

**MANAGEMENT 414
SANS +S™
TRAINING PROGRAM
FOR THE CISSP®
CERTIFICATION EXAM**

414.2

**Telecommunications,
Network and Internet
Security, and Security
Management Practices**

Copyright © 2006, The SANS Institute. All rights reserved. The entire contents of this publication are the property of the SANS Institute. User may not copy, reproduce, distribute, display, modify or create derivative works based upon all or any portion of this publication in any medium whether printed, electronic or otherwise, without the express written consent of the SANS Institute. Without limiting the foregoing, user may not reproduce, distribute, re-publish, display, modify, or create derivative works based upon all or any portion of this publication for purposes of teaching any computer or electronic security courses to any third party without the express written consent of the SANS Institute.

Preface

The cardinal rule for SANS training is that after you take a course you should be able to apply what you learned directly the day you get back into the workplace. My journey into writing about the 10 Domains started when Stephen Northcutt asked that I lead the development of adding the ISC2 10 Domains of Knowledge into SANS Security Essentials. That is SANS most popular training course and when taught bootcamp style it does an amazing job of helping students become capable of using hands-on techie tools. However, there had been a split in the community whether Security Essentials, which favored technical and pragmatic material, or the ISC2 10 Domains, which favors theory, should be the baseline standard for an information security professional. We were discussing this hotly debated issue in the SANS faculty speaker room over lunch one day and it suddenly dawned on us, why not add the 10 Domains into Security Essentials? Tony Cole, CISSP, was assigned to evaluate Security Essentials and determined that about 60% of the 10 Domains material was already covered in Security Essentials. Clement Dupuis and I were the leads on the project. This was a very successful edit and a number of students have passed their CISSP exams after going through SANS Security Essentials with the ISC2 10 Domains. However, when we added the additional material there was no longer time to cover the application of the material to the workplace; in addition, there are some students who prefer the more formal 10 domain structure.

To best meet the needs of the students, SANS authorized the creation of Management 414, SANS CISSP® 10 Domains +S™, which covers the 10 Domains of Knowledge in a formal 10 domain structure. In the meantime, Clement Dupuis, Stephen Northcutt, Marcus Sachs, Bill Stearns and Joshua Wright are removing some of the 10 Domains material from SANS Security Essentials and returning it back to the original vision for that track, to fully cover the essentials of technical security. Moreover, SANS has insisted that the course teach the application of the 10 Domains in the workplace - something no other 10 Domains course, including ISC2's does. This course meets the SANS promise: what you learn in the course you will be able to apply in the work place. One of the most important things I have learned from Alan Paller, Director of Research, in the years I have been involved with SANS is the importance of community consensus. In order to provide the highest quality training we have recruited experts to review the material and come to consensus on the course content and the application of the information. With the help of Zoe Dias, SANS Faculty Director, we enlisted a total of 68 reviewers from 10 countries. All but two of the reviewers are active CISSP's. The main author for the course, Eric Cole, has been a CISSP for almost 10 years.

SANS enthusiastically applauds the expert work of our technical reviewers/editors:

Monica Anklam, CISSP No. 31995, USA
Alex Arndt, CISSP No. 52343, Canada
Hank Askin, CISSP No. 40792, USA
Anjali Atanacio, CISSP No. 27039, USA
Jason Bevis, CISSP No. 35285, USA
Ron Black, CISSP No. 24245, USA
Anton Bojanec, CISSP No. 24560, Slovenia
Olufremi Bolanle, CISSP No. 51582, Nigeria
Jeff Bontsas, CISSP No. 39135, USA
Derek Browne, CISSP No. 26099, Canada
Sherry Callahan, CISSP No. 21760, USA
Ed Capizzi, CISSP No. 35909, USA
Jim Cate, CISSP No. 37031, USA
Patrick Chan, CISSP No. 40222, Canada
Jerry Chen, CISSP No. 47413, Canada
Daniel Cline, CISSP No. 31366, USA
Chris Cook, CISSP No. 38254, UK

Edwin Covert, CISSP No. 3597, USA
Phil Curran, CISSP No. 31708, USA
Edgar Danielyan, CISSP No. 42834,
UK/Armenia
David Dann, CISSP No. 51571, USA
Gary Delaney, CISSP No. 37636, USA
Sandeep Dhameja, CISSP No. 33585, USA
Joe Dial, CISSP No. 25358, USA
Heinz Durr, CISSP No. 42160, Switzerland
Darin Dutcher, CISSP No. 41299, USA
Rene Evers, CISSP No. 29057, USA
Chris Farrow, CISSP No. 45570, USA
Kenneth Fox, CISSP No. 42293, USA
Roger Fradenburgh, CISSP No. 28099, USA
Brian Freedman, CISSP No. 49504, USA
Donald Glass, CISSP No. 42244, USA
Mark Heinrich, CISSP No. 36190, USA

Lorna Hutcheson, USA
Lawrence Johnson, CISSP No. 25456, USA
Chaiw Kok Kee, CISSP No. 31589, Malaysia
Darrin Lau, CISSP No. 29948, USA
Eliot Leibowitz, CISSP No. 43782, USA
Steven Leong, CISSP No. 30313, Singapore
Chip Meadows, CISSP No. 10070, USA
Sean Mitchell, CISSP No. 36817, USA
Michael Morrell, CISSP No. 36227, USA
Pamela Nottage, CISSP No. 3758, USA
Sanjay Pandit, CISSP No. 44786, USA
John Pao, CISSP No. 29876, USA
Ariya Parsamanesh, CISSP No. 36074, AUS
Stephen Patton, CISSP No. 49746, USA
Robert Pfau, CISSP No. 21572, USA
Gabriel Proulx, CISSP No. 34018, Canada

Jim Purcell, CISSP No. 34519, USA
Andrew Salzman, CISSP No. 25162, USA
Amarottam Shrestha, CISSP No. 41671, AUS
Michael Solomon, CISSP No. 26517, USA
Robert Sorensen, CISSP No. 48304, USA
George Starcher, CISSP No. 34689, USA
Bruce Swartz, CISSP No. 46522, USA
David Taylor, CISSP No. 55890, USA
Brad Towers, CISSP No. 27957, USA
Jill Treu, CISSP No. 43196, USA
Tim Weil, CISSP No. 44250, USA
Deborah Weinstein, CISSP No. 44411, USA
Melody Wilson, CISSP No. 4130, USA
Steven Winterfield, CISSP No. 38096, USA
Kelli Wolfe, USA
Wayde York, CISSP No. 30404, USA

I have had the privilege of the best seat in the house and have really enjoyed working with the CISSP team. I sincerely hope that you benefit greatly from the information in these books and am very interested in your feedback. Please feel free to send me suggestions, corrections or questions to [mqt414\(q\)sans.orq](mailto:mqt414(q)sans.orq).

Eric Cole, Senior Instructor and Research Fellow
The SANS Institute

2. Telecommunications and Network Security

10 Domains of Knowledge

This section covers Domain 2, the Telecommunications, Network, and Internet Security domain.

2. Telecommunications and Network Security

- Internet/intranet/extranet: Firewalls, routers, gateways, and various protocols
- Communications security management and techniques that prevent, detect, and correct errors so that integrity, availability, and confidentiality of transactions over networks can be maintained

The Telecommunications, Network, and Internet Security domain includes the structures, transmission methods, transport formats, and security measures used to provide integrity, availability, authentication, and confidentiality for transmissions over private and public communications networks and media. This domain covers a breadth of topics relating to the critical elements of telecommunications, networking, and Internet security issues. As this domain has evolved, more and more emphasis is placed on the networking infrastructure items and how they relate to telecommunications. This section examines both the hardware and software issues relating to the communications protocol. It addresses how you take information in digital form and route it to another location in a timely and responsive fashion.

Domain 2: Agenda

- Defense and attacks
 - Intrusion detection and response
 - Firewalls
- Networking methodologies
 - Topologies
 - Transmission
 - Hardware
 - Theory and design

In this domain, we look at defense and attacks, including intrusion detection systems (IDSs), firewalls, and different methods of attack. To understand how to defend against something, you need two critical pieces of information. First, you need to understand how the technology works. Second, you need to know how the offense operates. Understanding the tactics and attack methods of attackers enables you to build strong, robust defensive measures.

This section starts by looking at the various security devices that you can deploy on a network. The main prevention device examined is the firewall. We look at the three types of firewalls and the strengths and weakness of each. We examine the strengths and weakness from both a security and functionality standpoint. We then look at the main detection device, which is an intrusion detection system. We also look at the security issues associated with tuning these devices.

After we look at each device, we look at a holistic approach of the network, examining how and why networks are set up in a specific fashion.

Intrusion Detection and Response

- Intrusion detection and response is the task of monitoring systems for evidence of an intrusion or inappropriate usage.
- Intrusion detection and response includes notifying the appropriate parties to take action to determine the extent of the severity of the incident and remediate the effects of the incident.

Intrusion detection and response is a two-fold operation. Intrusion detection deals with detecting an attack, and an intrusion response deals with how you handle the intrusion and recover your system in a timely manner. It is important to remember that intrusion detection and response plays a role in disaster recovery planning, which is a separate domain. Disaster recovery planning deals with all the steps to recover from a disaster, and intrusion detection and response plays a sub role in dealing with the incident. This is why it is sometimes called incident response.

Intrusion detection deals with all the activities of detecting an intrusion or attack on your network. IDSs play a key role in setting off alerts, but they are useful only if those alerts are properly tuned so that there is reliability in the results. Another critical part of intrusion detection is understanding your critical systems across your network and being able to detect any changes in a timely manner. This is more of a holistic approach than an IDS, but the two operate nicely together to provide a robust defense-in-depth level of protection.

Detecting an attack is only part of the puzzle. Being able to react to the attack in a timely manner and take action is just as critical. The ultimate goal of detecting an attack is to be able to prevent it in a timely manner. The quicker you can detect an attacker, the quicker you can prevent it and minimize the amount of damage to your systems.

Intrusion Detection and Response Goals

1. Creation and maintenance of intrusion detection systems and processes
2. Creation of a Computer Incident Response Team (CIRT)
 - Analysis of an event
 - Responding to an incident if the analysis warrants it
 - Escalation-path procedures
 - Resolution, post-incident follow up, and reporting to appropriate parties

Intrusion detection and response is a capability. When you build a capability, you need to follow certain steps. These steps are general and need to be customized for your organization; nonetheless, they provide a good roadmap for you to follow. It is important to note that even though they need to be customized, none of the steps can be skipped, and they all must be addressed.

The first step is to identify what systems you want to protect. This is the beginning of creating your plan. Even though you might have a lot of systems on your network, you cannot protect all of them. Being able to prioritize and pick the most critical ones to start with is important. After you determine which systems you are going to protect, you build a written plan containing checklists of what you are going to do to protect those systems. It is critical that you have a written process so that it is consistently followed across your organization. Some of the critical processes you need to account for are host (or network monitoring) and event notification.

The second step involves focusing on the intrusion response piece: what you do when you have an intrusion. The key part in this step is the formation of a Computer Incident Response Team (CIRT). This team is going to be the main focus for handling an incident at your organization. They need to be trained in the proper skills to be able to handle an incident in a timely manner.

Intrusion Detection Systems (IDSs)

- Types of IDSs:
 - Network-based
 - Host-based
- Methods of operation
 - Pattern matching
 - Anomaly detection
 - Protocol behavior

Intrusion detection system analysis is usually broken into two main categories: types and methods. Either type can work with either method, but the most common type and method combination is a network-based intrusion detection system (NIDS) with pattern matching.

The two types of systems covered here are network-based and host-based. A network-based IDS sits on the network like a sniffer, filtering out traffic that might be indicative of an attack. A host-based IDS sits on a computer and watches the system, looking for signs of attack.

The three common methods for detecting an attack are pattern matching, anomaly detection, and protocol behavior. Pattern matching, or signature-based IDSes, look for patterns of an attack and set off an alert. Anomaly-detection systems determine what is normal traffic on your network; anything that falls outside that norm is deemed an attack. Protocol behavior systems look at the RFC protocol standards to determine what is not normal traffic; anything that falls within that range is flagged.

These areas are covered in more detail over the next several slides.

Network-Based IDS (NIDS)

- Sits on a network and sniffs traffic
 - Essentially a sniffer with rules
- Looks for indication of an attack
- Operates in two modes
 - Passive: Sends alert, but does not stop the attack
 - Active: Stops the attack, usually by sending resets

A network IDS is essentially a device that sits on a network like a sniffer and gathers all traffic that passes over that network segment. Because it gathers all traffic across a network segment, it must be able to see all of the traffic. If it is connected via a hub, which is usually not a problem. If it is connected via a switch, however, it is critical that the switch be configured properly so that the IDS can see all the traffic without impacting the performance of the network.

By default, an IDS passively processes the traffic looking for signs of an attack. Usually an IDS operates in passive mode and sends off an alert but does not do anything to stop the attack. In essence, a console terminal must be monitored at all times; when an attack is detected, an alert is generated on the screen. It is then up to an operator or analyst to determine the extent of the problem and what to do. An IDS can also operate in active mode, in which it tries to stop the attack. That is, the IDS operates in more of an active way and requires less operator intervention. In active mode, when an attack is detected, the IDS automatically takes action to stop the attack. The most common way to do this is by sending resets to both the sender and receiver. Another method is to reconfigure the firewall to block the attack.

Host-Based IDS (HIDS)

- Sits on a host system
- Traditionally processes the log file in real time and looks for signs of an attack
- Sees exactly what the host sees
- Refers to anything that protects the host computer from attack

A HIDS sits on a host system and processes information to determine whether there is an attack. The positive aspect of a HIDS is that it sees whatever the host sees. Because a NIDS view is at the network level and not the host level, certain types of attacks can be used to bypass a NIDS. The drawback to a HIDS is that it is not as scalable as a NIDS because there is a one-to-one relationship between a HIDS and the system that it protects.

A traditional HIDS processes the log file in real time looking for signs of an attack and is able to see exactly what the host sees. However, most HIDS operate in passive mode and cannot do anything to stop an attack. A NIDS also typically operates in passive mode but can also be put into active mode. By nature of where a HIDS sits and how it operates, it is much harder to put a HIDS into active mode.

Over time, the meaning of HIDS has evolved into a more generic term. Today, most people use HIDS to refer to anything that protects the host computer from attack.

Pattern Matching

- It is the most common method of intrusion detection.
- It looks for patterns of an attack.
- It is also called signature detection.
- Signatures are rigid and prone to errors.

Pattern matching is the most common method of intrusion detection. It operates similarly to antivirus software in that it contains signatures for known types of attacks. This is why it is sometimes called signature detection rather than pattern matching. This method is fairly easy to deploy but has several limitations. The first issue is that the signatures are usually rigid and prone to errors or misclassification. The second problem is that because the system is looking for a signature, in most cases the system can detect only known attacks. Therefore, the system is less effective against new attacks, which are discovered on a regular basis.

This method is the most common method for attack detection for both HIDS and NIDS.

Anomaly Detection

- It understands and analyzes the traffic going over a network segment.
- Based on the analysis, it determines what is normal for a network.
- Anything falling outside the norm is deemed an attack.

In an ideal world, one way to detect an attack is to determine what is normal traffic and then anything that falls outside of that range is determined to be an attack. This is essentially what anomaly detection does. It first gathers all the traffic going across a network for a certain period of time. This is called the discovery stage. The remaining stages are only as good as the data gathered during the discovery stage. The problem with the discovery stage is that the assumption is that all traffic going across the network is legitimate and normal traffic and no attacks are present. If this is an active functioning network, that assumption is not true.

After the traffic is gathered, the next stage is the learning stage. This is where the system analyzes all the traffic, learning what normal traffic is usually based on protocol information. After the system has learned normal traffic, the system can be deployed on a network.

After the system is deployed, anything falling outside the range is deemed an attack. Because networks change over time, it is critical that these systems relearn the network at periodic intervals.

Protocol Behavior

- It uses the RFC protocol standards to determine what legitimate traffic is.
- It puts together profiles of illegal network traffic per protocol.
- Anything matching illegal traffic is deemed an attack.

The RFC standards documents contain what is required to write compliant network protocols. The standards are specific and straightforward. If you have not read an RFC standards document, you should. You might be surprised by how easy and straightforward they are to read.

These standards documents state what to do under normal circumstances. Therefore, by reading and analyzing the documents, it is fairly easy to determine what legitimate traffic is for a certain protocol and what type of traffic is illegal or not allowed. Based on this analysis, a system can determine any traffic that is illegal and deem it an attack. This method is considered less error prone than anomaly detection and is used more as a complement to pattern-matching systems.

IDS Events Defined

- True positive
- True negative
- False positive
- False negative

When classifying the accuracy of an IDS system, certain key terms are used. Therefore, you need to understand the following key terms with regard to IDSs:

- **True positive:** When the IDS sets off an alert and it is a real attack.
- **True negative:** When the IDS does not set off an alert and it is normal traffic.
- **False positive:** When the IDS sets off an alert and it is normal traffic.
- **False negative:** When the IDS does not set off an alert and it is attack traffic.

When measuring the effectiveness of an IDS, you want to increase the true positives and true negatives and decrease the false positives and false negatives. Most IDSs tend to have more false positives than false negatives when it comes to errors. Looking at the definitions, this should make sense. Having an IDS report an attack when there isn't one creates less of an impact than an IDS not reporting an attack when there actually is one and thereby allowing that attack to slip under the radar.

Firewalls

- **Firewalls sit between two networks and control the flow of traffic.**
- **There are three main main types:**
 - Packet filtering**
 - Stateful**
 - Proxy**
 - **Application level**
 - **Circuit level**

Firewalls

Probably the first thing any security analyst does when he designs a network is to plan for a firewall. It's almost impossible to have any kind of good internal security controls without first establishing a secure network perimeter. In fact, the principle of security in-depth practically demands that you be able to control the traffic entering and leaving your network. Fortunately, firewalls are visible components of today's information security scene. They're usually the first thing management thinks of when it writes out the security budget.

A good firewall (or at least a filtering router) can help prevent a variety of different types of attacks. In our scenario, it provides two helpful functions: It prevents outsiders from accessing internal network services and from using spoofed IP addresses, which should only appear inside your own network.

Blocking access to noncritical services probably is the single biggest benefit of any of the risk management techniques we're going to discuss. Why offer to the entire Internet every service that's running on your internal LAN? Offering such provides what the military would call a target-rich environment. If you narrow down to a select few the range of services you offer, you can concentrate on configuring those services in as secure a manner as possible, while simultaneously denying an attacker any possibility of using poorly managed secondary services against you.

Packet-Filtering Firewall

- Examines each packet independently and determines whether packets should pass or be dropped
- Has no idea of what traffic came before it
- Very fast, but not very secure
- Referred to as access control lists (ACLs) on some devices

A packet-filtering firewall is the most basic type of firewall. It is fairly simple in its processing capabilities, which means it is fast but not secure in protecting a network. A packet-filtering firewall works by examining each packet independently and determines whether it should pass or be dropped. This type of firewall has no idea of what traffic came before it. It essentially looks at only the network protocol information in each packet to determine whether the packet should be dropped or allowed on to the network. Because it has no idea of what other packets occurred on the network, it has to make assumptions, and those assumptions are not always correct.

