARS** IBM
- **RC6** RSA
- **Rijndael** John Daemen and Vincent Rijmen
- **Serpent** Ross Anderson, Eli Biham, and Lars Knudsen
- **Twofish** Bruce Schneier, John Kelsey, Doug Whiting, David Wagner, Chris Hall, and Niels Ferguson

In the fall of 2000, NIST picked Rijndael to be the new AES. It was chosen for its overall security as well as its good performance on limited capacity devices. Rijndael's design was influenced by Square, also written by John Daemen and Vincent Rijmen. Like Square, Rijndael is a block cipher separating data input in 128-bit blocks. Rijndael can also be configured to use blocks of 192 or 256 bits, but AES has standardized on 128-bit blocks. AES can have key sizes of 128, 192, and 256 bits, with the size of the key affecting the number of rounds used in the algorithm. Like DES, AES works in three steps on every block of input data:

1. Add round key, performing an XOR of the block with a subkey.
2. Perform the number of normal rounds required by the key length.
3. Perform a regular round without the mix-column step found in the normal round.

After these steps have been performed, a 128-bit block of plaintext produes a 128-bit block of ciphertext. As mentioned in step 2, AES performs multiple rounds. This is determined by the key size. A key size of 128 bits requires 9 rounds, 192-bit keys will re quire 11 rounds, and 256-bit keys use 13 rounds. Four steps are performed in every round:

1. **Byte sub**. Each byte is replaced by its S-box substitute.
2. **Shift row**. Bytes are arranged in a rectangle and shifted.
3. **Mix column**. Matrix multiplication is performed based upon the arranged rectangle.
4. **Add round key**. This round's subkey is cored in.

These steps are performed until the final round has been completed, and when the final step has been performed, the ciphertext is output.

DES

DES, the Data Encryption Standard, was developed in response to the National Bureau of Standards (NBS), now known as the National Institute of Standards and Technology (NIST), issuing a request for proposals for a standard cryptographic algorithm in 1973. NBS received a promising response in an algorithm called Lucifer, originally developed by IBM. The NBS and the NSA worked together to analyze the algorithm's security, and eventually DES was adopted as a federal standard in 1976.

NBS specified that the DES standard had to be recertified every five years. While DES passed without a hitch in 1983, the NSA said it would not recertify it in 1987. However, since no alternative was available for many businesses, many complaints ensued, and the NSA and NBS were forced to recertify it. The algorithm was then recertified in 1993. NIST has now certified the Advanced Encryption Standard (AES) to replace DES.

DES is what is known as a block cipher; it segments the input data into blocks of a specified size, typically padding the last block to make it a multiple of the block size required. In the case of DES, the block size is 64 bits, which means DES takes a 64-bit input and outputs 64 bits of ciphertext. This process is repeated for all 64-bit blocks in the message. DES uses a key length of 56 bits, and all security rests within the key. The same algorithm and key are used for both encryption and decryption.

3DES

Triple DES (3DES) is a variant of DES. Depending on the specific variant, it uses either two or three keys instead of the single key that DES uses. It also spins through the DES algorithm three times via what's called multiple encryption. Multiple encryptions can be performed in several different ways. The simplest method of multiple encryption is just to stack algorithms on top of each other—taking plaintext, encrypting it with DES, then encrypting the first ciphertext with a different key, and then encrypting the second ciphertext with a third key. In reality, this technique is less effective than the technique that 3DES uses, which is to encrypt with one key, then decrypt with a second, and then encrypt with a third.

HMAC

What is HMAC? HMAC is merely a specific type of MAC function. It works by using an underlying hash function over a message and a key. It is currently one of the predominant means to ensure that secure data is not corrupted in transit over unsecure channels (like the internet).

Any hashing functions could be used with HMAC, although more secure hashing functions are preferable. An example of a secure hash function (which is commonly used in HMAC implementations) is SHA-14. (Other common hashing functions include MD5 and RIPEND-160). As computers become more and more powerful, increasingly complex hash functions will probably be used. Furthermore, there are several generations of SHA hashing functions (SHA-256, SHA-384, and SHA-512) which are currently available but not very widely used as their added security is not yet believed to be needed in everyday transactions.