Several types of attacks can be used to bypass these firewalls. Packet-filtering firewalls complement *detailed* defense policies that include other firewalls but are not usually used by themselves.

Stateful Inspection Firewall

- Keeps a state table of all traffic going across a network
- Uses the state table to determine whether a packet should pass or be dropped
- More secure, but slower than a packet-filtering firewall

A stateful inspection or stateful packet-filtering firewall builds on top of a packet-filtering firewall and overcomes many of the limitations. The big drawback of a packet-filtering firewall is that it has to make assumptions because it does not keep track of what packets occurred before the packet that is being examined. A stateful packet-filtering firewall overcomes this by keeping a state table of all traffic that occurred on the network. By having a state table, assumptions no longer have to be made when filtering out or dropping packets.

Because a stateful packet-filtering firewall has to maintain a state table, it increases the resources that have to be used on the firewall and therefore these types of firewalls are slower than packet-filtering firewalls. This should make sense because there is no free lunch in network security: Whenever you increase the security, you decrease the speed.

Proxy Firewall

- Creates two TCP connections for each request
- Maintains one TCP connection with the client and one with the server
- Also called an application proxy because it processes packets at all seven layers

When you read this slide, it is important that you think "proxy firewall." The term *proxy* has many meanings across the security industry, and many people use it in different fashions depending on the context. In this context, we are talking about a true proxy firewall that can filter out and drop packets.

Unlike the other two firewalls, instead of just examining the packets, a proxy firewall is actually the termination point for the network communication. Therefore, a proxy firewall truly sits between two systems that communicate. The way a proxy firewall does this is by creating two TCP connections for each request. It maintains one TCP connection with the client and one with the server. It is also called an application proxy because it processes packets at all seven layers of the OSI model.

Application-Level Firewall

Application Level:

- Implemented on a computer by using proxy server software
- Hides the origin of packet
- Acts as intermediary and moves an accepted packet from one network to another network (proxy server)
- Referred to as application layer gateway
- Operates at layer 7

Typically used with a dual-homed host and records session history.

Proxy firewalls can be established for a variety of protocols, including HTTP, SMTP, and FTP.

Circuit-Level Firewall

Circuit Level:

- Operates as a proxy server
- Does not use application-level proxy software
- Develops a virtual connection between the host and destination
- Typically sits at the session layer

The following are some key characteristics of a circuit-level firewall:

- Operates as a proxy server
- Does not use application level proxy software
- Develops a virtual connection between the host and destination
- Typically sits at the session layer

Firewall Architectures

- Packet filtering
- Dual-homed host
- Screened host
- Screened subnet firewall

The following are four firewall architectures that are discussed over the next few slides:

- Packet filtering
- Dual-homed host
- Screened host
- Screened subnet firewall

Firewall Architectures (2)

- Packet-filtering router
 - Packet-filtering firewall
 - Manages connection to demilitarized zone (DMZ)
 - Separates organization's trusted network from external, untrusted network

The DMZ or perimeter network is placed between an internal network and an external network to provide protection to the internal network.

It is implemented by an outside router that protects against external penetrations and an inside router that controls the internal, trusted network's access to the DMZ.

Firewall Architectures (3)

- Dual-homed host
 - Uses two network interface cards (NICs)
 - One NIC attaches to a trusted network
 - The other NIC attaches to the untrusted network
 - The packets must pass through the host and be checked

In this architecture, the routing capabilities of the host must be disabled. The reason for this is that you do not want to accidentally set up internal routing that will pass traffic from one NIC to the other without any packet inspection.

Similarly, IP traffic forwarding should be disabled.

Firewall Architectures (4)

- Screened host firewalls
 - Uses two network interface cards (NICs)
 - One NIC attaches to a trusted network
 - The other NIC attaches to the untrusted network
 - Uses a screening router placed between the untrusted network and the trusted network

Some of the key characteristics of the screen host firewall follow:

- Uses two network interface cards (NICs)
- One NIC attaches to a trusted network
- The other NIC attaches to the untrusted network
- Uses a screening router placed between the untrusted network and the trusted network

Firewall Architectures (5)

- Screened subnet firewalls
 - Uses two network interface cards (NICs)
 - One NIC attaches to a trusted network
 - The other NIC attaches to the untrusted network
 - Uses two screening (packet-filtering routers)
 - One router screens incoming and outgoing messages to the Internet.
 - Second router screens local network traffic.

Some of the key characteristics of the screen subnet firewall follow:

- Uses two network interface cards (NICs).
- One NIC attaches to a trusted network.
- The other NIC attaches to the untrusted network.
- Uses two screening (packet-filtering routers).
- One router screens incoming and outgoing messages to the Internet.
- Second router screens local network traffic.

Firewall Architectures (6)

- Bastion host
 - It is a host computer in the public area of a DMZ.
 - It is exposed to attack from the Internet.
 - It must have functions to protect itself.
 - It can be a firewall or router.

Web, mail, and FTP servers can be considered bastion hosts. A bastion host is a host computer in the public area of a DMZ and is exposed to attack from the Internet.

Firewall Architectures (7)

- SOCKS
 - A circuit-level proxy server that is used to authenticate a client
 - Transport layer protocol
 - Replaces network system calls with socks calls
 - Network utilities have to be "socksified" to operate (ftp and telnet, for example).

SOCKS is a circuit-level proxy server that is used to authenticate a client. It supports hosts to connect through a firewall to an internal computer and it supports internal computer connections to external networks.

Incident Handling - Six Steps

- Preparation
- Identification
- Containment
- Eradication
- Recovery
- Lessons learned

This section covers the six steps in incident handling: preparation, identification, containment, eradication, recovery, and lessons learned. The steps serve you, the handler, as a compass or a roadmap, a way to keep in mind what you should try to do and the things you need to do next.

Some people think if they follow only some of the steps, they will be in good shape. That idea is wrong. To successfully handle an incident, you must follow all six steps. In addition, each step must be customized to the particular company and the industry in which you work. The following slides help you do that.

Preparation

- Planning is everything.
- Policy
 - Organizational approach
 - Interorganization
- Obtain management support.
- Select team members.
- Identify contacts in other organizations (legal, law enforcement).

When it comes to incident handling, planning is everything, and preparation plays a key role. It is important that you have a policy in place that covers an organization's approach to dealing with an incident. Among other things, the plan needs to cover the following:

- Whether the company is going to notify law enforcement agencies or run silent
- Whether the company is going to contain and clean up an incident or watch and learn

One thing you really want to avoid is having an incident happen and finding yourself in a debate about whether to contain the incident and clean up or to watch the attackers and try to gather more evidence. The time to make these (career-affecting) decisions is before the incident, keeping senior management and your legal staff apprised. The policy should also contain both an intra-organization approach and how a company works with other companies (inter-organizational) on an incident.

It is important that an incident-handling team has management support and buy-in. The last thing a company wants is for senior management to question or doubt the decisions that were made during an incident.

Not everybody makes a good incident handler. I have worked with some smart people whose personalities do not lend themselves to being good incident handlers. People who like to work solo and as heroes usually do not make good team members. You want someone who works well in a team environment, thinks out solutions, and does not make rash decisions.

Preparation (2)

- Update disaster recovery plan.
- Compensate team members.
- Provide checklists and procedures.
- Have emergency communications plan.
- Escrow passwords and encryption keys.
- Provide training.
- Have a jump bag with everything you need to handle an incident.

During preparation, a company needs to make sure it updates its organization's disaster recovery plan to include computer incident handling. A large organization with more than 10,000 computers is going to rack up some incidents. This can cause the incident handlers to burn out. Interestingly, they tend to burn out just as they get really good. After training and seasoning, they do a bang up job on a couple of hot problems, and the next thing you know they are suffering from various stress effects. The solution seems to be a set of things, including rewards and compensation (such as time off). This may run afoul of your organizational culture, but consider this. When do incidents occur? On Friday afternoons at 3:30 p.m? Do the handlers and administrators go home and wait until Monday to start on the clean up? No; in almost every case, they stay until the job is done. So you need to reward these people and let them get some rest.

Computing environments are complex, and no one knows every variant of Unix and so forth. Although we can try to make sure you have a solid grounding in the basics of handling systems, memory fades over time. Having a checklist to refer to that describes how to bring down or back up a system can help prevent errors and reduce the stress on the handler. If handlers are following the checklist and their attempts blow up, it is not their fault. It is simply time to update the checklist.

As a system administrator of a production system, I was never comfortable making privileged passwords available to others. However, in an emergency, a handler may need access to critical systems. One organization has a policy whereby the passwords are kept in sealed envelopes in locked containers. After several years of implementation, the organization reports that although sometimes cumbersome, this system has worked well for them. Note as well the two-fold responsibility here:

- The system administrators must make sure the envelopes are kept updated.
- The handlers must make sure they tread lightly on the systems, keep the administrators up-to-date on any changes they make, and, above all, never use a privileged password unless they are qualified on the affected operating system. One thing that will definitely make an incident worse is a clueless handler fumbling around as the administrator or root.

Not many of us can change the way our entire organization does business, but we can certainly be responsible for the way that we do business. Encourage people to write down critical passwords and encryption keys and store them safely so that they can be accessed if required. As encryption becomes more prevalent, an organization must set policy as to who owns the secret keys and passphrases and under what circumstances they can be used and accessed.

Being able to react when an incident occurs is important. Utilizing a "jump bag" that contains everything you need to handle an incident will allow you to react in a more timely fashion.

Identification

- How do you identify an incident?
- Be willing to alert early, but do not jump to a conclusion.
 - "Boy who cried wolf" syndrome
 - Look at all the facts
- Notify the correct people.
- Use the help desk to track trouble tickets and problems.

Bad things can happen when unqualified, unauthorized people make the call on an incident. Thousands of dollars later, after three days offline, you question that individual, "What were you thinking?" The answer is usually the same, "I, uh, thought it was nothing."

After a fire alarm is pulled, qualified firefighters who know the signs to look for come to the site and investigate. Only then does the person in charge at the scene authorize re-entry into the building. This should be the paradigm we work under. Be willing to alert early, have trained people look at the situation, and then stand down if nothing is wrong at a minimum of expense. Either way, make sure you have mechanisms in place to identify an incident.

There is nothing wrong with alerting early if you maintain situational awareness and everyone understands that it might not be an incident. You want to avoid screaming it is an incident and an hour later saying, " Oh, never mind." If you do this several times, you will be a victim of the "boy who cried wolf syndrome (meaning that when an incident actually occurs, no one will believe you because you were wrong so many times before).

Also, when it comes to identifying an incident, do what you are good at and utilize others in the organization. Why should an incident-handling team go through the trouble of tracking issues when the help desk is set up to do this on a regular basis? Let the help desk and others help you track issues; doing so will help ensure that all of the issues are resolved.

Identification (2)

- Assign a primary handler.
- Determine whether an event is an incident.
 - SMART guidelines
- Identify possible witnesses and evidence.
- Make a clean backup of the system

If one person isn't in charge, no person is in charge. For smaller incidents, often of the "would you check this out" category, there is no need to send core incident handlers. It is recommended practice to have a core team of well trained handlers and to have incident-handling skills and training as part of the job for security officers or system administrators. An organization that does this benefits by having multiple levels of trained "firefighters." However, in such a case, it is important to set up assignments in a way that encourages the system administrator to succeed.

Handlers who are not full-time should be given assignments in a way that clearly identifies what is expected of them: the quality of their investigation, their responsibility to preserve and collect evidence, what documentation they should produce, and when it is due. It is also important that they know who they can call for additional guidance or support.

After you determine whether an event is actually an incident, take the steps needed to build a criminal or civil case if appropriate. Immediately identify witnesses and get written statements of what they saw or heard while memory is fresh. Also be alert to information that could serve as evidence. According to the Federal Rules of Evidence, 702, "If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, if (1) the testimony is sufficiently based upon reliable facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case." **Application:** You are the expert. Therefore, you need to make sure you have the training and that as you handle the incident you apply accepted principles and methods.

At this point, you must make the decision whether to involve law enforcement. Senior management should always be involved in that decision unless you have a detailed policy to follow. The point where the handler validates that this appears to be an incident is also where the "contain and clean" or "watch and learn" decision is made. If possible, always make a clean binary backup of the system before you start making any modifications.

Containment

- An incident handler should not make things worse (liability and negligence).
- Secure the area.
- Make a backup.
- Pull the system off the network (optional).
- Change passwords.

Incident handlers should not make things worse; they should make things better. Above all, they should understand the basic principles of liability and negligence. As a handler, you are responsible for meeting the expectations of the prudent person rule. In a nutshell, this says that you should do what a reasonable person would be expected to do. Further, you should be aware that the corporate officers of your organization may be held liable for what you do, or what you do not do, if unlawful activities occur.

This can be one of the areas in which incident handlers can run into trouble. Nothing about incident handling allows you to break the law. If you suspect someone of downloading child pornography, for instance, you can't download these files to your computer to examine them. You also have to be careful to exercise due care, especially with privacy (Electronic Communications Privacy Act). For instance, if you are an Internet service provider (ISP), you cannot just release the personal information (home address, name, and credit card information) of a subscriber just because someone claims an attack.

Negligence for failure to meet a certain standard of care is generally determined by a court of law. Specifically, *negligence* is defined as the "failure to exercise the degree of care expected of a person of ordinary prudence in like circumstances in protecting others from a foreseeable risk of harm in a particular situation." In other words, a company that acts reasonably or with "due care" generally will not be found negligent.

When containing an incident, first secure the area. In doing so, make a backup of all infected systems; and if the original hard drive cannot be kept for evidence, multiple copies of the backups should be made. One should be kept for evidence and the other used to analyze the incident. At some point in the containment process, a decision needs to be made about whether the systems should be pulled off of the network or whether the entire network should be pulled from the Internet. In addition, passwords should be changed to make sure a compromised account cannot be used as a re-entry point into the system. However, a binary backup should be made prior to making any changes to the system.

Eradication

- Fix the problem before putting the system back online.
- Determine cause and symptom.
- Improve defenses.
- Perform vulnerability analysis.

Before the system goes back online, an incident handler must fix the problem; otherwise the vulnerability that the attacker used to compromise the system will still exist. Nuking the operating system from high orbit may be considered a shortcut in the handling process. Although it is certainly true that total destruction of the contents of the disk will take care of any malevolent code, the opportunity for re-infection via the same channel after you reload the operating system still exists. There are many cases in which handlers have taken systems down and reloaded the operating system only to have the box compromised again a couple of days later. The best course of action is to determine what the cause of the incident is, find the vector of infection, and act to prevent it from happening again.

When your system is hacked, word gets out and every hacker on the planet lines up to take another shot at you. It is not enough just to recover the system; the security of the affected system(s) needs to be upgraded. If it is a production system, you may hear arguments that the organization cannot take the risk of modifying it. This is an important and somewhat valid argument. The counter to this is that if the system has been compromised, it must have a vulnerability. If you do not remove the vulnerability, the system may become compromised again.

The simple trick of changing the name and IP address of the system can solve a lot of problems. If your organization has the time and resources, this can be a good opportunity to play with a "honeypot," a system that is designed to collect information about an attacker without yielding useful data. One of the most interesting sources of software for building a honeypot is Fred Cohen's deception toolkit, available from: <http://all.net>.

Vulnerability scanners, such as the NAI Cybercop, Internet Security Scanner, Cisco's Net Sonar, Nessus, Nmap, and Saint can identify weaknesses in your organization's internal network. The commercial software packages listed are somewhat expensive. If money is an issue, you can pay a consultant to run the software for you on a one-time or on a recurring basis and provide a report. Nmap, a free tool, is becoming one of my favorite tools; I have also had good success with Saint, another free tool.

After placing a suspect system on a small hub and doing the backup, it is helpful to run Nmap on the target computer from another system on the hub. This can give you insight into potential problems.

Running a security scanner on the neighboring systems in a compromise can help you ensure you have full and complete eradication. If one system is compromised, there is every chance the number is actually two or more.

Recovery

- Make sure you do not restore compromised code.
- Validate the system.
- Decide when to restore operations,
- Monitor the systems.

How do you restore from backups and ensure you are not reloading compromised code? There is no easy solution, but you can use file integrity software in reverse. Use a software package such as Tripwire on the compromised system and then do a restore from backups, possibly on a clean system. Run Tripwire again and compare the results. This can help you find the compromised code. For best results, mount the disk that you are running Tripwire on from a system with a known-good operating system. This way kernel modules will fail to protect the compromised code.

Remember again that after you have touched the machine, everything that breaks is your fault. Be sure to get the owner of the machine to sign that it is back in full operation. Make every effort to ensure the system works properly before leaving the scene.

The decision of when to put the system back into business has to be made by the system owner. As a handler, you can give owners advice, but this is their decision to let you know when to put the system back into business. They are the ones who depend on the system.

Needless to say, if the eradication is not complete or the infection vector has not been closed off, the earlier you detect re-infection, the better. It is also politically better if the handlers detect the problem and show up to fix it than if the problem comes to light because business operations are affected. This is a serious problem. Many times handlers take shortcuts, or there is something you undiscovered about the trust relationship, and the problem recurs.

- Develop a report.
 - Try to get consensus.
- Conduct a lessons-learned meeting
- Send recommendations to management.
- Conduct a follow-up meeting.

The only one who should write the report is the on-site handler. On-site handlers submit the draft, but you should allow everyone involved to review the draft. If someone has a strong disagreement about the facts involved, he can submit a statement that remains a part of the incident record. It is far better to find out that you have a lack of consensus before going to court than during court proceedings!

After the report has been reviewed, schedule a lessons-learned meeting. In general, the main purpose of such a meeting is to get consensus on the executive summary of the report.

What is the most important thing for an executive summary to cover? The answer is how much the organization saves using an effective incident-handling procedure!

With every incident, mistakes occur. You learn from these, improve your process for the future, and move on. Sometimes you run into policy or other organizational problems that hinder bringing the incident to a close. Note these and submit them to management for consideration.

Follow-up meetings are never the most popular events. People are tired; they have been under stress. The system is now back in operation, and the last thing anyone wants to do is have a meeting to rehash painful memories. However, this is a valuable tool for organizational improvement. This is the hardest time not to blame people. Remember that the focus should be on process improvement.

Internet, Intranet and Extranet

- Internet
 - Runs on TCP/IP protocol
 - Global network of public networks, network access points (naps), and service providers
 - Operated either for public access or private data exchange (with a VPN)
- Intranet
 - Internet-like logical network
 - Based on an organization's internal, physical network infrastructure
 - TCP/IP and http standards
 - Web browsers
- Extranet
 - Private network using Internet protocols
 - Accessible by partners and vendors outside of the organization, but not by the general public

Following are descriptions for Internet, intranet, and extranet:

Internet

- Runs on TCP/IP protocol
- Global network of public networks, network access points (naps), and service providers
- Operated either for public access or private data exchange (with a VPN)

Intranet

- Internet-like logical network
- Based on an organization's internal physical network infrastructure
- TCP/IP and HTTP standards
- Web browsers

Extranet

- Private network using Internet protocols
- Accessible by partners and vendors outside of the organization, but not by the general public

Common Data Network Services

- File services
- Mail services
- Print services
- Client/server services
- Domain name service

Following are common services offered on a network:

- File services provide data files on file servers.
- Mail services provide for e-mail internally or externally.
- Print service supports printing to a shared printer.
- Client/server services allocate computing resources among computers.
- Domain name service matches Internet uniform resource locator (URL) requests with the actual address or location of the server.

Types of Networks

- LANs (local area networks)
 - Discrete network designed to operate in a specific, limited geographic area (a single building or part of a building)
 - Connect workstations and file servers to provide sharing of network resources (printers, e-mail, and files)
- MANs (metropolitan area networks)
- WANs (wide area networks)
 - Network of subnetworks
 - Physically or logically interconnecting LANs over a large geographic area
 - Supports multiple communication protocols

Types of Computer Networks

A *local area network* (LAN) is a relatively small network confined to a small geographic area, such as a single office or a building. Laptops, desktops, servers, printers, and other networked devices that make up a LAN are located relatively close to each other.

The term *metropolitan area network* (MAN) is typically used to describe a network that spans a city-wide area or a town. MANs are larger than traditional LANs and predominantly use high-speed media, such as fiber-optic cable, for their backbones.

A *wide area network* (WAN) covers a significantly larger geographic area than LANs or MANs. A WAN may use public networks, telephone lines, and leased lines to tie together smaller networks over a geographically dispersed area.

LAN Transmission Methods

- **Unicast**
 - The packet is transmitted from the source to a single network destination address.
- **Multicast**
 - The packet is transmitted from a source to multiple, selected destination network addresses,
- **Broadcast**
 - The packet is transmitted from a source to all network addresses.

Following are three general ways that packets can be transmitted:

- **Unicast:** The packet is transmitted from a source to a single network destination address.
- **Multicast:** The packet is transmitted from a source to multiple, selected destination network addresses.
- **Broadcast:** The packet is transmitted from a source to all network addresses.

Physical versus Logical Topologies

- Physical topology
 - Defines how systems are connected together
 - Bus, ring, and star
- Logical topology
 - Defines the rules of communication across the logical topology
 - Ethernet, Fiber Distributed Data Interface (FDDI), and ATM

LAN Topologies and Protocols

There are several ways in which systems on a LAN can be interconnected, and there are several techniques you can employ to send signals to each other. This section examines some of the most common patterns for wiring a network, and it discusses how the network's physical properties relate to communication protocols used by its systems. In the process, we discuss LAN topologies such as bus, ring, and star and take a close look at media access technologies such as Ethernet, Token Ring, and ATM.

Physical Topologies

A *physical topology* describes how the network is wired together. It is the layout of how systems are connected via cables or wireless devices. Wire-based physical topologies are relatively easy to visualize because they are interconnected according to simple geometric patterns. This section covers some of the most common topologies you may see on a LAN:

- Bus
- Ring
- Star

Of the topologies that we cover, star is probably the most commonly seen on modern networks. If you've ever set up a cable modem or DSL router for several systems at home or connected workstations to a switch or a hub, you probably used the star topology for your internal network. However, other topologies are still in use and may be appropriate for the particular requirements of your organization. Let's begin with the simplest physical topology, the bus topology.

Bus Topology

Bus topology:

- All network nodes are connected by a common media bus.
- Data traversing the bus passes through all nodes.
- Packets are inspected by each node to determine if the packet is addressed to that particular node.

Bus Topology

The simplicity of the physical bus topology is, perhaps, its biggest advantage, especially if it is used on a small network. All systems in a bus topology are attached to the same cable segment. Using a bus topology to wire systems is relatively inexpensive when they are in close proximity. For example, if five workstations are located in one room, and the server is in an adjacent room, only a single cable needs to run between the two rooms. There is no need to provide individual cables for every system.

On the flip side, the bus topology is dated and used rarely today. Because all systems in a bus topology are attached to the same wire, the following limitations severely outweigh this topology's benefits:

- Low fault tolerance
- Poor traffic isolation
- Limited scalability

A break in the cable that interconnects systems on the bus impacts the entire network. This single point of failure significantly reduces reliability of networks based on a physical bus topology. Furthermore, having a single wire segment makes it hard to isolate and troubleshoot such problems. It is also difficult to expand such a network because adding a new system may require replacing the wire of the bus with a longer one. These disadvantages have inspired the adoption of other physical topologies.

Ring Topology

- The nodes on a network are connected by a medium in a closed, unidirectional loop.
- Multiple point-to-point connections form a ring."
- Systems transmit on one side and receive on another.
 - Dual ring can provide fault tolerance.

Ring Topology

Another physical topology that you may encounter follows the ring pattern, as illustrated in this slide. An easy way to think of the ring topology is to envision a group of kids standing in a circle, playing the telephone game. In the game, a message is passed from child to child around the circle by whispering into a neighbor's ear. Similarly, each system that is part of the physical ring topology transmits messages on one side and receives messages on the other. Each node in the ring is directly connected to two other nodes, one on each side of it. Messages travel around the ring from node to node in a closed loop. Unfortunately, if one of the cables in the loop develops a problem, this is likely to disrupt communications of the whole ring-based network segment. Essentially a ring topology is just a bus topology where you take the two ends and connect them together. Therefore, most of the disadvantages with a ring topology are also present in a bus topology.

The ring topology has a number of disadvantages that account for its relatively low popularity:

- Fault-tolerance limitations
- Cost of deployment
- Lack of Ethernet support

A physical ring topology can usually span larger distances than the bus topology because the network card of each ring node regenerates the signal when passing it along the loop. However, support for this functionality increases the cost of setting up the ring. Regardless of the cost, the major reason for the lack of popularity of this pattern is that it is not supported by Ethernet. Because Ethernet came out on top in the protocol wars, companies shy away from using rings because they cannot run Ethernet on them. However, Token Ring and FDDI can both utilize ring topologies and they provide the slight advantage of giving each station an equal amount of time to communicate.

Star Topology

- All network nodes are directly connected to a central host.
- High wiring costs for large installations.
- Multiple point-to-point connections to a central device (hub or switch):
 - Good fault tolerance
 - Traffic isolation provided by certain hardware
 - Scales well

Star Topology

Star is the most common physical topology in use today. All systems in this topology are connected directly to a central device, such as a hub or switch. A node that wants to send a message to another system on the star network directs the message to the central connection point, which is usually a hub or a switch, and it then relays the message to the appropriate recipient.

The star-wiring pattern helps provide fault isolation. If the cable leading to an individual system is faulty, the other systems can still exchange data. This is a significant improvement over physical bus and ring topologies, which can be impaired due to a problem with a single wire. However, this provides only fault tolerance from a faulty wire. If a computer has a faulty NIC (network interface card), an entire segment can still be flooded. The reliance on the central device in the star topology does create a single point of failure; however, a hub failure generally is easier to troubleshoot than cable-related problems (which undermine bus and ring topologies).

The main disadvantage of a star topology is probably the need to have a dedicated cable segment for each system. The total cost of wiring may become particularly evident if the networked systems are located far from the hub. For each system that needs to communicate over the network, a wire has to be run from the new node to the central location. In practice, however, the cost of running new cable in star topology usually is not sufficiently large enough to outweigh the ease with which new nodes can be added to the network and the fault tolerance that this pattern provides.

Traffic control capabilities in a star network are significantly better than that of the other physical topologies discussed. Because all circuits of the star are tied to a single device, the device can manage the flow of data between systems connected to it. To summarize, the following advantages of a star topology put it in the lead of the other physical topologies:

- Reasonable fault tolerance
- Scalability and ease of expansion
- Support for traffic isolation

Tree Topology

Tree topology:

- Branches with multiple nodes

The tree topology has multiple branches with nodes on each branch.

Mesh Topology

Mesh topology:

- All network nodes are connected to every other node.

The mesh topology has the highest level of redundancy because each node is connected to every other node.

Logical Topologies

- Independent of physical topologies
- Logical topologies
 - Ethernet
 - Token Ring
 - Fiber Distributed Data Interface (FDDI)
 - Asynchronous Transfer Mode (ATM)
 - High-Level Data Link Control (HDLC)
 - Integrated Services Digital Network (ISDN)
 - X.25

Logical Topologies

After the systems have been interconnected, they need to know the rules for sending signals to each other. These rules are specified by media access protocols, which are examined in this section:

- Ethernet
- Token Ring
- FDDI
- ATM

These protocols are responsible for making sure that a signal sent by a system finds its way to its destination. The process that the protocol follows to send data over the cable, regardless of how it is physically wired, can be described using a *logical topology*. There are three common logical topologies, which actually have the same names and properties as physical topologies: bus, ring, and star.

Note

A logical topology describes how a signal travels across the wires, which have been arranged according to a physical topology.

Physical and logical topologies generally are independent of each other. As you will see, a Token Ring network, which uses a logical ring topology, is usually wired according to a physical star topology. There often is a relationship between a physical and a logical topology that results in some pairings being used more often than others.

Note

Regardless of the logical topology choice, the underlying physical topology that describes how the wires are connected can be different.

LAN Transmission Protocols

- Carrier sense multiple access (CSMA)
 - Computer continuously monitors the common transmission line
 - Transmits when the line appears to be unused
 - If the transmission conflicts with another transmission, one of the following two behaviors occur:
 1. **Persistent carrier sense:** If there is no acknowledgment from the destination that it received the packet from, the computer assumes a collision has occurred and resends the packet.
 2. **Non-persistent carrier sense:** The computer waits a random amount of time before resending the packet.

The objective is to use it on wire to transmit multiple messages that can desire to transmit at any time. There are different methods of arbitrating or assigning who is going to transmit at what time.

LAN Transmission Protocols (2)

- Carrier-sense multiple access with collision avoidance (CSMA/CA)
 - Computers are attached to two coaxial cables.
 - Each cable carries data signals in one direction only.
 - Computer monitors its receive cable to see if the carrier is busy.
 - Computer transmits on its transmit cable if no carrier busy was detected.
 - Precedence based on pre-established tables

The following are key characteristics of CSMA/CA:

- Computers are attached to two coaxial cables.
- Each cable carries data signals in one direction only.
- The computer monitors its receive cable to see if the carrier is busy.
- The computer transmits on its transmit cable if no carrier was detected.
- The precedence is based on pre-established tables.

LAN Transmission Protocols (3)

- Carrier-sense multiple access with collision detection (CSMA/CD)
 - Prior to transmitting, the computer monitors common cable to determine if another computer is transmitting.
 - If no traffic is detected, the computer transmits.
 - Computer monitors the line to see if there was a collision with another computer's transmission at the same time
 - If the computer detects a collision, it transmits an extended jam signal that causes all nodes on the segment to stop sending data.
 - Computers respond to jam signal by waiting a random amount of time before attempting to transmit again.

The following are key characteristics of CSMA/CD:

- Prior to transmitting, the computer monitors common cable to determine if another computer is transmitting.
- If no traffic is detected, the computer transmits.
- The computer monitors the line to see if there was a collision with another computer's transmission at the same time.
- If the computer detects a collision, it transmits an extended jam signal that causes all nodes on the segment to stop sending data.
- Computers respond to a jam signal by waiting a random amount of time before attempting to transmit again.

LAN Transmission Protocols (4)

- Polling
 - Secondary computers (polled devices) are assigned a specific period of time to transmit by a primary computer.
 - If a secondary computer does not have any data to transmit, the primary computer moves on and provides another secondary computer with the opportunity to transmit.

Polling is commonly used in mainframe environments.

The primary computer can assign higher priorities to specific secondary computers by providing them with the opportunity to transmit more frequently than other secondary computers.

Ethernet

- Ethernet is "baseband" or shared media.
- Only one station is allowed to transmit at any given time within a single collision domain.
- All stations are required to listen before they transmit.
- All stations are required to monitor their transmission to check for collisions.
- Data transmitted using CSMA/CD bus technology.
- Three cable standards:
 - Thinnet - 10 base2, 10 mbps
 - Thicknet - Known as 10 base5, 10 mbps
 - Unshielded twisted pair 10 base t, 10 mbps
 - Three types:
 - 10 base t -10 mbps,
 - 100 base t (Fast Ethernet) -100 mbps
 - 1000 base t (Gigabit Ethernet) - 1 gbps

Ethernet

Ethernet is by far the most popular media access protocol currently used on LANs. A chunk of data transmitted by Ethernet over the wire is called *a frame*. On an Ethernet network, only a single node should transmit a frame at a time. If multiple systems transmit simultaneously, a *collision* will occur, which can cause both signals to fail and require the systems to retransmit their frames.

To keep the number of collisions to a minimum, a system is required to check whether anyone else is already transmitting before placing a frame on the wire. If another system's signal is already on the wire, the system is expected to wait according to the algorithm designed to give each node a fair shot at using the network. If the line is clear, the system generates a signal and monitors the transmission to make sure there was no collision. These properties are summarized under Ethernet's designation as a carrier sense multiple access/collision detection (CSMA/CD) protocol.

Ethernet specifications actually define more than just protocols for sending signals over the wire. Other properties include cabling requirements for transferring data at desired rates and the maximum length of the wire segment. In addition, Ethernet standards specify which physical topology should be used for a particular type of Ethernet communication.

LAN Transmission Protocols

Token-passing:

- A small message frame is circulated around the network.
- Possession of the token grants a network device the right to transmit.
- When the device completes its transmission, it passes the token to the next computer in the network to receive the token.
- If a device on the network has no data to transmit, it passes the token to the next network device.
- Each device has a maximum amount of time it can hold the token.

Problems that must be handled on a token ring network include:

- The token is lost or destroyed for some reason.
- Multiple tokens are generated by mistake.

Token networks avoid the collision-handling requirements of CSMA approaches.

Token passing is amenable to networks supporting high-bandwidth applications.

Token Ring and ARCNET

- Communication is token-based.
- Each station conditions or amplifies the token as it is passed.
 - Devices on networks are connected in a ring formation
 - Possession of a token in the form a frame grants a device the right to transmit over the ring.
 - One node on the network is an active monitor to ensure proper operation of the network.
 - Loss of token
 - Multiple tokens
 - Typically, only one token is allowed per ring, but there may be two (early release).
- ARCNET
 - Token-passing network implementation
 - Star topology
 - Early LAN access method

Token Ring and ARCNET

Token Ring offers an alternative method to sending signals across the network media. Originally developed by IBM in the 1970s, it is still in use on some networks, although it is not as popular as Ethernet. Token Ring networks, classified as logical ring topology systems, can only communicate with their immediate neighbors, and the data travels in a closed loop. A specialized frame called a *token* is used to carry data around the Token Ring network. To prevent collisions, only the system that possesses the token is allowed to transmit to the network. (This is drastically different from the CSMA/CD technique employed by the Ethernet.) Each system receives and examines the token to see whether it contains data for that system. If there is such data, the system processes it and passes the token along to its immediate neighbor. If contents of the token are not destined for the system, it simply transmits the token to the next node.

Note

Each node on a Token Ring network amplifies the token as it passes it along the loop.

To begin communicating on the Token Ring network, a system puts data into an empty token and passes it to the neighbor. Eventually, the token makes its way to the destination and loops back to the sender. The originating system then removes the data, marks the token as empty, and passes it along the ring. When a node receives an empty token, it can fill it if it needs to send data or simply pass it to the neighboring system.

Although logically Token Ring is a ring topology, its wiring usually follows a physical star topology. In this configuration, systems that form the Token Ring network are connected to a central device called a *multistation access unit* (MSAU). To pass a token to a neighboring node, the system actually sends the signal to the MSAU, which retransmits it to the originator's neighbor.

FDDI

Fiber distributed data interface (FDDI):

- Token-passing media access method
- Fiber optic 100 mbps transmission rates
- Comprises two counter-rotating rings:
 - FDDI adds a second ring for fault tolerance.
 - One ring is active at any given time.
- Because data is represented by light pulses, FDDI is resistant to noise and other types of electromagnetic interference.

Fiber Distributed Data Interface (FDDI) is similar to Token Ring in that it uses a token to pass data along a logical ring. However, FDDI has a second ring for redundancy purposes. The second ring is not used for normal communications if the primary ring operates properly. If the primary ring fails, FDDI can switch to using the second ring. In any case, only one ring is used for communications at a given time.

Asynchronous Transfer Mode (ATM)

- Encapsulates common protocols
- Uses virtual path identifiers (VPI) to create end-to-end connectivity
- Uses a fixed data cell size (48 bytes) for better quality of service (QoS)
- Like combining Ethernet and IP

ATM

ATM provides yet another way for sending signals over the wire. Because ATM is relatively expensive to set up, it is not frequently seen on LANs. However, its traffic predictability and support for high bandwidth make it a good fit for networks that need to carry low-latency traffic, such as video streaming. ATM is more commonly used for establishing high-speed backbones that interconnect smaller networks and can carry signals over significant distances, as discussed in the "WAN Technologies" section.

ATM has properties attributable to media access technologies, such as Ethernet and Token Ring, as well as properties of higher-level protocols, such as IP and IPX. If you don't have a pure ATM environment, you can still use it to encapsulate traffic based on other protocols. ATM's capability to encapsulate a wide range of network protocols allows it to be integrated with most existing WAN and LAN implementations.

ATM is connection-oriented. This means that before systems can communicate over an ATM network, they must establish a virtual circuit between each other. The circuit can span across multiple ATM switches that also handle communications for other systems at the same time. The circuit is considered to be "virtual" because its communication channel traverses a shared network medium.

The virtual circuit is torn down at the end of the connection. This concept is similar to the way telephone calls are established. When you dial a number, the phone company sets up a virtual circuit from your phone to the phone of the person whom you are calling. The telephone circuit between the two phones ceases to exist when the call is complete.

802.11

- 802.11 standard supports two physical layers:
 - Infrared
 - Radio frequency
 - FHSS (Frequency Hopping Spread Spectrum)
 - DSSS (Direct Sequence Spread Spectrum)
- Branched into 802.11a, 802.11b, and 802.11g
- 802.11b supports up to 11 Mbps at 2.4 GHz
- 802.11a supports up to 54 Mbps at 5 GHz
- 802.11g supports up to 54 Mbps at 2.4 GHz

802.11

The www.extremetech.com website provides a good introduction to 802.11. The information here is reproduced from a few papers found at the website. 802.11b is almost always the protocol in use on wireless LANs:

"IEEE 802.11b is the most common and established wireless network protocol in use today, referred to as the IEEE 802.11b standard. The 802.11b standard defines, among other things, the radio frequency bandwidth wireless signals can use, throughput rates over that signal, and how wireless endpoints communicate with one another.

802.11b signals function in the 2.4000 GHz to 2.4835 GHz range, and have a maximum theoretical throughput of 11 Mbps (though testing suggests that actual throughput is more like 4-6 Mbps) and can even step down to 5.5 Mbps, 2 Mbps, and 1 Mbps to allow a more robust signal. 802.11b uses only Direct Sequence Spread Spectrum (DSSS) radio signaling, as opposed to Frequency Hopping Spread Spectrum (FHSS), which was part of the original 802.11 specifications. DSSS allows for greater throughput, but is more susceptible to radio signal interference. Interestingly, many DSSS-based 802.11 products are interoperable with current 802.11b networks, but only at 802.11b's 2 Mbps or 1 Mbps. Wireless endpoints have a coverage area that depends on antenna strength and the ability and clarity of the local environment to transmit radio signals - typically ranging from 75 to 150 feet for an office environment."