RSA

RSA is one of the first public key cryptosystems ever invented. It can be used for both encryption and digital signatures. RSA is named after its inventors, Ron Rivest, Adi Shamir, and Leonard Adleman, and was first published in 1977. This algorithm uses the product of two very large prime numbers and works on the principle of difficulty in factoring such large numbers. It's best to choose large prime numbers from 100 to 200 digits in length and that are equal in length. These two primes will be P and Q. Randomly choose an encryption key, E, so that E is greater than 1, E is less than P * Q, and E must be odd. E must also be relatively prime to (P – 1) and (Q – 1). Then compute the decryption key D:

$D = E^{-1} \bmod ((P - 1)(Q - 1))$

Now that the encryption key and decryption key have been generated, the two prime numbers can be discarded, but they should not be revealed. To encrypt a message, it should be divided into blocks less than the product of P and Q. Then,

$Ci = Mi^{E} \bmod (P * Q)$

C is the output block of ciphertext matching the block length of the input message, M. To decrypt a message take ciphertext, C, and use this function:

$Mi = Ci^{D} \bmod (P * Q)$

The use of the second key retrieves the plaintext of the message.
This is a simple function, but its security has withstood the test of more than 20 years of analysis. Considering the effectiveness of RSA's security and the ability to have two keys, why are symmetric encryption algorithms needed at all? The answer is speed. RSA in software can be 100 times slower than DES, and in hardware it can be even slower.

RC4

RC4 was created before RC5 and RC6, but it differs in operation. RC4 is a stream cipher;whereas all the symmetric ciphers we have looked at so far have been block-mode

ciphers.A stream-mode cipher works by enciphering the plaintext in a stream, usually bit by bit.This makes stream ciphers faster than block-mode ciphers. Stream ciphers accomplishthis by performing a bitwise XOR with the plaintext stream and a generated key strea .

One-time-pads

One-time pad (OTP), also called Vernam-cipher or the perfect cipher, is a crypto algorithm where plaintext is combined with a random key. It is the only known method to perform mathematically unbreakable encryption. Used by Special Operations teams and resistance groups in WW2, popular with intelligence agencies and their spies during the Cold War and beyond, the one-time pad gained a reputation as a simple but solid encryption system in the intriguing world of intelligence, espionage, covert operations and military and diplomatic communications, with an absolute security which is unmatched by modern crypto algorithms.

We can only talk about one-time pad if four important rules are followed. If these rules are applied correctly, the one-time pad can be proven unbreakable (see Claude Shannon's "Communication Theory of Secrecy Systems"). Even infinite computational power and infinite time cannot break one-time pad encryption, simply because it is mathematically impossible. However, if only one of these rules is disregarded, the cipher is no longer unbreakable.

- The key is as long as the message or data that is encrypted.
- The key is truly random (not generated by a simple computer function or such)
- Each key is used only once, and both sender and receiver must destroy their key after use.
- There should only be two copies of the key: one for the sender and one for the receiver (some exceptions exist for multiple receivers)

Important note: one-time pads or one-time encryption is not to be confused with one-time keys (OTK) or one-time passwords (sometimes also denoted as OTP). Such one-time keys, limited in size, are only valid for a single encryption session by some crypto-algorithm under control of that key. Small one-time keys are by no means unbreakable, because the security of the encryption depends on the crypto algorithm they are used for.

CHAP

CHAP is used to periodically verify the identity of the peer using a 3-way handshake. This is done upon initial link establishment, and MAY be repeated anytime after the link has been established.

1. After the Link Establishment phase is complete, the authenticator sends a "challenge" message to the peer.
2. The peer responds with a value calculated using a "one-way hash" function.
3. The authenticator checks the response against its own calculation of the expected hash value. If the values match, the authentication is acknowledged; otherwise the connection SHOULD be terminated.
4. At random intervals, the authenticator sends a new challenge to the peer, and repeats steps 1 to 3.

CHAP provides protection against playback attack by the peer through the use of an incrementally changing identifier and a variable challenge value. The use of repeated challenges is intended to limit the time of exposure to any single attack. The authenticator is in control of the frequency and timing of the challenges. This authentication method depends upon a "secret" known only to the authenticator and that peer. The secret is not sent over the link. Although the authentication is only one-way, by negotiating CHAP in both directions the same secret set may easily be used for mutual authentication. Since CHAP may be used to authenticate many different systems, name fields may be used as an index to locate the proper secret in a large table of secrets. This also makes it possible to support more than one name/secret pair per system, and to change the secret in use at any time during the session.