—<http://www.extremetech.com/article/0,3396,s%253D1034%2526a%253D13521,00.asp>

"Tests of products that use the new 802.11a standard shows that 802.11a networks are as much as five times faster than 802.11b nets, providing an average throughput of 28 Mbps in practice. The problem with 802.11a is that it is not interoperable with previously deployed 802.11b products due to the difference in frequency bands."

—<http://www.eweek.com/article/0,3658,s%253D7142526a%253D21038,00.asp>

WAN Devices

- Routers
- Multiplexers
 - Connect one of a number of input signals to an output

Binary input on select lines will route one of four input lines to the output in the above example.

WAN Devices (2)

- WAN switches
 - Use data link layer
 - Send data over public network circuits
 - Multi-port device
 - Operate at the data link layer
- Access servers
 - Asynchronous device
 - Provides dial-in and dial-out connections to the network
- Modem
 - Modulator/demodulator
 - Converts digital signals to analog signals
 - Transmits analog signals over conventional telephone lines
 - At the receiving end, analog signals are converted back to digital signals.

The following are key attributes of WAN switches, access servers, and modems:

WAN switches

- Use data link layer
- Send data over public network circuits
- Multi-port device
- Operate at the data link layer

Access servers

- Asynchronous device
- Provides dial-in and dial-out connections to the network

Modem

- Modulator/demodulator
- Converts digital signals to analog signals
- Transmits analog signals over conventional telephone lines
- At receiving end, analog signals are converted back to digital signals

Circuit-Switched Networks

- Switches establish a dedicated physical circuit between sender and receiver for a communication session.
- Preceded packet-switching technology
- Ideal for communications that need to have a constant connection
- Provides a single transmission path

On a circuit-switched network, switches establish a dedicated physical circuit between sender and receiver for a communication session. This technology preceded packet-switching technology and is ideal for communications that need to have a constant connection.

Packet-Switched Network

- Data to be transmitted is partitioned into packets.
- Packets are assigned sequence numbers as they are transmitted.
- Packets are sent to destination through router.
- Router tries to establish the best route.
- Packets are reassembled at the destination based on originally assigned sequence numbers.
- If a path is not available, the packet is routed to the destination through a different path.
- Fault-tolerant network

Packet-switched networks are good for bursty communications. Data to be transmitted is partitioned into packets, where each packet is assigned sequence numbers as they are transmitted. These packets are sent to the destination through a router, which tries to establish the best route.

Packets are reassembled at the destination based on originally assigned sequence numbers. If a path is not available, the packet is routed to a destination through a different path.

Virtual Circuits

- Virtual circuits (VC)
 - Path through intermediate devices and bridges to set up communication with a partner station
 - Path used for duration of session

Virtual circuits are unlike a bridged network where forwarding decisions are made on a frame by frame basis and there is no concept of a communication session. The path is through intermediate devices and bridges to set up communication with a partner station, and the path used is for the duration of the session.

Virtual Circuits (2)

- **Switched virtual circuits (SVC)**
 - Dynamically established as required
 - Disconnected when transmission is complete
 - Three phases:
 - Circuit establishment
 - Data transfer
 - Circuit termination
- **Permanent virtual circuits (PVC)**
 - Permanently connected
 - Eliminates overhead associated with circuit establishment and break down

There are two general types of virtual circuits: SVC and PVC.

Switched virtual circuits (SVC)

- Dynamically established as required
- Disconnected when transmission is complete
- Three phases
 - Circuit establishment
 - Data transfer
 - Circuit termination

Permanent virtual circuits (PVC)

Permanently connected

Eliminates overhead associated with circuit establishment and break down

WAN Technologies

- **Dedicated lines**
- **Frame Relay**
- **X.25**
- **HDLC and SDLC**
- **Voice over IP**
- **Integrate Services Digital Network**
- **DSL and cable modems**
- **SMDS**
- **ATM**
- **VOIP**

WAN Technologies

Token Ring and Ethernet protocols are great for sending signals over relatively short distances; however, they are not effective at carrying data between geographically distant sites. The connections among these greater distances comprise and are better served by a WAN, and this section takes a brief look at technologies designed for WAN communications:

Dedicated lines

Frame Relay and X.25

HDLC and SDLC

Voice over IP (VoIP)

ISDN, DSL, and cable modems

Private Circuit Technologies

- **Dedicated/leased line**
 - Dedicated line reserved by a communications carrier for the private use of a customer
 - Point-to-point link
- **Leased line types**
 - T1: DS-1 formatted data transmitted at 1.544 mbps through the telephone network.
 - T3: DS-3 formatted data transmitted at 44.736 mbps through the telephone network.
 - E1: Wide-area digital data transmission at 2.048 mbps (predominantly used in Europe).
 - E3: Wide-area digital data transmission at 34.368 mbps.

Memorize

With private circuits, organizations typically utilize either dedicated or leased lines.

Dedicated/leased line

- Dedicated line reserved by a communications carrier for the private use of a customer
- Point-to-point link

Leased line types

- T1: DS-1 formatted data transmitted at 1.544 mbps through the telephone network
- T3: DS-3 formatted data transmitted at 44.736 mbps through the telephone network
- E1: Wide-area digital data transmission at 2.048 mbps (predominantly used in Europe)
- E3: Wide-area digital data transmission at 34.368 mbps

Private Circuit Technologies (2)

- Serial line IP (SLIP)
 - De facto standard developed in 1984 to support TCP/IP-based asynchronous dial-up connections
 - Does not have error detection
 - Being replaced by point-to-point protocol (PPP)
- Point-to-point protocol (PPP)
 - Used for transmitting over dial-up and dedicated links
 - Allows multi-vendor operability
 - Improves on the Serial Line Internet Protocol (SLIP)
 - Builds on slip by adding login, password, and error correction
 - Data link layer protocol
 - Incorporates authentication methods:
 - Challenge handshake authentication protocol (chap)
 - Password authentication protocol (pap)

SLIP and PPP are provided by some ISPs for accessing the Internet.

Private Circuit Technologies (3)

- Extensible authentication protocol (EAP)
 - Authentication mechanism
 - Extension to PPP
 - Supports a variety of authentication mechanisms
 - New authentication methods used as desired
 - Defined in RFC 2284

EAP is used to provide authentication for a variety of applications, such as wireless. EAP is defined in RFC 2284.

X.25 and SMDS

- X.25
 - Used for packet switching
 - Robustness
 - Cost-effective
 - Standard
 - Point-to-point communication: Communication between DTE (data terminal equipment) and a DCE (data circuit-terminating equipment), usually a modem or communication between DTE and DSU/CSU (data service unit/channel service unit)
- Switched multimegabit data service (SMDS)
 - Connectionless
 - High speed
 - Uses packet-switched public networks
 - Applicable to WANs

X.25

X.25 is another packet-switching WAN protocol and is similar to Frame Relay. Actually, X.25 is a precursor to Frame Relay and was accepted as a standard by the International Telecommunication Union (ITU) in 1976. One of the major differences between the two protocols is that X.25 provides error checking, windowing, and retransmission services that are not available in Frame Relay. This difference weighs in favor of Frame Relay, making it a faster protocol than X.25 because of Frame Relay's lower transmission overhead. To compensate for deficiencies of WAN infrastructure, low-level error checking and correction capabilities that are considerably less relevant today were built in to X.25.

Both X.25 and Frame Relay protocols rely on two main types of equipment to establish WAN communications: data terminal equipment (DTE) and data communications equipment (DCE). The DTE connects the company's network to the X.25 or Frame Relay cloud and is usually owned by the customer. DCE devices are carrier-owned and provide switching and clocking services for transmitting data across the WAN network.

HDLC and SDLC

- High-Level Data Link Control
 - Controls data flow
 - Error correction
 - Operates at the data link layer, layer 2
 - Data encapsulation method
 - Uses synchronous serial links
- Synchronous Data Link Control
 - Operates at data link layer, layer 2
 - Created by IBM
 - Uses a polling media-access method
 - Primary station controls all communications with secondary stations

HDLC and SDLC

High-Level Data Link Control (HDLC) protocol is an ISO standard that supports point-to-point and multipoint communications. It is typically used by X.25 and Frame Relay to move packets across the WAN cloud. HDLC designates a primary station and a secondary station in its communications, and it can operate in three modes:

- **Normal response mode** (NRM), in which the primary station initiates communications with the secondary station. The secondary station transmits only as a responder when instructed by the primary station.
- **Asynchronous response mode** (ARM), in which the secondary station can transmit without explicit permission from the primary station and has the ability to initiate communications. The primary station retains the responsibility of error recovery, link setup, and link termination.
- **Asynchronous balanced mode** (ABM), in which both stations have equal responsibilities and can transmit and receive messages independently over a duplex line, such as X.25.

HDLC is an extension of the *Synchronous Data Link Control* (SDLC) protocol, which was developed by IBM in the 1970s. SDLC is mainly used in IBM's proprietary *Systems Network Architecture* (SNA) environments. Unlike HDLC, SDLC supports only the NRM mode of operation.

Voice over IP

- Combining data
- Cost-effective
- Redundancy
- Security issues
- Exposure

VoIP

Voice over IP (VoIP) technologies, sometimes described as IP telephony, enable the use of data networks for carrying voice communications. When connected to the Internet, VoIP equipment enables individuals to talk without using the traditional telephone network. Instead, sound is converted into digital form and is transported over the network as data. Because traffic conditions on the Internet are irregular, the quality of sound on the receiving end depends on the state of the connection between the conversing parties. For greater reliability, companies may send VoIP packets over private WAN links, such as point-to-point T lines, ATMs, or Frame Relays.

Organizations that already have invested in a high-capacity LAN and WAN infrastructure can use that same network to carry voice, video, and data traffic. Companies that want to add a level of redundancy to voice communication systems can deploy standard telephone lines in parallel with VoIP. In addition, it is common for companies to use traditional telephone lines for interfacing with the outside world and to deploy IP telephony for communicating within the company.

Security Issues with Telephony

Do not forget to account for information security when deploying or maintaining your company's telephone system, regardless of whether it is traditional or IP-based. One of the most prominent threats in this area is the abuse of Private Branch Exchange (PBX) systems, which frequently are relied upon to support telephone services within a company. PBX systems often contain maintenance hooks for providing remote maintenance capabilities over the phone line. If an attacker learns the number for connecting to the backdoor and is able to authenticate using the PBX's default password, the attacker may be able to make calls on the company's account or access sensitive voicemail messages. Furthermore, when a PBX is connected to the company's data network, the PBX system may act as a gateway to internal computer systems.

PBXs are not the only concern for a security practitioner, though; faxes and fax machines bring forth issues as well. For instance, a person may receive a sensitive fax and leave it on the machine—for the cleaning crew or anyone else to view. A more recent development is the use of multipurpose fax machines that, in addition to faxing, can copy, print, and even scan documents. It is unclear how robustly these mechanisms are separated from each other within the device. If the machine is connected to the network, it may act as a gateway to the company's internal systems in a manner similar to a PBX.

ISDN

- Integrated Services Digital Network
- Digitization of the telephone network
- Voice and digital data transmitted over existing telephone wires

ISDN, DSL, and Cable Modems

ISDN and DSL are offered by telephone companies as consumer-oriented technologies to provide high-speed Internet connectivity over existing telephone lines. These signals are carried over copper wires that traditionally have been used solely for voice communications. DSL, and to some extent ISDN, devices compete with cable modems that are offered by television cable operators. Cable modems rely on existing coaxial and fiber lines that have been used to carry TV signals to subscribers of the cable services.

The widespread use of ISDN did not take place, mainly because of high connectivity costs coupled with advancements in DSL and cable modem technologies. The use of DSL is not limited to home users, however, and is gaining acceptance in businesses. High-end connectivity options for DSL are capable of sustaining bandwidth equivalent to a T1 and usually are less expensive. However, unlike T1 lines, which can be used to establish point-to-point links between the company's sites, DSL lines are used to connect only a single site to the Internet.

DSL

- Digital subscriber line (DSL)
 - Point-to-point public network
 - Uses existing twisted pair copper telephone lines
 - Provides high-bandwidth data service
 - Different types available
- DSL types
 - Asymmetric digital subscriber line (ADSL) *
 - Single-line digital subscriber line (SDSL)
 - High-rate digital subscriber line (HDSL)
 - Very-high-data-rate digital subscriber line (VDSL)

DSL is a point-to-point network that uses existing phone lines.

There are four general types:

- Asymmetric digital subscriber line (ADSL) *
- Single-line digital subscriber line (SDSL)
- High-rate digital subscriber line (HDSL)
- Very-high-data-rate digital subscriber line (VDSL)

ADSL and SDSL

- Asymmetric digital subscriber line (ADSL)
 - Has higher download data rates than upload
 - Downstream data rates: 1.5 mbps to 9 mbps
 - Upstream data rates: 16 kbps to 640 kbps
 - Maximum distance: 18,000 ft
- Single-line digital subscriber line (SDSL)
 - Symmetrical download and upload rates
 - 1.544 mbps
 - Single twisted pair used for both directions
 - Operating range: 10,000 ft

Asymmetric digital subscriber line (ADSL)

- Has higher download data rates than upload
- Downstream data rates: 1.5 mbps to 9 mbps
- Upstream data rates: 16 kbps to 640 kbps
- Max distance: 18,000ft

Single-line digital subscriber line (SDSL)

- Symmetrical download and upload rates
- 1.544 mbps
- Single twisted pair used for both directions
- Operating range: 10,000 ft

HDSL and VDSL

- High-rate digital subscriber line (HDSL)
 - Symmetrical download and upload rates
 - 1.544 mbps
 - Uses two copper twisted pairs
 - Used by local phone companies to provide tl service
 - Operating range: 12,000 ft
- Very-high-data-rate digital subscriber line (VDSL)
 - Downstream data rate: 13 mbps to 52 mbps
 - Upstream data rate: 1.5 to 2.3 mbps upstream
 - Operating range: 1,000 ft to 4,500 ft

High-rate digital subscriber line (HDSL)

- Symmetrical download and upload rates
- 1.544 mbps
- Uses two copper twisted pairs
- Used by local phone companies to provide tl service
- Operating range: 12,000 ft

Very-high-data-rate digital subscriber line (VDSL)

- Downstream data rate : 13 mbps to 52 mbps
- Upstream data rate : 1.5 to 2.3 mbps upstream
- Operating range : 1,000 ft to 4,500 ft

Cable Modem

- Cable modem
 - The cable company provides them.
 - They provide a broadband Internet connection.
 - Cable modems in a geographical area share a single coaxial cable to access the Internet.
 - The data rate is a function of the number of concurrent users.
 - Operating range: 1,000 ft to 4,500 ft

Cable modem

- Provided by cable company.
- Provides broadband Internet connection.
- Cable modems in a geographical area share a single coaxial cable to access the Internet.
- Data rate is a function of the number of concurrent users.
- Operating range: 1,000 ft to 4,500 ft

LAN Cabling

- Coaxial cable
 - Used for high speed analog and digital transmission with high immunity to interference
 - 50-ohm cable for digital signaling
 - 75-ohm cable for high speed data and analog signals
 - Can transmit for longer distances without amplification
- Coaxial cable transmission schemes
 - Baseband: Single channel of information
 - Broadband: Multiple channels of information
- Twisted-pair
 - Relatively low-speed transmission medium
 - Two insulated wires twisted in opposing rotations
 - Counter wrapping provides for cancellation of common mode noise and interference
 - Two types: Shielded (STP) or unshielded (UTP)

Network Cabling

Cables are the primary means of carrying data on computer networks. Twisted-pair, coax, and fiber-optic cable are some of the more common cable types. The vast majority of network implementations still use unshielded twisted-pair (UTP) cables.

UTP cabling is frequently used for connecting telephones and computers at homes and offices and consists of at least one pair of ordinary copper wires. The wires in a pair are insulated and twisted around each other with each pair at a different twist rate to reduce the effects of electromagnetic interference. This is why this cable type is referred to as "twisted pair."

It also is possible to enclose the wires in a special shielding to further protect them from electromagnetic interference. This type of cable is called shielded twisted-pair (STP). The extra shielding makes UTP cables unwieldy, and you will not see it often in network installations. UTP cable does not have the extra shielding that is used in STP cables.

LAN Cabling (2)

- Fiber optic cable
 - Cable comprising bundles of optical fibers
 - Information transmitted as light signals
 - Resistant to electromagnetic interference
- Cross-over cable
 - Wire two Ethernet devices without hub, switch, or bridge
 - Two and only two devices
 - Cross:
 - +TX to +RX
 - -TXto-RX

Fiber optic cable

- Cable comprising bundles of optical fibers
- Information transmitted as light signals
- Resistant to electromagnetic interference

Cross-over cable

- Wire 2 Ethernet devices without hub, switch, or bridge
- 2 and only 2 devices
- Cross:
 - +TXto+RX
 - TX to -RX

Twisted-Pair Cabling Categories

Unshielded twisted pair types:

- Category 1
 - Standard telephone wiring
- Category 2
 - 4mbps
 - EIA/TIA-586 standard
- Category 3
 - 10mbps
 - Applied in 10base t networks
- Category 4
 - 16 mbps
 - Applied in token ring networks
- Category 5
 - 100 mbps
 - Was standard for local area networks
- Category 6
 - 155 mbps
 - Now being specified instead of category 5
- Category 7
 - Specification of up to 1 Gbps

UTP Cable Categories

Network wiring, particularly UTP cable, is categorized according to the bandwidth that it can sustain. The following table lists standard categories of UTP cable and outlines their capacities. (In this context, a cable category is typically referred to as *CAT*.) Capacity of a UTP cable mostly depends on the number of wire twists per segment. More twists provide better interference isolation, which allows the cable to sustain higher throughput rates. Unfortunately, additional twists also increase the cost of the cable.

Capacities of UTP Cable Categories

Cable Category	Capacity
Category 1 and 2	Voice, low-speed data
Category 3	Data 10 Mbps
Category 4	Data 16 Mbps
Category 5	Data 100 Mbps to 1 Gbps

Analog Versus Digital



Analog

- Analog signals are representative of most real world situations.
- Analog systems are subject to interference, noise, and distortion during amplification and processing.
- Analog signals can be sampled at regular intervals and the sampled values can be represented by numbers, then the numbers can be processed in digital systems.
- Processing digital representations provides high immunity to noise and interference.

Asynchronous Communications

- Data is sent by changes in levels of voltage or current in a sequential fashion.
- Start bit or bits indicate the beginning of the sequence.
- Stop bits indicate the end of the sequence.
- Transmitting and receiving equipment operate at the same data rate.

Modems and dial-up remote devices operate asynchronously.

With asynchronous communications, data is sent by changes in levels of voltage or current in a sequential fashion.

Start bit or bits indicate the beginning of the sequence and stop bits indicate the end of the sequence.

Transmitting and receiving equipment operate at the same data rate.