CHAP requires that the secret be available in plaintext form. Irreversably encrypted password databases commonly available cannot be used. It is not as useful for large installations, since every possible secret is maintained at both ends of the link.

The Challenge packet is used to begin the Challenge-Handshake Authentication Protocol. The authenticator MUST transmit a CHAP packet with the Code field set to 1 (Challenge). Additional Challenge packets MUST be sent until a valid Response packet is received, or an optional retry counter expires.

A Challenge packet MAY also be transmitted at any time during the Network-Layer Protocol phase to ensure that the connection has not been altered.

The peer SHOULD expect Challenge packets during the Authentication phase and the Network-Layer Protocol phase.

Whenever a Challenge packet is received, the peer MUST transmit a CHAP packet with the Code field set to 2 (Response). Whenever a Response packet is received, the authenticator compares the Response Value with its own calculation of the expected value. Based on this comparison, the authenticator MUST send a Success or Failure packet.

PAP

PAP can authenticate an identity and password for a peer resulting in success or failure.

PAP provides a simple method for the peer to establish its identity using a 2-way handshake. This is done only upon initial link establishment.

After the Link Establishment phase is complete, an Id/Password pair is repeatedly sent by the peer to the authenticator until authentication is acknowledged or the connection is terminated.

PAP is not a strong authentication method. Passwords are sent over the circuit "in the clear", and there is no protection from playback or repeated trial and error attacks. The peer is in control of the frequency and timing of the attempts.

Any implementations which include a stronger authentication method (such as CHAP) MUST offer to negotiate that method prior to PAP.

This authentication method is most appropriately used where a plaintext password must be

available to simulate a login at a remote host. In such use, this method provides a similar level of security to the usual user login at the remote host.

PAP Configuration Options:

00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
Option								Length							
Data :::															

Option. 8 bits.

Option	Length	Description
3	4	Authentication-Protocol.

Length. 8 bits.

Data. Variable length.

NTLM

NTLM is a suite of authentication and session security protocols used in various Microsoft network protocol implementations and supported by the NTLM Security Support Provider ("NTLMSSP"). Originally used for authentication and negotiation of secure DCE/RPC, NTLM is also used throughout Microsoft's systems as an integrated single sign-on mechanism. It is probably best recognized as part of the "Integrated Windows Authentication" stack for HTTP authentication; however, it is also used in Microsoft implementations of SMTP, POP3, IMAP (all part of Exchange), CIFS/SMB, Telnet, SIP, and possibly others.

The NTLM Security Support Provider provides authentication, integrity, and confidentiality services within the Window Security Support Provider Interface (SSPI) framework. SSPI specifies a core set of security functionality that is implemented by supporting providers; the NTLMSSP is such a provider. The SSPI specifies, and the NTLMSSP implements, the following core operations:

1. Authentication -- NTLM provides a challenge-response authentication mechanism, in which clients are able to prove their identities without sending a password to the server.
2. Signing -- The NTLMSSP provides a means of applying a digital "signature" to a message. This ensures that the signed message has not been modified (either accidentally or intentionally) and that that signing party has knowledge of a shared secret. NTLM implements a symmetric signature scheme (Message Authentication Code, or MAC); that is, a valid signature can only be generated and verified by parties that possess the common shared key.
3. Sealing -- The NTLMSSP implements a symmetric-key encryption mechanism, which provides message confidentiality. In the case of NTLM, sealing also implies

signing (a signed message is not necessarily sealed, but all sealed messages are signed).

NTLM has been largely supplanted by Kerberos as the authentication protocol of choice for domain-based scenarios. However, Kerberos is a trusted-third-party scheme, and cannot be used in situations where no trusted third party exists; for example, member servers (servers that are not part of a domain), local accounts, and authentication to resources in an untrusted domain. In such scenarios, NTLM continues to be the primary authentication mechanism (and likely will be for a long time).

Blowfish

Blowfish is a keyed, symmetric cryptographic block cipher designed by Bruce Schneier in 1993 and placed in the public domain. Blowfish is included in a large number of cipher suites and encryption products, including SplashID. Blowfish's security has been extensively tested and proven. As a public domain cipher, Blowfish has been subject to a significant amount of cryptanalysis, and full Blowfish encryption has never been broken. Blowfish is also one of the fastest block ciphers in public use, making it ideal for a product like SplashID those functions on a wide variety of processors found in mobile phones as well as in notebook and desktop computers.

Schneier designed Blowfish as a general-purpose algorithm, intended as a replacement for the aging DES and free of the problems associated with other algorithms.

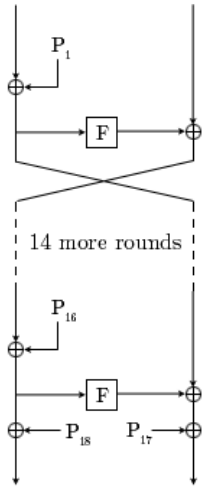
Notable features of the design include key-dependent S-boxes and a highly complex key schedule.

The Blowfish algorithm

Blowfish has a 64-bit block size and a key length of anywhere from 32 bits to 448 bits. It is a 16-round Feistel cipher and uses large key-dependent S-boxes. It is similar in structure to CAST-128, which uses fixed S-boxes.

The diagram below shows the action of Blowfish. Each line represents 32 bits. The algorithm keeps two subkey arrays: the 18-entry P-array and four 256-entry S-boxes. The S-boxes accept 8-bit input and produce 32-bit output. One entry of the P-array is used every round, and after the final round, each half of the data block is XORed with one of the two remaining unusedP-entries.

The diagram below shows Blowfish's F-function. The function splits the 32-bit input into four eight-bit quarters, and uses the quarters as input to the S-boxes. The outputs are added modulo 232 and XORed to produce the final 32-bit output.



Since Blowfish is a Feistel network, it can be inverted simply by XORing P17 and P18 to the ciphertext block, then using the P-entries in reverse order.

Blowfish's key schedule starts by initializing the P-array and S-boxes with values derived from the hexadecimal digits of pi, which contain no obvious pattern. The secret key is then XORed with the P-entries in order (cycling the key if necessary). A 64-bit all-zero block is then encrypted with the algorithm as it stands. The resultant ciphertext replaces P1 and P2. The ciphertext is then encrypted again with the new subkeys, and P3 and P4 are replaced by the new ciphertext. This continues, replacing the entire P-array and all the S-box entries. In all, the Blowfish encryption algorithm will run 521 times to generate all the subkeys - about 4KB of data is processed.

PGP/GPG

PGP/GPG are tools for encrypting and signing files and e-mail messages. Encrypting to prevent others for being able to view it, signing so that the receiver can be certain from whom the original file came from and that it did not alter**.**

Whole disk encryption

Full Disk Encryption or Whole Disk Encryption is a phrase that was coined by Seagate to describe their encrypting hard drive. Under such a system, the entire contents of a hard drive are encrypted. This is different from Full Volume Encryptionwhere only certain partitions are encrypted.

It is a kind of disk encryption software or hardware, which encrypts every bit of data that goes onto a disk or disk volume. The term "full disk encryption" is often used to signify that everything on a disk is encrypted, including the programs that can encrypt bootable operating system partitions. Full-disk encrypt is in contrast to Filesystem-level encryption, which is a form of disk encryption where individual files or directories are encrypted by the Filesystem itself. The enterprise-class full-disk encryption for Linux goes by the name dm-crypt. There is also an extension to it called LUKS (Linux Unified Key Setup)which enables us to do fancy things like key management for example. dm-crypt and LUKS, both are free-software working together in order to provide data encryption on storage media thus allowing that

what is a secret stays a secret. dm-crypt is a device-mapper and part of the Linux operating system kernel. LUKS is a hard disk encryption specification, represented by cryptsetup, its actual implementation.

TwoFish
Twofish is a 128-bit block cipher that accepts a variable-length key up to 256 bits. The cipher is a 16-round Feistel network with a bijective F function made up of four key-dependent 8-by-8-bit S-boxes, a fixed 4-by-4 maximum distance separable matrix over GF(2^8), a pseudo-Hadamard transform, bitwise rotations, and a carefully designed key schedule. A fully optimized implementation of Twofish encrypts on a Pentium Pro at 17.8 clock cycles per byte, and an 8-bit smart card implementation encrypts at 1820 clock cycles per byte. Twofish can be implemented in hardware in 14000 gates. The design of both the round function and the key schedule permits a wide variety of tradeoffs between speed, software size, key setup time, gate count, and memory. We have extensively cryptanalyzed Twofish; our best attack breaks 5 rounds with $2^{22.5}$ chosen plaintexts and 2^{51} effort.