Synchronous Communications

- Transmitting and receiving stations are synchronized.
- Each unit of data, usually a byte, does not need a start or stop bits.
- Transmission is more efficient (less overhead).
- Transmits at high data rates are synchronized with electronic clock signals.

Synchronous is more efficient and higher speed than asynchronous.

Data sent by changes in levels of voltage or current are in a sequential fashion.

Start bit or bits indicate the beginning of the sequence.

Stop bits indicate the end of the sequence.

Transmitting and receiving equipment operate at the same data rate.

Network Devices

- **Hubs**
- **Switches**
- **Bridges**
- **Routers**
- **CSU/DSU**

Network Devices

Several types of devices are commonly used at the core of the network to provide a reliable and flexible communication medium. This section looks at several network devices to make sure that you understand what they are and when they should be used:

Hubs

Bridges

Switches

Routers

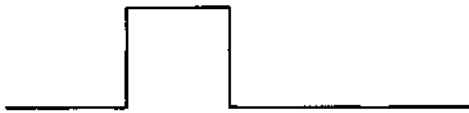
CSU/DSU

Repeaters

Repeaters:

- Signals deteriorate with distance.
- Repeaters amplify and shape signals before retransmitting.

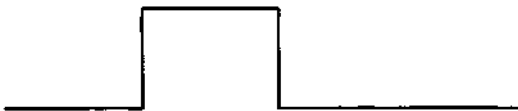
Repeaters do not add intelligence or process data; they reshape a pulse representing data that starting out looking like the original.



AND AFTER TRAVELING A LONG DISTANCE OVER WIRES WITH CUMULATIVE CAPACITANCE, IT LOOKS LIKE THIS:



A REPEATER RESHAPES IT TO THIS:



Hubs

- Most common network hardware
- Replicate or "amplify" data
- Provides no traffic control
- Concentrator
- Connects multiple LAN devices together
- Operates as a multi-port repeater
- No security

Hubs

A hub is one of the simplest devices you will find on a network. To build a basic LAN, simply connect the network interface cards (NICs) of your systems to ports on the hub via straight-through cables and, voila, you are networked! Hubs can vary in size from 4-port devices, used for home office networks, to chassis hubs that support the insertion of multiple 12-port cards for interconnecting large enterprises.

A hub operates by "repeating" data that it receives on one port to its other ports. As a result, a data frame transmitted by one system is retransmitted to all other systems connected to the hub. A classic hub does not have traffic-monitoring capabilities and cannot control which ports should or should not receive the frame, forming a large collision domain. This property of a hub has significant security implications, because a system connected to the hub may be able to intercept a data frame destined for someone else.

The process of gleaning data from a network transmission is called *sniffing*. For now, it will suffice to say that if you rely on hubs to connect your network segments, then your network is a sniffer's dream.

Note

You may find it useful to have a simple 2- or 4-port hub in your networking toolkit. A basic hub may come in handy if you want to "tap" into a data stream with a sniffer to troubleshoot a networking problem or to ensure that security mechanisms function as expected. With the proliferation of switches, though, hubs are becoming harder to find. Even if the device claims to be a hub, it may be a good idea to verify that it is not actually an inexpensive switch.

Switch

- Sometimes called a switching hub.
- Send the data packet only to the specific port where the destination MAC address is located, rather than all ports attached to the hub or bridge.
- Switches can be considered fast multi-port bridges.
- Primarily operate at layer 2, the data link layer.
- Mixture of hub and bridge technology.
- A switch will keep track of the MAC addresses attached to each of its ports.
- Makes sniffing ineffective.
- Layer 3 switch
 - Operates at network layer
 - Operates as a router
 - Uses routing tables
- Virtual LAN (VLAN)
 - Logical grouping of ports on switches
 - LAN traffic limited to members of the group
 - A single broadcast domain

Switches

A network switch combines the functionality of a hub and a bridge into a single device. If you think of a switch as a bridge with more than two ports, you'll get the idea. Like a hub, a switch can retransmit data to multiple ports. In addition, an Ethernet switch keeps track of MAC addresses attached to each of its ports, which grants it the traffic control capabilities of a bridge. By monitoring and controlling traffic between its ports, a switch will only direct a data frame to the system or network segment for which it is destined, narrowing each port to its own collision domain.

To speed up forwarding of data across ports, some switches employ the technique called *cut-through* switching. In cut-through switching, the device only reads the initial portion of the frame to obtain its destination MAC address and immediately forwards the frame to the appropriate port. A slower but sometimes more reliable alternative called *store-and-forward* reads the entire frame, verifies its integrity, and only then directs it to the destination.

Because a switch typically does not replicate frames to all ports, it offers a powerful way to defend against sniffing attacks. If your network is set up so that every system is connected to a dedicated port of the switch, a sniffer's field of vision is severely restrained. On a fully switched network, each system will usually see just traffic that is destined for it and will not be able to intercept other peer-to-peer traffic on the segment.

Bridges

- Learn where all the network systems are located.
 - Construct a table, listing which MAC addresses are off of each port
- Useful in breaking a large LAN into smaller LANs connected by bridges
- Connect LANs with the same technology
- Operate at layer 2, the data link layer
- Amplify the data signals
- Sends data according to media access control address

Bridges

A bridge is used to connect two physical segments of a network, much like an over-the-water bridge connects two sections of a road. When a bridge receives a data frame on one of its ports, it makes a decision whether the data should be sent to the other port. This functionality allows a bridge to automatically control the flow of data between network segments that it connects.

To decide when to replicate frames from one port to another, the bridge learns which systems reside on which network segment. It accomplishes this by automatically recording the MAC addresses of frames that pass through it to construct a table that maps MAC addresses to network segments. If a bridge needs to process a frame destined to a MAC address that is not in the table, it forwards the frame anyway.

A bridge is moderately helpful at reducing the effectiveness of network sniffers because it splits the larger network that could be monitored by a sniffer into smaller segments. However, traffic to and from systems that exist on the same side of the bridge as the sniffer can still be intercepted. In practice, bridges are used less and less on modern networks; inexpensive switches are taking their place.

Routers

- Operate at layer 3, the network layer
- Basis of the Internet
- Examine IP source and destination addresses in packets and forward packets to the intended network
- Maintain tables with routing information that point to all reachable networks

Routers

Routers are often considered to be perimeter devices because they interconnect logical networks. A switch or a bridge, on the other hand, connects physical segments that reside on the same logical network. Much of the Internet relies on routers for determining what paths packets should take to get from one network to another. Like a switch or a bridge, a router makes decisions about where to direct data that passes through it. However, whereas a switch makes its decisions by tracking MAC addresses, a router operates on a layer higher by looking at IP addresses when forwarding packets.

Note

Routers are flexible devices and can handle a variety of protocols. In this section, however, the main focus is on routers that process IP traffic that originates from or is destined to an Ethernet-based network.

Overview of Numbering Systems

• Decimal	Sample Notation
-Base 10	123456
• Binary	
- Base 2	11110001001000000
• Hex	
-Base 16	1E240

Numbering Systems

Most human beings have 10 fingers and 10 toes, so it's probably only natural that our preferred system of counting has 10 digits as well. It's no coincidence that the English word for a single numeral is also the word for a single finger: *digit*. We refer to this numbering system as *base 10* because, well, it's based on 10 distinct digits. We're so used to base 10 that many people don't even know that there are other possibilities, although one hopes that not many IT professionals fall into this trap!

In reality, there are as many different bases as there are numbers with which to express them. If your friend Nancy comes up with a string of 387 distinct symbols, there's no reason why she couldn't compute something in base 387. A base is just a method of representing a number, which after all is just an abstract idea that really directly can't be written down anyway.

What Is a Protocol?

- Protocol
 - Standard set of rules.
 - Define the format and order of messages and actions taken upon receipt of the messages
- Network Protocol
 - Determine how computers communicate with each other
 - Describe a standard format and method for communication by adhering to a layered architecture model
- Layered Architecture
 - Divides networking processes in manageable layers
 - Supports interoperability through industry standards
 - Can modify one layer without affecting the others
 - Easier to understand communication functions

Protocols

In the broadest sense, *a. protocol* is nothing more than an agreement of how different entities will act and react in certain circumstances. A medical protocol prescribes a course of treatment for a certain disease. A diplomatic protocol is the basis for a formal treaty that, for example, may specify how two nations will allow free trade along a common border. Similarly, a communications protocol establishes the parties in an exchange of information. It dictates the format of such communication and also the allowable responses to various situations that can occur.

Real-Life Protocols

For clarity's sake, you might think of a protocol as a conversation, perhaps between more than just two parties. As an analogy, consider what happens when you approach the counter at your favorite coffee shop.

```
CLERK: Hello, may I take your order?  
YOU: I'd like a triple vente (large) latte.  
CLERK TO BARRISTA (coffee bartender): Order in.  
BARRISTA: Ready.  
CLERK TO BARRISTA: Triple vente latte.  
BARRISTA: Triple vente latte.  
CLERK TO YOU: That will be $4.13, please.  
YOU: [pay]
```

See how that worked? Standard corporate protocol dictates that the clerk first greet you and ask for your order, and then wait for you to reply. After the clerk hears your order, he turns to the barista and notified her to prepare to hear the order. The barista confirmed her readiness, and only then did the clerk pass along your order. The barista even confirmed that she heard the order correctly by repeating it back to the clerk. After the subtransaction with the barista was complete, the clerk could then turn back to you and ask for your money, which you cheerfully hand over.

Encapsulation

- Divide network communications into layers.
- Divide task of communication into pieces for easier implementation.
- Data encapsulation is the process of appending data around the information from one data packet to the data of another packet.
- Each layer encapsulates information around the packet it received from the layer immediately above it, then sent to the layer below.
- When the packet is received, the information that pertains to each layer is removed (stripped) from the packet as it works its way up the protocol stack.

Organizing Protocols into Stacks

We have just presented a very simple requirement for protocols. Computers must be able to communicate. However, the Internet is a very complex thing, and to meet that simple requirement we actually need a wide range of protocols for hardware, software, and communications media. The model we use to organize these protocols is called the *protocol stack*.

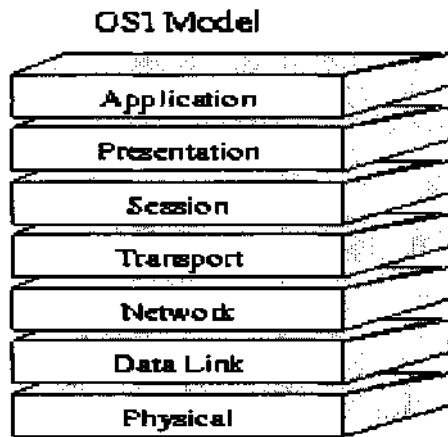
Imagine, if you will, a five-story apartment building. This building, however, is very special (or very strange, depending on your point of view). First of all, the really important things happen on the top floor, floor 5. Second, the only way the people on the fifth floor can get anything done is by asking the people on the floors below them to do it. For example, the people on the fifth floor want to eat dinner. They tell this to the people on the fourth floor. The people on the fourth floor figure out that dinner requires a soup, salad, main course, and dessert. They tell this to the people on the third floor. The people on the third floor decide that the courses will be onion soup, a garden salad, fish stew, and apple pie, and they tell this to the people on the second floor. The people on the second floor figure out what ingredients will be needed for this dinner (for example, broth, lettuce, vegetables, fish, etc.) and give this information to the people on the first floor. The people on the first floor actually go to the store, buy all the ingredients, and bring them back to the apartment building.

After the ingredients are purchased, the process goes in reverse. The first floor gives the raw ingredients to the second floor. The second floor checks that all the ingredients are there and then hands them off to the third floor. The third floor prepares the various courses by making the soup, tossing the salad, cooking the fish, and baking the pie. After all this is done, they hand the food off to the fourth floor. The fourth floor people package all the food up into nice courses and bring it up to the fifth floor residents, so they can eat a delicious meal.

In essence, that's how protocol stacks work. Protocol stacks divide network communications into different layers, like the floors in the apartment building. Each layer in the stack works on the packet in different ways. Some layers make sure the packet has all the information it needs; some layers make sure the packet is ready for an application to work with; and some layers make sure the packet gets on to the network properly. Each layer works directly with the layer above and below it, just as in the apartment building example. As packets are passed from one layer to the next, each layer examines or modifies the packet in some way. After the packet has reached the "ground floor" of the network, it is sent to its destination.

The use of protocol stacks in network communications makes the task of implementing protocols much easier. By making communications more modular, a service, process, or application need only concern itself with the layers it needs, leaving the other layers to someone else.

The Open Systems Interconnect (OSI) Model



- Defines how data and network information are communicated from one computer through the network media to another computer
- Comprises seven distinct layers:
 - Each layer has a unique set of properties.
 - Interfaces with its adjacent layers.

The Standard OSI Model

The standard reference model for protocol stacks is the International Standards Organization's (ISO) Open Systems Interconnection (OSI) model. The OSI model divides network communications into seven layers:

- The Physical Layer handles transmission across the physical media. This includes such things as electrical pulses on wires, connection specifications between the interface hardware, and the network cable and voltage regulation.
- The Data Link Layer connects the physical part of the network (for example, cables and electrical signals) with the abstract part (packets and data streams).
- The Network Layer handles interaction with the network address scheme and connectivity over multiple network segments. It describes how systems on different network segments find and communicate with each other.
- The Transport Layer actually interacts with your information and prepares it to be transmitted across the network. It is this layer that ensures reliable connectivity from end to end. The Transport Layer also handles the sequencing of packets in a transmission.
- The Session Layer handles the establishment and maintenance of connections between systems. It negotiates the connection, sets it up, maintains it, and makes sure that information exchanged across the connection is in sync on both sides.
- The Presentation Layer makes sure that the data sent from one side of the connection is received in a format that is useful to the other side. For example, if the sender compresses the data prior to transmission, the Presentation Layer on the receiving end would have to decompress it before the receiver could use it.
- The Application Layer interacts with the application to determine which network services will be required. When a program requires access to the network, the Application Layer will manage requests from the program to the other layers down the stack.

The Seven Layers of the OSI Model

- Application layer (layer 7)
 - Interface to the user
 - Deals with the communication aspects of an application
 - Establishes the availability of the intended communication partner
- Presentation layer (layer 6)
 - Presents data to the application layer
 - Provides translation services, such as EBCDIC to ASCII
 - Data encryption, compression, and decompression, are performed in this layer.
 - Defines how applications can enter the network

The Application Layer interacts with the application to determine which network services will be required. When a program requires access to the network, the Application Layer will manage requests from the program to the other layers down the stack.

The Presentation Layer makes sure that the data sent from one side of the connection is received in a format that is useful to the other side. For example, if the sender compresses the data prior to transmission, the Presentation Layer on the receiving end would have to decompress it before the receiver could use it.

The Seven Layers of the OSI Model (2)

- Session layer (layer 5)
 - Establishes initial contact with other computers
 - Establishes lines of communication
 - Maintains the session from end-to-end
 - Formats data for transfer between end nodes
 - Provides session restart and recovery
 - Supports simplex, half-duplex, and full-duplex communications
- Transport layer (layer 4)
 - Maintains end-to-end integrity
 - Maintains control of the session
 - Defines how to make connections between nodes
 - Defines managing the networking of messages
 - Establishes the logical connection between the sending host and receiving host

The Session Layer handles the establishment and maintenance of connections between systems. It negotiates the connection, sets it up, maintains it, and makes sure that information exchanged across the connection is in sync on both sides.

The Transport Layer actually interacts with your information and prepares it to be transmitted across the network. It is this layer that ensures reliable connectivity from end to end. The Transport Layer also handles the sequencing of packets in a transmission.

The Seven Layers of the OSI Model (3)

- Network layer (layer 3)
 - Transmits packets from the source network to the destination network
 - Performs packet routing
 - Performs error detection
 - Manages node traffic
- Data link layer (layer 2)
 - Sets up communications link between individual devices over a physical link or channel
 - Ensures that messages are delivered to the proper device
 - Translates messages received into bits for the physical layer
 - Formats the message into data frames
 - Inserts customized header containing the hardware destination and source address
 - Includes the media access control (MAC) sublayer

The Network Layer handles interaction with the network address scheme and connectivity over multiple network segments. It describes how systems on different network segments find and communicate with each other.

The Data Link Layer connects the physical part of the network (for instance, cables and electrical signals) with the abstract part (for instance, packets and data streams).

The Seven Layers of the OSI Model (4)

- Physical layer (layer 1)
 - Converts bits into electrical signals or light impulses for transmission
 - Defines the physical connection between the computer and the network
 - Specifies the electrical and mechanical aspects of the interface from the computer to a physical transmission medium, such as twisted-pair, coaxial cable, or fiber optic cable

The Physical Layer handles transmission across the physical media. This includes such things as electrical pulses on wires, connection specifications between the interface hardware, and the network cable and voltage regulation.

The TCP/IP Protocol Stack

- TCP/IP is a suite of protocols developed by the U.S. Department of defense in the 1970s for the ARPANET to support the development of reliable, long distance, interactive networks.
- TCP/IP is the protocol of the Internet.

The TCP/IP Stack

In comparison to the OSI protocol stack, the *Transmission Control Protocol/Internet Protocol* (TCP/IP) stack is much simpler. This model predates the OSI model and, as the name implies, is the underlying protocol of the Internet. As such, it's much more widely used than OSI-based protocols. In fact, although the stack usually is referred to as the TCP/IP stack, a more accurate name is IP stack. TCP is only one of the several protocols typically offered by an IP stack.

OSI versus TCP/IP

OSI	Application	7	Application	TCP/IP
	Presentation	6		
	Session	5		
	Transport	4	Transport (TCP)	
	Network	3	Internet (IP)	
	Data Link	2	Network	
	Physical	1		

Comparing the Two Models

This slide shows a comparison between the OSI model and the TCP/IP model. As you can see, the OSI model is more granular. The OSI model splits apart some functionality that was combined in the TCP/IP model. The Network Layer in the TCP/IP model comprises both the Physical Layer and the Link Layer in the OSI model, and the Application Layer in TCP/IP encompasses the Application, Presentation, and Session Layers of OSI. The OSI model is more detailed because it was designed to support protocols other than just TCP/IP. By creating more layers, the designers made it easier to break down the functionality of each protocol and build more specific interfaces and linkages between the layers.

Even though each model breaks down the functionality a bit differently, you should realize that no matter which model you use, it must perform all the functions required to take a piece of application data, place it into a packet, put that packet on the wire, and deliver it safely and efficiently to its destination.

TCP/IP Model

- Application layer
 - Combines application, presentation, and session layer functions of the OSI model
 - Applications communicate using sockets and ports
- Host-to-host layer - *ivM\$ptr+*
 - Similar to OSI transport layer
 - Provides packet sequencing
 - Defines protocols for setting up the level of transmission service
 - Provides for reliable end-to-end communications
 - Ensures the error-free delivery of data
 - Maintains the integrity of the data

Application layer

- Combines Application, Presentation, and Session layer functions of the OSI model
- Applications communicate using sockets and ports

Host-to-host layer

- Similar to OSI Transport layer
- Provides packet sequencing
- Defines protocols for setting up the level of transmission service
- Provides for reliable end-to-end communications
- Ensures the error-free delivery of data
- Maintains the integrity of the data.

HOST-TO-HOST LAYER PROTOCOLS INCLUDE:

TCP, UDP

TCP/IP Model (2)

- Internet layer
 - Equivalent to the OSI network layer
 - Manages network connections
 - Invokes protocols for the logical transmission of packets over the network
 - Assigns IP addresses to network nodes
- Network access layer
 - Equivalent of the data link and physical layers of the OSI model
 - Maps IP addresses to MAC addresses
 - Encapsulation of IP datagrams into transmission frames

Internet layer

- Equivalent to the OSI Network layer
- Manages network connections
- Invokes protocols for the logical transmission of packets over the network
- Assigns IP addresses to network nodes

Network access layer

- Equivalent of the Data Link and Physical layers of the OSI model
- Maps IP addresses to MAC addresses
- Encapsulation of IP datagrams into transmission frames

IP (Internet Protocol)

- Works at the IP Layer of the TCP/IP stack
- Deals with transmission of packets between end points
- The fundamental protocol of the Internet

The *Internet Protocol* (IP) is the protocol by which information is sent from one computer to another on the Internet. Each computer on the Internet has at least one IP address that uniquely identifies it from all other computers on the Internet. When you send or receive data (for example, an e-mail note or a web page), the message gets divided into little chunks called *packets*. Each of these packets contains both the sender's Internet address and the receiver's address. Any packet is sent first to a gateway computer that understands a small part of the Internet. The gateway computer reads the destination address and forwards the packet to an adjacent gateway that in turn reads the destination address, and so forth across the Internet until one gateway recognizes the packet as belonging to a computer within its immediate neighborhood or domain. That gateway then forwards the packet directly to the computer whose address is specified.

Because a message is divided into a number of packets, each packet can, if necessary, be sent by a different route across the Internet. Packets can arrive in a different order than the order they were sent in. The Internet Protocol just delivers them. It's up to another protocol, the *Transmission Control Protocol* (TCP) to put them back in the right order.

IP is a connectionless protocol, which means that there is no established connection between the endpoints that communicate. Each packet that travels through the Internet is treated as an independent unit of data without any relation to any other unit of data. (The reason the packets do get put in the right order is because of TCP; the connection-oriented protocol keeps track of the packet sequence in a message.)

The most widely used version of IP today is *Internet Protocol version 4* (IPv4). However, *IP version 6* (IPv6) is also beginning to be supported. IPv6 provides for much longer addresses and therefore the possibility of many more Internet users. IPv6 includes the capabilities of IPv4, and any server that can support IPv6 packets can also support IPv4 packets.

Packets Are Positional

0		15				31					
VER	IHL	TOS		Length in Bytes							
ID Field				Frag Offset							
TTL		Protocol		Header Checksum							
Source IP Address											
Destination IP Address											

IP header with no options shown, 20 bytes total

IP Headers in Detail

Here you see a diagram of how the bits inside an IP packet header are laid out. Pay particular attention to the way the diagram is labeled, because this is the standard way of looking at a packet header. Across the top, the bits are numbered from 0 on the left to 31 on the far right, for a total of 32 bits. 32 bits equal 4 bytes, and there are 5 rows, so you know that the total length of the header shown is 20 bytes. When dealing with packet headers, always start counting bits and bytes with 0. The first byte here is byte 0, and the last is byte 19.

Sockets

- Uniquely identify a connection
- Consist of the following:
 - Source IP address
 - Source port number
 - Destination IP address
 - Destination port number

Sockets

Taken as a pair, an address and a port number often are referred to as a *socket*. The socket tells a host's IP stack where to plug in a data stream so that it connects to the right application. Each stream actually has two sockets: a source and a destination. Together, this *socket pair* uniquely identifies the connection among all other connections on the network.

Addressing

- Every network and every computer must have an address.
- Every NIC card has a number burned in at the factory.
- Every software process that communicates must be addressable.

We all have many ways of identifying ourselves to other people depending on what type of communication we want with them. If we meet someone on the street, we use names. If we call someone on the phone, we use his phone number. If we mail a letter, we use the person's street address. Computers have identifiers as well. In fact, like humans, most computers have several of them depending on the type of communication.

Addressing (2)

- Used to uniquely identify a computer or network
- A device may have multiple types of addresses:
 - MAC address (00:60:1D:F0:EA:AF)
 - IP address (192.168.100.105)
 - Machine name (Enterprise)
 - Domain name
 - enterprise.federation.org
 - enterprise.geek.com

The MAC Address

Pretend for a moment that you're going to write a letter to a friend. You dash off your note, stuff it into an envelope, and stick a stamp on it. Now all that remains is for you to tell the post office whom the letter is for. You could just write "John Smith" on the front and hope the mailman knows who Smith is; unless you live in a very small town, however, that's not likely. In addition to your friend's name, you have to write a street address, like "1218 Parsell Street, Fredericksburg, Virginia, USA." You need to tell the post office which house to deliver the mail to, and it's a good idea to include any applicable postal or Zip code for the address. When the occupants in the house receive the mail and look at the name, they know to give the letter to John Smith.

Network addressing works in a similar fashion. When a computer wants to talk to another computer on the same LAN, not only must it put the destination's IP address in the packet but also the Link Layer hardware address. After all, the IP address is part of the Network Layer, and the physical Link Layer doesn't know anything about it. When the interface hardware needs to know whether a given packet is destined for its own machine, it can only look at the link-level headers. An IP address would be meaningless to the hardware.

This is why every piece of networking equipment is required to have a unique *Media Access Control* (MAC) address. The MAC address is a 48-bit number that uniquely identifies that particular piece of equipment. Each device has its own MAC address that is unique to the entire world. Even if two computers are the same model made by the same manufacturer, each will have its distinct MAC address. If a device has more than one network interface, such as a firewall, it will have more than one MAC address (one per interface).

MAC addresses usually are written as 6-byte hex strings with colons inserted between the bytes: for instance, 00:60:1D:F0:EA:AF. When an interface listens to the network for traffic, it looks inside each Link Layer header to see whether the destination MAC address matches its own. Only if the address matches does the hardware pass the packet up the IP stack for processing.

IP Address Classes

- 4-byte address
- The address is broken down into a network and host portion.
- Class A
 - First byte must begin with 0.
 - Class A: 1.0.0.0 through 127.255.255.255
 - N.H.H.H; 255.0.0.0; /8
- Class B
 - First byte must begin with 10.
 - Class B: 128.0.0.0 through 191.255.255.255
 - N.N.H.H; 255.255.0.0; /16
- Class C
 - First byte must begin with 110.
 - Class C: 192.0.0.0 through 223.255.255.255
 - N.N.N.H; 255.255.255.0; /24

IP Addresses and Subnets

The Internet relies on a number of assumptions to function properly. One of these assumptions is that each organization connected to the Internet will use unique IP addresses for its computers. These days, when you get connected to the Net, your ISP assigns you a block of IP addresses to use for hosts at your site. If you're a home user, you typically only get one address (or maybe two or three if you have that many computers). Companies and other organizations usually get substantially more. In this section, you learn how IP addresses are allocated to make the most efficient use of this limited resource.

IP Broadcast Addresses

- Special type of address
- Will be sent to all hosts on a given network segment
- Broadcast address (directed broadcast) is when the host portion is set to all Is.
 - For example, 10.255.255.255 is the broadcast address for the 10 network.
- Limited broadcast stays on local segment
 - 255.255.255.255

Broadcast Addresses

As mentioned previously, you should never use host addresses of all 0s or all 1s. In a /24 subnet, these would be the addresses x.x.x.0 and x.x.x.255. That's because these are reserved for a special kind of network transmission known as a *broadcast*. The two addresses are known as *broadcast addresses*. A broadcast packet is a single packet that is processed by every IP stack on the LAN.

As astute as you are, you are no doubt asking yourself, "Why are there two broadcast addresses?" It's just a quirk of history. Certain older IP implementations used broadcast addresses of all 0s. Most modern stacks use all 1s instead, although some will understand both for the sake of backward compatibility. Because it's sometimes tough to know which is which, assume computers use 1s unless otherwise indicated, but avoid using the 0s just in case. For the rest of this section, we assume the 1s convention.

Types of Broadcast Packets

There actually are two types of broadcast packets. The first is called a *net-directed broadcast*, which is a fancy way to say that the network number bits in the broadcast address are the same as those in the host's IP address. The host bits are still all 1s, however, to differentiate these packets from regular traffic. Net-directed broadcasts are, as the name implies, intended for all hosts with a specific network number. Routers and gateways usually pass these along to other parts of the same network that might happen to reside on different physical segments of cable.

The second type of broadcast is referred to as a *limited broadcast*. Packets of this type contain a destination address composed entirely of 1s, which is 255.255.255.255. This is referred to as limited because routers or gateways never pass on these sorts of broadcast packets. They are only intended for a single network segment. Limited broadcasts are mostly used when computers boot, so they can obtain DHCP leases or otherwise configure their network interfaces.

Note

Remember, there's also a broadcast MAC address involved here. The special FF:FF:FF:FF:FF:FF address causes all hardware interfaces to send the packet up through their IP stacks for processing.

Private Network Addressing

- Address space is scarce.
- Advisable to hide internal address structure
- Used to handle "private" address space
- Makes more efficient use of IP addresses
- Makes it difficult to trace information back to source
- Non-routable addresses (RFC 1918):
 - 10.X.X.X
 - 172.16.0.0-> 172.31.255.255
 - 192.168.X.X

Private Addressing

Not every host capable of accessing the Internet has a direct connection. These days, computers are (or should be!) behind firewalls of some type. They also may use *Network Address Translation* (NAT), so that the IP addresses in use on the internal LAN are automatically mapped to a different set of addresses when they traverse the firewall and go out to the Internet. If no one on the Internet can see these addresses, why should an organization bother to request an address block from its ISP? Even more to the point, why should the ISP waste addresses by allocating them to a customer when these addresses never will be routed over the Internet?

It turns out that the answer to each of these questions is, "They don't have to." The *Internet Assigned Numbers Authority* (IANA), the ultimate authority for IP address assignments, has designated three sets of *private address blocks* that never can be routed over the Internet and therefore are free for anyone to use as they want within their own networks. Because these addresses can't traverse the Internet, it doesn't matter whether 2, 5, or 10,000 different sites pick the same address to use on their internal networks. So long as the traffic is translated to routable IP addresses before it goes out on to the Internet, the actual internal network numbers used don't matter. The following figure shows the three private network blocks. Feel free to use these at will in any of your private networks.

Figure: Private Network Allocations

10.0.0.0/8
172.16.0.0/16-172.31.0.0/16
192.168.0.0/16

Types of NAT

Three general types of NAT that are used

- One-to-one NAT
 - Used on DMZs with public accessible systems
- Pool NAT
 - Set of public addresses that are mapped
- Many-to-one NAT
 - Formally referred to as PAT (port address translation)

Besides being a good neighbor and not using more than your share of addresses, using NAT means that your host systems are shielded from the Internet from a reconnaissance point of view (in addition to the filtering that your firewall provides). There are a number of variations of NAT. RFC 2623 defines the standards for NAT used on the Internet. Generally, we use NAT in the outbound direction, from your network to the Internet. We might also use NAPT, *Network Address and Port Translation*. This is best explained with a common example. Suppose your site has NAT and you also choose to use an outbound proxy for HTTP. You would need to give your web browser the internal IP address and port number for your proxy server. This is done in Internet Explorer by selecting Tools, Internet Options, Connections, LAN Settings and then selecting the appropriate proxy settings.

Name Resolution

- Host table
 - Static entries in a file
 - Used on small networks
- DNS
 - Resolving domain name to IP address
 - Hierarchical-based system
 - Used on large networks

At one point, the IP addresses and names were kept in tables and they were downloaded nightly. As the Internet kept growing, this became impractical for a number of reasons related to the size of the table and issues surrounding a single point of failure.

Naming a thing is not the same as knowing a thing, but it is often a first step. I remember when I first started hearing about the Domain Name Service (DNS). At this time, the major database vendors were all talking about their distributed database products that would be available "real soon now," and then the next thing I knew I was running distributed database software. It didn't cost me a thing and it worked pretty well from day one. DNS is a *distributed database* because the entire address table is not stored on a single host; instead, it is distributed across many servers.

Domain Name Service (DNS)

- Protocol for translating IP addresses to domain names (and back again)
- Hierarchical system of domain names
- Root-level servers for top-level domains (.com, .org, .edu, .gov, and so on)

As we have seen, computers on the Internet are identified by IP addresses, which are nothing more than very large numbers. If you have a very small network, say no more than a handful of machines, you may be able to remember the IP address of every machine on your network. But what about the networks of your friends? Can you remember all of them as well? When you interact with new people on the Internet, do you need to remember their addresses as well? How about the whole Internet? How do you keep all those addresses straight?

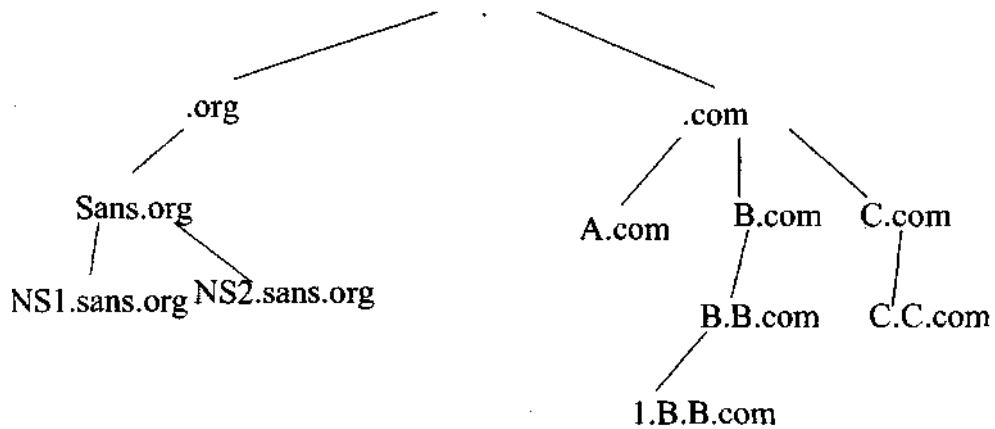
The sad fact is that IP addresses are hard to remember. Humans find it much easier to remember names rather than numbers. But IP addresses are, of course, expressed as numbers. We need a way to associate easy to remember names with hard-to-remember numbers.

Enter the *Domain Name System*, or DNS. DNS is a protocol for translating IP addresses to names and back again. This is how you can enter a name such as www.microsoft.com in your web browser and it can find the computer at IP address 127.46.131.137.

DNS is not a single server that tracks all the names and addresses on the Internet. Rather, it is comprised of several master servers and thousands of smaller servers around the globe, each handling a small part of the Internet. Here's how it works:

The various networks on the Internet are divided up into groups called *domains*. The domains are structured in a hierarchy like a tree. The top level of the tree is called the *root* or *top-level domain*. There are a handful of these such as .com, .edu, .gov, and .org. Each level down the hierarchy tree adds another level to the domain. Each level can be another domain (called a *subdomain*) or a host computer itself. The diagram on the next slide illustrates this point better.

DNS Hierarchy



Here you see a pictorial representation of a typical DNS hierarchy. The structure is called a "tree" structure because it looks a bit like an upside-down tree. If you hold the picture upside down you will see what I mean. The structure here is an example of a portion of the .com domain. In reality, the .com domain is much, much bigger, but this small example will be perfect for our purposes.

At the top of the structure is root servers. Each root server would then direct you down the tree.

DNS Queries

- **gethostbyname**: When you have the fully qualified domain name (pokey.sans.org) or the local name within your private network (pokey) and need the address
- **gethostbyaddr**: When you have the address and need the name

On your slide, you see two commands: `gethostbyaddr` and `gethostbyname`. If you have one piece of information, you can often acquire the other. If you have an NT or Unix system, try this. Think of a well-known host, such as www.sans.org. Then type:

```
nslookup <name of the well-known host>
```

It should return the IP address. Now try an IP address and you should get a host name. `Nslookup` then, is an application that does the `gethostbyaddr` and `gethostbyname` functions for you.

`Gethostbyname` is by far the more common lookup and is called a *forward lookup*. `Gethostbyaddr` would of course be a *reverse lookup*.

Domain Hijacking

- Most Internet-based services are based on DNS naming.
- Most assume DNS servers are correct.
- DNS has no built-in security.
- Domain hijacking allows an attacker to "take over" a domain.
- It can redirect communications from a "good" domain to "bad" domain.

Now that we know how DNS works, we can discuss the security aspects of DNS. DNS is one of the most important functions that make the Internet what it is today. This is not because of its importance to the technical infrastructure, because we have seen that it is not absolutely required for the Internet to work. However, most Internet services in use today rely on DNS to enable users and programs to easily locate and connect to any host on the Internet. Without DNS, Internet users would have to rely on using IP addresses to locate computers. How long do you think that would last?

DNS has no built-in security mechanisms. There is no authentication of either the user, the requesting computer, or the DNS server. And there is no verification that the machine name or IP address the DNS server gives as a reply to a query is, in fact, correct. When something becomes that important and has no built-in security checks or controls, it is ripe for attack by evildoers. DNS is no exception.

In recent years we have seen a phenomenon called *domain hijacking*. Simply put, an attacker takes over a domain by first blocking access to a domain's DNS server and then putting up his own server in its place. So, for example, if an attacker wanted to take over the xyz.com domain, he would have to remove the xyz.com DNS server from operation using a denial-of-service attack to block access to the DNS server. Then he would put up his own DNS server, advertising it to everyone on the Internet as xyz.com. So, when an unsuspecting user went to connect to xyz.com, he would get the attacker's domain instead of the real one.

Why would someone want to do this? Well, the attacker could put up a bogus web server as www.xyz.com and advertise that XYZ Corp was giving away millions of dollars worth of merchandise free, or that XYZ Corp was announcing massive layoffs in the coming months. Because users have no way of distinguishing false information from real information, this type of activity can have serious consequences for the victim of a DNS hijacking attack.

IPv6

- IPv4 accommodates 4.2 billion unique addresses (32-bit address).
- New technology growth requires more address space.
- IPv6 is designed to meet addressing growth.
 - 128 bits = 340 undecillion addresses (7 addresses for each atom of every human)
 - Offers greater flexibility in allocating addresses

IPv6

The IPv6 protocol was designed to supersede IPv4 addressing while supporting the growth of the Internet. While the IPv4 protocol accommodates 4.2 billion unique IP addresses with a 32-bit address, the allocation of IP addresses on the Internet was not completed in the most effective manner, leaving a shortage of available IP addresses. With technology such as NAT, the Internet continued its growth, but it was somewhat limited without the widespread availability of globally unique IP addresses. New technology such as mobile phones and PDAs connecting to the Internet has increased demand for addresses, as well as the spread of Internet technology to populous countries such as China and India. As a result, a new mechanism was needed to accommodate continued growth and adoption of Internet-connected technology.