Twofish can:

- Encrypt data at 285 clock cycles per block on a Pentium Pro, after a 12700 clock-cycle key setup.
- Encrypt data at 860 clock cycles per block on a Pentium Pro, after a 1250 clock-cycle key setup.
- Encrypt data at 26500 clock cycles per block on a 6805 smart card, after a 1750 clock-cycle key setup.

Diffie-Hellman
Whitfield Diffie and Martin Hellman conceptualized the Diffie-Hellman key exchange. They are considered the founders of the public/private key concept; their original work envisioned splitting the key into two parts. This algorithm is used primarily to send keys across public networks. The process isn't used to encrypt or decrypt messages; it's used merely for the creation of a symmetric key between two parties.

An interesting twist is that Malcolm J. Williamson of the British Intelligence Service had actually developed the method a few years earlier, but it was classified.
On the Security+ exam, if you are asked about an algorithm for exchanging keys over an insecure medium, unless the context is IPSec, the answer is always Diffie-Hellman.

dhe
Adding an ephemeral key to Diffie-Hellman turns it into DHE (which, despite the order of the acronym, stands for Ephemeral Diffie-Hellman). Adding an ephemeral key to Elliptic Curve Diffie-Hellman turns it into ECDHE (again, overlook the order of the acronym letters, it is called Ephem- eral Elliptic Curve Diffie-Hellman). It is the ephemeral component of each of these that provides the perfect forward secrecy.

Use of algorithms/protocols with transport encryption

SSL and TLS
Secure Sockets Layer (SSL) is used to establish a secure communication connection between two TCP-based machines. This protocol uses the handshake method of establishing a session. The number of steps in the handshake depends on whether steps are combined and/or mutual authentication is included. The number of steps is always between four and nine, inclusive, based on who is doing the documentation.
One of the early steps will always be to select an appropriate cipher suite to use. A cipher suite is a combination of methods, such as an authentication, encryption, and message authentication code (MAC) algorithms used together. Many cryptographic protocols such as TLS use a cipher suite.

Netscape originally developed the SSL method, which has gained wide acceptance throughout the industry. SSL establishes a session using asymmetric encryption and maintains the session using symmetric encryption.

IPSec
IP Security (IPSec) is a security protocol that provides authentication and encryption across the Internet. IPSec is becoming a standard for encrypting virtual private network (VPN) channels and is built into IPv6. It's available on most network platforms, and it's considered to be highly secure.

One of the primary uses of IPSec is to create VPNs. IPSec, in conjunction with Layer 2 Tunneling Protocol (L2TP) or Layer 2 Forwarding (L2F), creates packets that are difficult to read if intercepted by a third party. IPSec works at layer 3 of the OSI model.
The two primary protocols used by IPSec are Authentication Header (AH) and Encapsulating Security Payload (ESP). Both can operate in either the transport or tunnel mode. Protocol 50 is used for ESP, while protocol 51 is used for AH.

SSH
Secure Shell (SSH) is a tunneling protocol originally used on Unix systems. It's now available for both Unix and Windows environments. The handshake process between the client and server is similar to the process described in SSL. SSH is primarily intended for interactive terminal sessions.

HTTPS
Hypertext Transport Protocol over SSL (HTTPS), also known as Hypertext Transport Protocol Secure, is the secure version of HTTP, the language of the World Wide Web. HTTPS uses SSL to secure the channel between the client and server. Many e-business systems use HTTPS for secure transactions. An HTTPS session is identified by the https in the URL and by a key that is displayed in the web browser

Cipher suites
The system may be considered weak if it allows weak keys, has defects in its design, or is easily decrypted. Many systems available today are more than adequate for business and personal use, but they are inadequate for sensitive military or governmental applications.

Cipher suites, for example, work with SSL/TLS to combine authentication, encryption, and message authentication. Most vendors allow you to set cipher suite preferences on a server to determine the level of strength required by client connections. With Sybase, for example, you

set the cipher suite preference to Weak, Strong, FIPS, or All. If you choose Strong, you are limiting the choices to only encryption algorithms that use keys of 64 bits or more. Choosing Weak adds all the encryption algorithms that are less than 64 bits, while choosing FIPS requires encryptions, hash and key exchange algorithms to be FIPS- compliant (AES, 3DES, DES, and SHA1). Apache offers similar choices but instead of the words Strong and Weak, the names are changed to High, Medium, and Low.