The IPv6 protocol was designed to meet these growth demands, expanding the address size from 32-bits to 128-bits. A 128-bit address is approximately 340 undecillion addresses or 340,282,366,920,938,463,463,374,607,431,768,211,456. With this many unique addresses, the IPv6 protocol can accommodate seven unique IP addresses for each atom in every human on earth.

Of course, all of our atoms don't need that many IP addresses (two or three would suffice). Instead, the sheer volume of available IP addresses accommodates for more flexible deployment of address space on the Internet. For example, ISPs will be able to geographically assign IPv6 prefixes to different parts of the world, allowing for the simplified routing of traffic on the Internet. Organizations can obtain an IPv6 prefix with sufficient available addressing to accommodate all present and future addressing needs.

IPv6 Features

- Extended address space
 - Route aggregation, improved delegation/management, hierarchy
- Autoconfiguration support
- Support for IPv6 over IPv4 (tunneling)—
- Support for IPv4 over IPv6 (translation)-
- Flexible embedded protocol support

IPv6 Features

The "killer feature" of IPv6 is the expansion of address space, permitting route aggregation on core Internet routers through geographic address space allocation, improve delegation and management of addresses to organizations and ISPs alike, as well as a hierarchical distribution of address space that makes troubleshooting and Internet routing simpler.

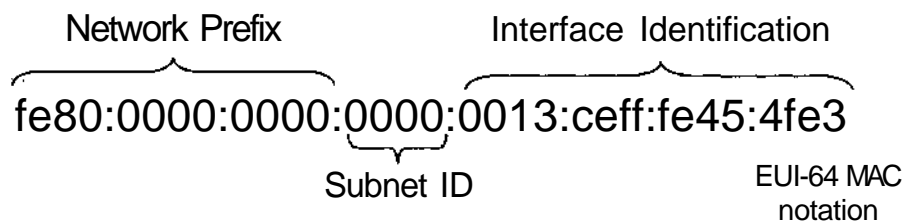
Another valuable feature of IPv6 is support for addressing autoconfiguration. Anyone who has been responsible for manually assigning IP addresses to hosts understands that this is a problematic and cumbersome process. With 128 bits of address space, it becomes possible to use the globally unique MAC addresses on all network cards as IP addresses. In this way, administrators can simply introduce a new node to an IPv6 network without manually specifying an IP address; the IP address is configured automatically based on the local MAC address and advertisement information from the default gateway on the network.

During the transition process between IPv4 and IPv6, it is possible to establish IPv6 tunnels over the existing IPv4 Internet using IPv4 protocol 41 or one of several tunneling protocols, such as AYIYA (Anything in Anything) or Teredo (Tunneling IPv6 over UDP through NAT). Further, it is also possible to continue supporting IPv4 traffic on an IPv6 backbone using gateway services that translate IPv4 packets into an IPv4 format.

Another significant change in the IPv6 protocol is the use of a fixed IP header. While the IPv4 header could expand to include additional information such as strict or loose source routing, the IPv6 protocol has a fixed header length of 40 bytes. In order to accommodate additional flexibility in the protocol, IPv6 introduces a "next header" field that indicates the embedded protocol contained in the packet payload. This is similar to IPv4's embedded protocol field, but unlike this field, the next protocol can include multiple embedded protocol fields, one right after another. Currently supported IPv6 next header protocols includes the encapsulating security protocol (ESP) and authentication header protocol (AH) for IPsec, destination options header to specify processing options at the destination system, and upper-layer protocols such as UDP, TCP, and ICMP.

IPv6 Addressing

- Addresses specified in hex are colon-delimited.
- Autoconfiguration uses local MAC address with router prefix/subnet ID.
- Groups of repeating 0000's are simplified with "::".



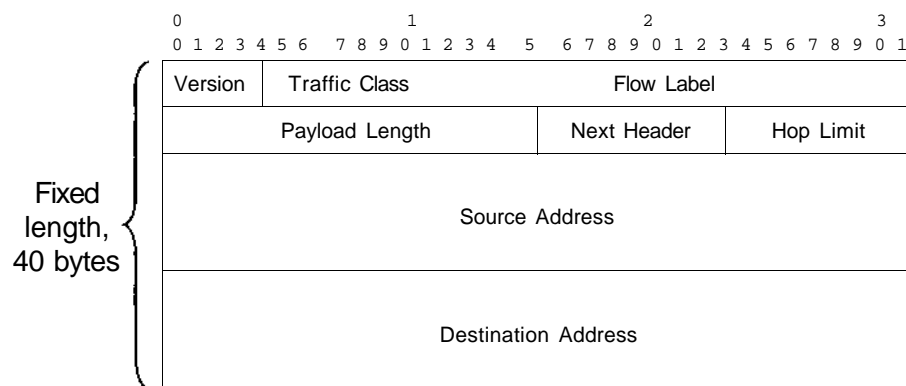
IPv6 Addressing

Because of the longer address space, changes have been made to how IP addresses are represented from IPv4. Where it is simple to specify a 32-bit address in dotted-decimal notation, remembering an address that is four times longer in the same format can become overwhelming quickly. In IPv6, IP addresses are represented using hexadecimal notation, with values separated by colons instead of dots. For the foreseeable future, many IPv6 addresses will include strings of repeating four repeating 0s ("0000"), which can be condensed to just a single colon to represent one or more groups of four 0s.

IPv6 addresses are broken up into three major sections: the network prefix, the subnet ID, and the interface identification section:

- **Network Prefix** — The network prefix is represented in the first 48-bits (6 bytes) of the IPv6 address. This is the address portion that is allocated to organizations that need to address IPv6 clients, or to preserve other network functionality. Some fixed network prefix allocations include "fe80::" for local network use, "ff00::" for multicast traffic, "2001::" for large ISP inter-domain routing, and "2002::" for IPv6-to-IPv4 gateway networks.
- **Subnet ID** — The subnet ID is configured according to the addressing needs of the organization. For flat IPv6 networks, this value will usually be "0000," but can be any value selected by the organization that has been allowed the network prefix.
- **Interface Identification** — The interface identification section uniquely identifies the IPv6 node. With IPv6 autoconfiguration, the MAC address of the client populates the interface identification portion of the IPv6 address. Because a MAC address is a 48-bit value, but the interface identification portion of the IPv6 address is 64 bits, the MAC address is expanded to fill the space by converting it to the Extended Unique Identifier (EUI) format specified by the IEEE. The EUI expansion takes the first three octets of the MAC address, appends the constant value "ff:fe", and then appends the last three bytes of the MAC address to form the interface identification portion of the IPv6 address.

IPv6 Header



Traffic Class + Flow Label provide QoS, Next Header indicates embedded protocol data, and Hop Limit prevents routing loops

IPv6 Header

To accommodate the changes in the IPv6 protocol, the header information has changed by removing superseded functionality from the IPv4 header and introducing some new fields:

Version: 4-bits, the version field indicates the packet is IPv6 and is always a "6".

Traffic Class: 1 byte/8 bits, the traffic class field is used to specify the priority of the packet for QoS.

Flow Label: 20 bits, the flow label field is used for QoS management to convey special handling functions for the packet.

Payload Length: 2 bytes/16 bits, the payload length fields specifies the length of the packet in a quantity of bytes.

Next Header: 1 byte/8 bits, the next header field specifies the next encapsulated protocol in the payload of the packet. The values that are assigned to IPv4 embedded protocols (such as TCP, UDP, and ICMP) are forward-compatible with the IPv6 next header field.

Hop Limit: 1 byte/8 bits, the hop limit field is used to prevent routing loops by decrementing the hop limit value at each router. This is similar to the TTL field used in the IPv4 header.

Source Address: 16 bytes/128 bits, the source address of the IPv6 station transmitting the packet.

Destination Address: 16 bytes/128 bits, the destination or recipient of the IPv6 packet.

UDP (User Datagram Protocol)

- Connectionless communications
- Sends packets out with no guaranteed delivery
- Much less "overhead"
- Good if small amount of packet loss is acceptable

The User Datagram Protocol (UDP)

UDP is the simpler of the two Transport Layer protocols typically used with IP, which is why we cover it first. A trick to remember it is to think of the *Unreliable Damn Protocol*. However, do not get tricked because it is really not unreliable; it is just not guaranteed delivery. In fact, UDP is a useful, important protocol and used by many applications today.

UDP Ports

- Same port concept as TCP (trusted port and ephemeral ports)
- Some common applications that use UDP:
 - DNS (53)
 - NTP(123)
 - BOOTP(67and68)

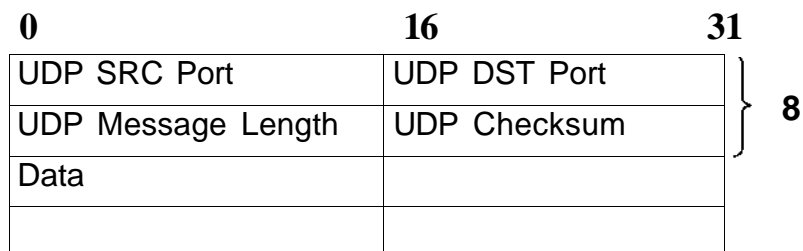
Typical Uses for UDP

UDP typically is used in situations where it's okay if some packets are lost or reordered. In a streaming audio application, for example, each packet contains such a miniscule amount of audio data that the client probably can afford to lose one or two, or even several packets in succession without suffering a noticeable lack of quality. By doing without some level of error checking, the application can push the audio data around the network much more quickly, which gives better quality overall, even if a few packets don't make it through.

Also UDP is often used for applications that don't send much data, perhaps just a handful of bytes, so they don't mind retransmitting the data if it happens to get lost. In most cases, the packets will go through fine, but the loss of one, two, or even several packets poses no great problem. The time it takes to recover from the occasional dropped packet is more than made up for by the time saved by not checking for errors that rarely happen anyway. It's easy to retransmit a query if the client doesn't get a response in a reasonable amount of time.

Other important UDP-based protocols include the *Network Time Protocol* (NTP) and the BOOTP/DHCP protocols used by hosts to automatically configure their network interfaces and load their operating systems via the network when they start up.

UDP Header



The UDP Header

Even a featherweight protocol such as UDP needs *some* kind of packet header because the Transport Layers on each host need a way to communicate essential information. This slide diagrams the layout of the UDP header. This looks like a short header, and it is; remember, however, that these are Transport Layer headers. The Network Layer just below this will also add its own headers, encapsulating the UDP headers.

As packet headers go, UDP is simple. There are only four fields: source port, destination port, datagram length, and checksum. Each field is exactly 2 bytes long. A mere 8 bytes of overhead per packet is pretty good! Let's examine these fields in detail.

Source Port and Destination Port

UDP uses the concept of *ports* to help get datagrams to and from the proper applications. You have learned that ports are just ID numbers associated with certain applications running on a host. When one host wants to send datagrams to a server process running on another host, it needs to know what port that process is listening to. If a computer is like an apartment building, the applications running on it are like its residents, and the port numbers are like the apartment numbers in which the residents live. W. E. B. Smith lives in apartment 80, for example, so messages (packets) going to that apartment number clearly are meant for him.

Most server ports are *well known*, like web servers that always listen to port 80 no matter how many times they are restarted or the machine is rebooted. Clients usually use *ephemeral* ports—that is, ports that change each time the client application runs. Why the difference? Well, clients usually poll servers, and not the other way around. Because the client almost always initiates the communication, it needs to know in advance what port the server runs on. After the client contacts the server, however, the server can look in the packet headers to see what port that particular client is using, so having a predictable port on the client side isn't important.

As with all the IP protocols, the *source port* indicates the port the sender is bound to, whereas the *destination port* indicates the service on the receiver to which the packet should be delivered. Valid port numbers are 1 through 65,535.

TCP (Transmission Control Protocol)

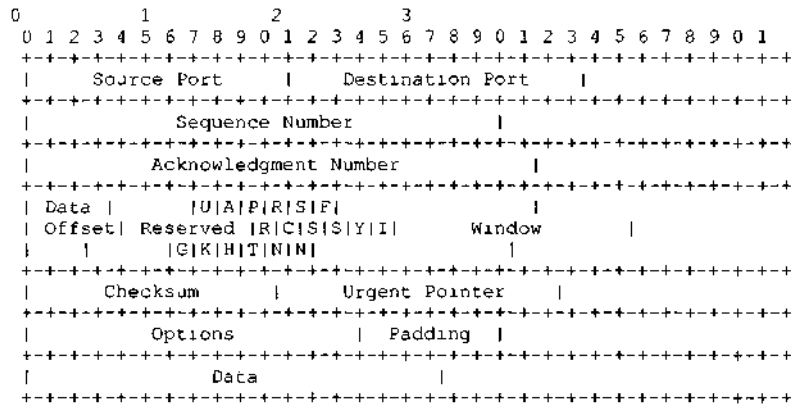
- Connection-oriented communications
- Ensures reliable packet delivery
- "Expensive" overhead
- Three-way handshake:
 - **SYN**
 - **SYN ACK**
 - **ACK**

TCP

TCP is probably the most commonly used Transport Layer protocol today. It establishes a virtual connection, often referred to as a *session*, between the hosts. The protocol is designed to provide reliable connections over possibly unreliable networks. Unlike UDP, which blindly sends datagrams and hopes they arrive, TCP can guarantee that the packet will arrive or at least that it will notify you of a problem. Because of this guarantee, TCP often is a network programmer's protocol of choice. It's probably the easier of the two protocols to program for, too, because most of the error handling is down inside the Transport Layer and out of sight from the application code. TCP is especially useful for any application in which there are more than one or two network hops between two computers, because more hops equals more chances for errors to be introduced into the communication.

Most of the Internet protocols you use every day are based on TCP. Some examples include HTTP (Hypertext Transfer Protocol, used by web servers and browsers), FTP (File Transfer Protocol, used to transfer files to and from servers) or POP3 (Post Office Protocol version 3, used to download e-mail).

The TCP Header



The TCP Header

Because TCP is a much more heavyweight protocol than UDP, it requires a much larger header. The normal TCP header is a whopping 20 bytes, more if any options are specified. From a security standpoint, some of these fields are more important than others. Let's take a look at some of the key elements of the TCP header.

Key Fields of a TCP Header

- Source port
- Destination port
- Sequence number
- Acknowledgement number
- SYN bit
- ACK bit

Now let's look at some of the key fields in the TCP header in detail. When you connect to a system, you not only connect to an IP address, you also connect to a specific port on a computer. Think of this as going to an apartment building. The address will only get you to the apartment building, not inside. To get inside, you have to not only know the address, you also have to know a specific apartment number. The same logic holds when connecting to a computer; you have to connect to an IP address and a port number. The destination port is the port or application you are connecting to on a remote computer. The source port is the port that you are connecting from.

Because TCP is reliable, it uses sequence numbers to track packets and provide reliable delivery of information. Sequence numbers are used by the host computer when sending out data, and the acknowledgment numbers are used to acknowledge the receipt of information.

The SYN, or synchronization, bit is used when establishing a connection and is used only in the first two legs of the three way handshake. The ACK, or acknowledgement, bit is used when a system is acknowledging the receipt of information.

TCP Ports

- Well known ports, < 1024
 - 20 - FTP data
 - 21 - FTP
 - 23 - Telnet
 - 25 - SMTP
 - 53 - DNS
 - 79 - finger
 - 80 - HTTP
- Source ports, >= 1024

TCP utilizes ports numbered 1 through 65,535 to communicate. Ports 1 through 1023 are considered trusted or well known ports. These ports are each assigned and reserved for a specific function. Ports 1024 and above are called the ephemeral ports, which means they could be used by any service for any reason.

Some of the more common port assignments are as follows:

20 - FTP data
21 - FTP
23 - Telnet
25 - SMTP
53 - Domain Name System (DNS)
79 - finger
80 - HTTP

It is important to point out that you can run a service on any port you want, but if you use your own port assignments, no one will be able to communicate with you. For example, mail servers by default try to connect to each other on port 25. You can run your mail program on port 200, but when any mail system tries to connect to port 25 and it is not open, it will not be able to send you mail because it doesn't know that you are running mail on a different port.

When a host connects to a server, it usually connects on a well known port number. But what does it use as the source port? What it usually uses is an ephemeral port or a port greater than 1023. It is chosen so that the port is randomly generated. If you connect to a server several times in a row, the way TCP works is that the source port must be different each time.

TCP Code Bits

- Also called TCP flags
- Control data flow and signal information to receiving host

```
      2   1 | | 8   4   2   1
+--+--+--+--+ +--+--+--+--+--+--+--+--+--+
| U | A | | P | R | S | F |
+--+--+--+--+ +--+--+--+--+--+--+--+--+--+ .
```

Upper bits are used for ECN (explicit congestion notification).

Flags

TCP stacks sometimes need to communicate about the data they exchange. The catch is they can't insert their own information into the payload because that would corrupt the data stream and might confuse the applications. Instead, the TCP protocol provides six 1-bit flags that can be specified in the packet headers. Some of these are more common than others, but the unusual use of TCP flags is a good indicator of suspicious traffic, so you should become familiar with all of them.

- **URG (Urgent).** The Urgent flag is used by applications such as Telnet and Rlogin. An application can set this bit to let the other end of the connection know that important data is coming, but it's up to the client and server to decide what's urgent and what to do about it. There are some ambiguities inherent in this implementation, too, such as the fact that there's no way to tell the receiver where the urgent data starts in the stream. It could begin at any byte in that packet's payload. There's also no way to specify where urgent data ends. That's why most legitimate applications never use the URG flag.
- **ACK (Acknowledgement):** The Acknowledgement flag is used to indicate that the sender is acknowledging receipt of some data. The receiver should look in the Acknowledgment Number field to see what data is being ACKed, as discussed previously.
- **PSH (Push):** TCP stacks usually buffer incoming data until a certain amount has been collected, and then pass it in a chunk to the application. When data is being transmitted in bulk, this usually is the most efficient way to handle the stream; for interactive processes (such as Telnet or SSH), however, it's more important that data be processed as soon as it comes in, even byte by byte. To ask for this behavior, the sender can set the PSH flag on a packet to indicate that it shouldn't be buffered but instead should be passed immediately to the remote application for processing.
- **RST (Reset):** Immediately upon receipt of a packet with the Reset flag set, a host should terminate the connection that contained that packet.
- **SYN (Synchronize):** The Synchronize flag indicates a connection request.
- **FIN (Finish):** The FIN flag is just the opposite of SYN. It indicates that a connection is being shut down in an orderly fashion. It contrasts with RST, in that FIN is a much more graceful way to close a connection.

TCP and UDP

- TCP
 - Reliable
 - Connection-oriented
 - More protocol overhead (slower)
 - Optimized for sessions
- UDP
 - Unreliable
 - Connectionless
 - Less protocol overhead (faster)
 - Optimized for query responses

We hope by now you have a much better idea of how much work TCP puts into making sure your connections are as efficient and as error-free as possible. Having all that built-in, error-correction capability means a tradeoff between raw speed and reliable communications; for most applications, however, especially those designed for the Internet, it's probably well worth it. It's easier to program for, too, because TCP takes care of a lot of the grungy details for you.

ICMP (Internet Control Message Protocol)

Two purposes:

- To report errors (troubleshoot) rather than transfer information.
- To provide network information. Ping and traceroute are the best known ICMP applications.