Key stretching
Key stretching refers to processes used to take a key that might be a bit weak and make it stronger, usually by making it longer. The key (or password/passphrase) is input into an algorithm that will strengthen the key and make it longer, thus less susceptible to brute-force attacks. There are many methods for doing this; two are discussed here:

PBKDF2 PBKDF2 (Password-Based Key Derivation Function 2) is part of PKCS #5 v. 2.01. It applies some function (like a hash or HMAC) to the password or passphrase along with Salt to produce a derived key.

Bcrypt bcrypt is used with passwords, and it essentially uses a derivation of the Blowfish algorithm, converted to a hashing algorithm, to hash a password and add Salt to it.

Section 6.3- Given a scenario, use appropriate PKI, certificate management and associated components.

Certificate authorities and digital certificates

Certificate Authorities
The CA is the trusted authority that certifies individuals' identities and creates electronic documents indicating that individuals are who they say they are. The electronic document is referred to as a digital certificate, and it establishes an association between the subject's identity and a public key. The private key that is paired with the public key in the certificate is stored separately. It is important to safeguard the private key, and it typically never leaves the machine or device where it was created.

The CA is more than just a piece of software, however; it is actually made up of the software, hardware, procedures, policies, and people who are involved in validating individuals' identities and generating the certificates. This means that if one of these components is compromised, it can negatively affect the CA overall and can threaten the integrity of the certificates it produces.

Every CA should have a certification practices statement (CPS) that outlines how identities are verified; the steps the CA follows to generate, maintain, and transmit certificates; and why the CA can be trusted to fulfill its responsibilities. It describes how keys are secured, what data is placed within a digital certificate, and how revocations will be handled. If a company is going to use and depend on a public CA, the company's security officers, administrators, and legal department should review the CA's entire CPS to ensure that it will properly meet the company's needs, and to make sure that the level of security claimed by the

CA is high enough for their use and environment. A critical aspect of a PKI is the trust between the users and the CA, so the CPS should be reviewed and understood to ensure that this level of trust is warranted.

Digital Certificates

A digital certificate binds an individual's identity to a public key, and it contains all the information a receiver needs to be assured of the identity of the public key owner. After an RA verifies an individual's identity, the CA generates the digital certificate, but how does the CA know what type of data to insert into the certificate? The certificates are created and formatted based on the X.509 standard, which outlines the necessary fields of a certificate and the possible values that can be inserted into the fields. As of this writing, X.509 version 3 is the most current version of the standard. X.509 is a standard of the International Telecommunication Union (www.itu.int). The
IETF's Public-Key Infrastructure (X.509), or PKIX, working group has adapted the X.509 standard to the more flexible organization of the Internet, as specified in RFC 3280, and is commonly referred to as PKIX for Public Key Infrastructure (X.509).

The following fields are included within a X.509 digital certificate:

- Version number Identifies the version of the X.509 standard that was followed to create the certificate; indicates the format and fields that can be used.
- Subject Specifies the owner of the certificate.
- Public key Identifies the public key being bound to the certified subject; also identifies the algorithm used to create the private/public key pair.
- Issuer Identifies the CA that generated and digitally signed the certificate.
- Serial number Provides a unique number identifying this one specific certificate issued by a particular CA.
- Validity Specifies the dates through which the certificate is valid for use.
- Certificate usage Specifies the approved use of certificate, which dictates intended use of this public key.
- Signature algorithm Specifies the hashing and digital signature algorithms used to digitally sign the certificate.
- Extensions Allow additional data to be encoded into the certificate to expand the functionality of the certificate. Companies can customize the use of certificates within their environments by using these extensions. X.509 version 3 has extended the extension possibilities.

The subject of a certificate is commonly a person, but it does not have to be. The subject can be a network device (router, web server, firewall, and so on), an application, a department, a company, or a person. Each has its own identity that needs to be verified and proven to another entity before secure, trusted communication can be initiated.

If a network device is using a certificate for authentication, the certificate may contain the network address of that device. This means that if the certificate has a network address of 10.0.0.1, the receiver will compare this to the address from which it received the certificate to make sure a man-in-the-middle attack is not being attempted**.**

CA

Four main types of certificates are used:

- End-entity certificates
- CA certificates
- Cross-certification certificates
- Policy certificates

A CA issues end-entity certificates to a specific subject, such as Joyce, the Accounting department, or a firewall. An end-entity certificate is the identity document provided by PKI implementations.

A CA certificate can be self-signed, in the case of a standalone or root CA, or a superior CA within a hierarchical model can issue it. The superior CA gives the authority and allows the subordinate CA to accept certificate requests and generate the individual certificates it. In these situations, a representative from each department requiring a CA registers with the higher trusted CA and requests a Certificate Authority certificate. Cross-certificates, or cross-certification certificates, are used when independent CAs establish peer-to-peer trust relationships. Simply put, they are a mechanism through which one CA can issue a certificate allowing its users to trust another CA. Within sophisticated CAs used for high-security applications, a mechanism is required to provide centrally controlled policy information to PKI clients. Placing the policy information in a policy certificate often does this.

CRLS

Certificate revocation is the process of revoking a certificate before it expires. A certificate may need to be revoked because it was stolen, an employee has moved to a new company, or someone has had their access revoked. A certificate revocation is handled either through a certificate revocation list (CRL) or by using the Online Certificate Status Protocol (OCSP). A repository is simply a database or database server where the certificates are stored.

The process of revoking a certificate begins when the CA is notified that a particular certificate needs to be revoked. This must be done whenever the private key becomes known. The owner of a certificate can request that it be revoked at any time, or the administrator can make the request.

OCSP

The CA marks the certificate as revoked. This information is published in the CRL and becomes available using the OCSP. The revocation process is usually very quick; the time required is based on the publication interval for the CRL. Disseminating the revocation information to users may take longer. Once the certificate has been revoked, it can never be used—or trusted—again.

The CA publishes the CRL on a regular basis, usually either hourly or daily. The CA sends or publishes this list to organizations that have chosen to receive it; the publishing process occurs automatically in the case of PKI. The time between when the CRL is issued and when it reaches users may be too long for some applications. This time gap is referred to as latency. OCSP solves the latency problem: If the recipient or relaying party uses OCSP for verification, the answer is available immediately. Currently, this process is under evaluation and may be replaced at some time in the future.

When a key is compromised, a revocation request should be made to the CA immediately. It may take a day or longer for the CRL to be disseminated to everyone using that CA.

Pki

The Public-Key Infrastructure (PKI) is intended to offer a means of providing security to messages and transactions on a grand scale. The need for universal systems to support e-commerce, secure transactions, and information privacy is one aspect of the issues being addressed with PKI.

PKI is a two-key, asymmetric system with four main components: certificate authority (CA), registration authority (RA), RSA (the encryption algorithm), and digital certificates. Messages are encrypted with a public key and decrypted with a private key. As an example, take the following scenario:

1. You want to send an encrypted message to Jordan, so you request his public key.
2. Jordan responds by sending you that key.
3. You use the public key he sends you to encrypt the message.
4. You send the message to him.
5. Jordan uses his private key to decrypt the message.

The main goal of PKI is to define an infrastructure that should work across multiplevendors, systems, and networks. It's important to emphasize that PKI is a framework and not a specific technology. Implementations of PKI are dependent on the perspective of the software manufacturers that implement it. This has been one of the major difficulties with PKI: Each vendor can interpret the documents about this infrastructure and implement it however they choose. Many of the existing PKI implementations aren't compatible with each other, but this situation should change over the next few years because customers demand compatibility**.**

Public key

Public key infrastructures (PKIs) are becoming a central security foundation for managing identity credentials in many companies. The technology manages the issue of binding public keys and identities across multiple applications. The other approach, without PKIs, is to implement many different security solutions and hope for interoperability and equal levels of protection.

Private key

PKIs comprise components that include certificates, registration and certificate authorities, and a standard process for verification. PKI is about managing the sharing of trust and using a third party to vouch for the trustworthiness of a claim of ownership over a credential document, called a certificate.

The Public-Key Infrastructure (PKI) is intended to offer a means of providing securityto messages and transactions on a grand scale. The need for universal systems to support e-commerce, secure transactions, and information privacy is one aspect of the issues being addressed with PKI.

PKI is a two-key, asymmetric system with four main components: certificate authority (CA),

registration authority (RA), RSA (the encryption algorithm), and digital certificates. Messages are encrypted with a public key and decrypted with a private key. As an example, take the following scenario:

1. You want to send an encrypted message to Jordan, so you request his public key.
2. Jordan responds by sending you that key.
3. You use the public key he sends you to encrypt the message.
4. You send the message to him.
5. Jordan uses his private key to decrypt the message.

The main goal of PKI is to define an infrastructure that should work across multiple vendors, systems, and networks. It's important to emphasize that PKI is a framework and not a specific technology. Implementations of PKI are dependent on the perspective of the software manufacturers that implement it. This has been one of the major difficulties with PKI: Each vendor can interpret the documents about this infrastructure and implement it however they choose. Many of the existing PKI implementations aren't compatible with each other, but this situation should change over the next few years because customers demand compatibility.

The Basics of Public Key Infrastructures

A PKI provides all the components necessary for different types of users and entities to be able to communicate securely and in a predictable manner. A PKI is made up of hardware, applications, policies, services, programming interfaces, cryptographic algorithms, protocols, users, and utilities. These components work together to allow communication to take place using public key cryptography and asymmetric keys for digital signatures, data encryption, and integrity. Although many different applications and protocols can provide the same type of functionality, constructing and implementing a PKI boils down to establishing a level of trust.

If, for example, John and Diane want to communicate securely, John can generate his own public/private key pair and send his public key to Diane, or he can place his public key in a directory that is available to everyone. If Diane receives John's public key, either from him or from a public directory, how does she know it really came from John? Maybe another individual is masquerading as John and replaced John's public key with her own. If this took place, Diane would believe that only John could read her messages and that the replies were actually from him. However, she would actually be communicating with Katie. What is needed is a way to verify an individual's identity, to ensure that a person's public key is bound to their identity and thus ensure that the previous scenario (and others) cannot take place.

In PKI environments, entities called registration authorities and certificate authorities (CAs) provide services similar to those of the Department of Motor Vehicles (DMV). When John goes to register for a driver's license, he has to prove his identity to the DMV by providing his passport, birth certificate, or other identification documentation. If the DMV is satisfied with the proof John provides (and John passes a driving test), the DMV will create a driver's license that can then be used by John to prove his identity. Whenever John needs to identify himself, he can show his driver's license. Although many people may not trust John to identify himself truthfully, they do trust the third party, the DMV.

What does the "infrastructure" in "public key infrastructure" really mean? An infrastructure provides a sustaining groundwork upon which other things can be built. So an infrastructure works at a low level to provide a predictable and uniform environment that allows other higher-level technologies to work together through uniform access points. The environment that the infrastructure provides allows these higher level applications to communicate with each other and gives them the underlying tools to carry out their tasks.

Registration

A key pair (public and private keys) can be generated locally by an application and stored in a local key store on the user's workstation. The key pair can also be created by a central key-generation server, which will require secure transmission of the keys to the user. The key pair that is created on the centralized server can be stored on the user's workstation or on the user's smart card, which will allow for more flexibility and mobility. In most modern PKI implementations, users have two key pairs. One key pair is often generated by a central server and used for encryption and key transfers. This allows the corporate PKI to retain a copy of the encryption key pair for recovery, if necessary. The user to make sure that she is the only one with a copy of the private key usually generates the second key pair, a digital signature key pair. Nonrepudiation can be challenged if there is any doubt about someone else obtaining a copy of an individual's signature private key. If the key pair was created on a centralized server, that could weaken the case that the individual was the only one who had a copy of her private key. If a copy of a user's signature private key is stored anywhere other than in her possession, or if there is a possibility of someone obtaining the user's key, then true nonrepudiation cannot be provided**.**

Trust models

We need to use a PKI if we do not automatically trust individuals we do not know. Security is about being suspicious and being safe, so we need a third party that we do trust to vouch for the other individual before confidence can be instilled and sensitive communication can take place. But what does it mean that we trust a CA, and how can we use this to our advantage?

When a user chooses to trust a CA, she will download that CA's digital certificate and public key, which will be stored on her local computer. Most browsers have a list of CAs configured to be trusted by default, so when a user installs a new web browser, several of the most well-known and most trusted CAs will be trusted without any change of settings.

In the Microsoft CAPI environment, the user can add and remove CAs from this list as needed. In production environments that require a higher degree of protection, this list will be pruned, and possibly the only CAs listed will be the company's internal CAs. This ensures that the company will automatically install digitally signed software only if it was signed.