The Internet Control Message Protocol is a fascinating lightweight set of applications that were originally created for network troubleshooting. The packets are close to the IP layer, only with no notion of a transport.

The most well known ICMP application is certainly the echo request/echo reply, or ping. Ping is used to find whether a given Internet host is reachable or not. Traceroute is built on ping and used to plot out the path a packet took through the network.

Other ICMP applications are used for flow control, to reroute packets, and to collect network information.

ICMP packets can serve multiple functions (such as echo request/reply and other error messages). To identify the purpose of each ICMP packet, each ICMP packet has a Code and Type field. Each type of ICMP packet serves a different purpose. Some are essential to Internet communication and some can be used for malicious purposes. For a detailed listing of ICMP type and code information, refer to <http://www.iana.org/assignments/icmp-parameters>.

Ping

- Ping is used to see whether a host is active.
- Ping sends ICMP echo request and waits for ICMP echo reply.
- With security concerns, some sites are blocking ping, so this does not always work. You might have to use TCP scans.

Ping checks to see whether a host is active on the network. It does this by sending out an ICMP echo request. If the system is alive, it sends back an ICMP echo reply. When the sending system receives the echo reply, it knows the remote host is on the network.

With security concerns, some sites or firewalls block ICMP or ping traffic, so just because you do not receive a reply does not mean that the system is not on the network. It might mean that the traffic is being blocked. Most operating systems come with versions of ping built in, or you can get third-party ping software.

Traceroute

- It shows you the path a packet took to get to its destination.
- It can tell you the external router for a site and therefore be used to map a network.
- Normal traceroute lists the routers.
- As a general rule, all hosts on the same network have to go through the same external router and potentially the same firewall.
- By performing traceroutes and looking at the last couple of hops, you can spot similarities.

Traceroute

From an administrative or security point of view, probably the second most useful application of ICMP is the traceroute (or tracert.exe) command. We talked about traceroute briefly when we discussed the Time-To-Live (TTL) value in the IP header. Traceroute uses a clever combination of TTL values and ICMP replies to map out the route packets take from one computer to another, sometimes through many hops. The command works by sending a series of packets all going to the same destination but with TTL values starting at 1. When the first packet is sent, its TTL expires at the first hop, so the router usually replies with an ICMP "Destination Unreachable" or "Time Exceeded" message. The traceroute command eventually receives this reply and looks inside its payload for the IP address of the sender, which it assumes is the first hop's router.

Traceroute then sends a second packet, this time with a TTL of 2, which expires at the second hop, generating another ICMP reply. Traceroute now knows the second hop's router as well. It keeps sending packets this way, incrementing the TTL by 1 each time and getting replies from each hop until one of the packets finally is delivered to the destination host. By continually incrementing the TTL, traceroute can record all the routers in the path the packets take between your machine and some other machine on the Internet.

Protocols: A Review

- **TCP:** Although slower, TCP is reliable and is the basis for most Internet applications (for instance, FTP and Telnet).
- **UDP:** UDP is faster and less reliable and is often the basis for query-type applications (for instance, NFS, RPCs, and DNS).
- **ICMP:** Helps troubleshoot errors.

There are, as mentioned previously, other protocols than ICMP, UDP, and TCP, including IPSec and routing protocols. There are still other protocols that are primarily known by the numeric ID they use in the Protocol ID field. It can be instructive to run a network analyzer or a software sniffer such as TCPdump and filter out the main protocols, and then examine what is left in the output from the captured data. There are versions of TCPdump for Unix and for Windows (Windump). If you are running TCPdump, you could use a simple filter such as "not TCP and not UDP and not ICMP and not IGMP." This often uncovers odd traffic that can be quite interesting to examine.

TCPdump is a freeware protocol analyzer for Unix/Linux. It can be downloaded from <http://www.tcpdump.org>.

Windump is a version of TCPdump that has been ported to Windows (9x, NT, and 2000). It can be downloaded from <http://netgroup-serv.polito.it/windump/>.

Application Layer Security Protocols

- S/MIME
 - The secure MIME
- Secure electronic transaction (SET)
 - Originated by Visa and Mastercard as an Internet credit-card protocol
 - Supports the authentication of sender and receiver
 - Uses des symmetric key encryption
 - Uses RSA public key for symmetric key exchange and digital signatures
- Secure hypertext transfer protocol (S-HTTP)
 - Early standard for encrypting http documents
 - Used for Web transactions
 - Supports a variety of encryption algorithms
 - Provides authentication, confidentiality, integrity, and nonrepudiation
- SSH (secure shell)
 - Supports authentication, compression, confidentiality, and integrity
 1. Rsa certificate exchange for authentication
 2. Triple des for session encryption

Security Protocols

All these different protocol layers are great, but you may be wondering about their roles in securing your information. As you'll see later in this section, none of them really provides much security. Some have basic integrity checking to make sure data isn't accidentally modified by faulty network equipment, but IP lacks good support for confidentiality and integrity.

All is not lost, however. Many solutions to this problem have cropped up over the years, and to provide these resolutions there are a plethora of protocols you can utilize. Typically they fit in either the Application or the Network Layer of an IP stack. Let's look at just a few of them.

Application Layer Security Protocols

Application Layer protocols are the easiest to understand. In fact, you probably already use some of them. Security protocols in the Application Layer rely on a program's developers to explicitly code support for the protocol into their product. Probably the most common example of an Application Layer protocol is the *Secure Sockets Layer* (SSL). SSL started life as a way to enable secure communication between web browsers and servers, but today you can find it embedded in a wide variety of applications. Its flexibility and security make it a good fit for a wide variety of communication security needs.

Two other examples of common application layer security protocols are the *Secure Multipurpose Internet Mail Extensions* (S/MIME) and *Privacy Enhanced E-mail* (PEM) standards for secure e-mail. Both S/MIME and PEM easily allow users to exchange encrypted and/or digitally signed messages, even if they use different e-mail programs. Both protocols format messages in such a way as to pass harmlessly through standard e-mail servers, so support for this protocol need only be present on the users' desktops. This flexibility makes the protocols compatible with virtually any mail server an organization might choose to use. Of the two, PEM has fallen somewhat out of favor, whereas S/MIME's popularity continues to rise.

Web Security

- SSL
 - Secure sockets layer (SSL) protocol
 - Developed in 1994 by Netscape
 - Protects the confidentiality and integrity of information transmitted between two applications
 - Provides for authentication in both directions
 - Uses public and private key encryption and a message authentication code (MAC)
 - Independent of application

SSL

- Secure sockets layer (SSL) protocol
- Developed in 1994 by Netscape
- Protects the confidentiality and integrity of information transmitted between two applications
- Provides for authentication in both directions
- Uses public and private key encryption and a message authentication code (MAC)
- Independent of application

Web Security (2)

- TLS
 - Transport layer security
 - Upgraded version of SSL
 - Protects the confidentiality and integrity of information transmitted between two applications above the transport level
 - Provides for authentication in both directions
 - Uses public and private key encryption and a message authentication code (MAC)
 - Independent of application

TLS

- Transport layer security
- Upgraded version of SSL
- Protects the confidentiality and integrity of information transmitted between two applications above the transport level
- Provides for authentication in both directions
- Uses public and private key encryption and a message authentication code (MAC)
- Independent of application

Web Security (3)

- SSL 3.0
 - Secure sockets layer
 - Uses cryptography to protect confidentiality
 - Extensible by allowing a variety of encryption algorithms to be used
 - Provides for authentication in both directions using digital signal standard and RSA
 - Uses public and private key encryption and a keyed message authentication code (MAC)

SSL 3.0

- Secure sockets layer
- Uses cryptography to protect confidentiality
- Extensible by allowing a variety of encryption algorithms to be used
- Provides for authentication in both directions using digital signal standard and RSA
- Uses public and private key encryption and a keyed message authentication code (MAC)

Other TCP/IP Protocols

- Telnet
 - Terminal emulation
 - Gives a user on a remote client machine access to the resources of another computer
 - Limited to running applications
- File transfer protocol (FTP)
 - Provides for file transfer between two computers
 - Supports access to directories and files
 - Cannot execute remote files as programs
- Trivial file transfer protocol (TFTP)
 - Reduced version of ftp
 - Sends and receives files
 - Has no directory browsing abilities
- Simple mail transfer protocol (SMTP)
 - Used to send and receive Internet e-mail

Telnet does not support downloading of files.

FTP supports authentication.

TFTP does not support authentication.

When a message is sent, it's sent to a mail queue.

The SMTP server regularly checks the mail queue for messages and delivers them.

Other TCP/IP Protocols (2)

- Simple network management protocol (SNMP)
 - Provides for the exchange of management information among network devices
 - Polls network devices from a management station

SNMP can notify network managers of any network events by employing agents that send an alert called a trap to the management station.

The database of these traps are called MIBS (management information bases).

Network Attacks and Abuses

- Logon abuse
 - Unauthorized access to sensitive network services by authorized personnel
 - Personal use or illegal or inappropriate content of material
 - Unauthorized non-business use
- Eavesdropping
 - Tapping network communications
 - Tapping into cable
 - Using an induction loop to pick up electromagnetic emanations from wires

Logon abuse

- Unauthorized access to sensitive network services by authorized personnel
- Personal use or illegal or inappropriate content of material
- Unauthorized non-business use

Eavesdropping

- Tapping network communications
- Tapping into cable
- Using an induction loop to pick up electromagnetic emanations from wires

Network Attacks and Abuses (2)

- Smurf attack
 - Uses IP spoofing
 - Uses Internet Control Message Protocol (ICMP) to "ping" packets
 - Attack site sends ping messages to bounce sites.
 - Attack site spoofs its source address, making it the address of the target site.
 - Bounce sites return ping messages to target site, saturating it.

Smurf attack

- Uses IP spoofing.
- Uses Internet Control Message Protocol (ICMP) "p^mg" packets.
- Attack site sends ping messages to bounce sites.
- Attack site spoofs its source address — makes it the address of the target site.
- Bounce sites return ping messages to target site, saturating it.

Network Attacks and Abuses (3)

- Denial of service (DOS)
 - Service outages are caused by saturation of networked resources.
 - Overloading or saturating network resources.
 - Network and computers cannot provide required services to users.

Denial of service (DOS)

- Service outages caused by saturation of networked resources.
- Overloading or saturating network resources.
- Network and computers cannot provide required services to users.

Network Attacks and Abuses (4)

- Network intrusion
 - Unauthorized intrusion into network, usually from an unknown, external attacker
 - Penetration attack
- Spoofing
 - Providing incorrect and false information to gain unauthorized access to network resources

Network intrusion

- Unauthorized intrusion into network, usually from an unknown, external attacker
- Penetration attack

Spoofing

- Providing incorrect and false information to gain unauthorized access to network resources

Network Attacks and Abuses (5)

- Piggy backing
 - Gaining unauthorized access to a system by using an authorized user's access
- Back-door attack
 - Attack through asynchronous connections, usually dial-up
- Probing
 - Type of eavesdropping in which the potential attacker acquires information of various services available on a network
 - Scan of hosts to determine active systems and open ports

Piggy-backing

- Gaining unauthorized access to a system by using an authorized user's access

Back-door attack

- Attack through asynchronous connections, usually dial-up

Probing

- Type of eavesdropping in which potential attacker acquires information of various services available on a network
- Scan of hosts to determine active systems and open ports

Network Attacks and Abuses (6)

- **SYN attack**
 - Attacker sends connection requests in a TCP session to a target system (beginning of initialization handshake).
 - Attacker does not respond when target system replies to connection requests.
 - Target system waits for responses but does not receive them. With many open-connection requests, target systems time out and crash.
- **IP spoofing attacks**
 - Attacker sends a packet with an IP source address of a known, trusted host.
 - Target accepts packet and acts accordingly.

SYN attack

- Attacker sends connection requests in a TCP session to a target system (beginning of initialization handshake).
- Attacker does not respond when target system replies to connection requests.
- Target system waits for responses but does not receive them. With many open connection requests, target system times out and crashes.

IP spoofing attacks

- Attacker sends a packet with an IP source address of a known, trusted host.
- Target accepts packet and acts accordingly.

Routing

This space intentionally left blank.

Two Addresses

At minimum, a computer has two addresses:

- MAC address
 - 48-bit address (12 hexadecimal digits)
 - First half vendor code (00-00-0c Cisco or 08-00-20 Sun)
 - Usually hard coded in to NIC
 - Does not change
- IP address
 - 32-bit address
 - Part network and part host
 - Configured by user
 - Changes based on location

To understand how routing works, you have to understand that any computer connected to a network has a minimum of two addresses. Usually there are two addresses per network interface. So, if a server has four network interface cards, or NICs, each interface would have two addresses: a MAC address and an IP address. The reason you need two addresses goes back to the OSI model and how communication is broken down into multiple layers. Layer 3 is responsible for routing traffic across a network, and IP operates at Layer 3 and needs an address in order to route the traffic. So there is an IP address that Layer 3 uses to determine how to get a packet from source to destination. As we go down the OSI stack, however, the Layer 3 information gets encapsulated by Layer 2 before it goes out on the wire. So Layers 1 and 2 need some way to directly send information to a given host. This is done via a *Media Access Control* (MAC) address that operates at the lower layers. Now let's look at each address in more detail:

- **MAC addresses:** A MAC address is a 48-bit address that is usually written as 12 hexadecimal digits grouped in pairs of two. So a typical address might look like the following: 00-00-0c-34-15-43. Because a MAC address is usually hard coded into the NIC card and does not change, it is the vendor's responsibility to make sure that every card has a unique MAC address. The way this is done is the MAC address is broken into two pieces. The first half, or 6 hexadecimal digits, is assigned to a specific vendor, and the second half is a unique number assigned by that vendor. Now as long as the vendor uses the first half of their code, it is their responsibility to make sure every card has a unique MAC address and that there are no duplicates. So by looking at a MAC address, you can tell what vendor the NIC came from. For example, if the first half is 00-00-0c, you know the card was produced by Cisco; if it starts with 08-00-20, you know it was produced by Sun.

- IP addresses: An IP address is a 32-bit address, or 4 bytes, and usually written with a period between each byte. So a typical IP address might be 15.5.10.35. An IP address is broken into two pieces: a network piece and a host piece, depending on the type of address it is (Class A, B, or C) and whether subnet masks are being used. You cannot tell where the division is just by looking at the address. You must also look at the subnet mask to see which piece identifies the network and which piece identifies the host. The IP address is configured by the user, and as the computer moves around or changes location, the IP address must also change.

Just to summarize the two addresses, let's look at an example. I travel around the world and check my e-mail from various locations. Each time I go to a new state or country, I have to reconfigure my machine with a new IP address, but my MAC address never changes. Actually, for my home network I know my IP address by heart because I change it so often, but I have no idea what my MAC address is because it never changes and it operates at a layer in the protocol stack that most people do not get that involved with.

MAC and IP Addresses

- No direct relationship exists between the two addresses.
- Given one address, a computer must send out a packet to find out the other address:
 - ARP (Address Resolution Protocol)
 - RARP (Reverse Address Resolution Protocol)

Okay, so now we know that there are two addresses—a MAC and an IP address—but how do we tie the two together? Unfortunately, no direct relationship exists between the two addresses. Looking only at an IP address, there is no way that you can determine what the MAC address is and vice versa. If I give you a MAC address of 00-00-0c-45-56-32, there is no way that you can tell me what the IP address is. You could make a totally random guess, but that would not be a good way to link the two together. Therefore, given one of the addresses, the only way to find out the address is to send out a packet saying, "Hey I know one address! Can you let me know what the other address is?" Actually, there is a protocol that will take care of this for us.

Address Resolution Protocol (ARP), given an IP address, will find out what the corresponding MAC address is.

Reverse Address Resolution Protocol (RARP), given a MAC address, will find out what the corresponding IP address is.

Address Resolution Protocol (ARP)



42.1 broadcasts a packet with 42.2's IP address and asks it to respond with its physical address.

0		16	31
HARDWARE TYPE		PROTOCOL TYPE	
HLEN	PLEN	OPERATION	
SOURCE MAC		SOURCE MAC	
SOURCE MAC		SOURCE IP	
SOURCE IP		TARGET MAC	
TARGET MAC		TARGET MAC	
TARGET IP		TARGET IP	

Now let's take a look at ARP and how it works. The Internet protocols are specified by standards documents called *Requests For Comments* (RFCs). ARP is specified by RFC 826. It is not an Internet protocol per se, because it is not carried in an Internet packet (or an IP packet). It is an Ethernet frame that is sent to all systems on a network segment. (This is what we call a *Layer 2 broadcast*.) If a message is a broadcast message, that means it is sent to all the machines on part or all of the network.

The source host sends the ARP request and includes its source MAC and IP address, and then presumably the destination host will pick it up and reply. Of course, the reply will contain the destination host's MAC and IP address. After this is done, the two systems can talk IP to one another. If you see an ARP, you are probably on the same physical cable segment as the sending computer, because ARPs will not be passed through a router.

Routing Protocols

- Distance vector
 - **RIP**
- Link state
 - **OSPF**
- Hybrid
 - **EIGRP**

We have seen how routing works and how packets get from source to destination, but how do routers actually determine the best path a packet should take through the network? The way routers do this is by communicating information with each other, giving each router information about possible paths through a network. As with everything, with computers you want things done in a uniform fashion, so protocols are developed. Routing protocols are the rules that routers use to communicate information with each other. There are two general types of routing protocols: distance vector and link state.

Distance Vectors

- Identifies neighbors and figures out distance metrics to each network
- Problems
 - Routing loops
- Solutions
 - Defining a maximum
 - Split horizon
 - Poison reverse
 - Hold-down timers

Distance-vector protocols work by each router identifying all of its neighbors or routers to which it has a direct connection. Any router that it is directly connected to has a distance of 0. Then by using the information it receives from its neighbors, it builds a routing table based on metrics to determine how many hops it would take to get to a destination network. They iterate on the number of hops to find the shortest-path spanning tree. To get the information they need, routers typically share the entire routing table with each of its neighbors. These algorithms tend to be simpler than link-state algorithms, but by sending the entire table can not only generate additional bandwidth but can be slow to converge, which means it leaves the routing table open to having routing loops develop.

With distance-vector routing protocols, slow convergence on a new configuration can cause inconsistent entries to exist, which cause a routing loop to be created. An example of a routing loop is this:

1. Router A sends all of its traffic to router B.
2. Router B sends all of its traffic to router C.
3. Router C sends all of its traffic to router A.

Now all the traffic is caught in an endless loop. This could be caused by the convergence problem. Suppose that router A has a direct connection to a network, and router B has an indirect connection through many hops to the same network. At this point, both router C and B will send their traffic through router A. Let's say the link router A has gone down. Well, router A knows that there is a slower connection through router B, so it sends its traffic to router B. Router C is slow in processing the information, so it still thinks that it can get to the network via router A. So it tells router B that it can get to the network in a small number of hops. Router B, knowing that the link for router A is down, thinks this is a better link and sends its traffic to router C; little does it know that router C is still sending it to router A. So now A sends its traffic to B, B sends it to C, and C sends it back to A. See how quickly a routing loop can be created?

There are many different ways that routing loops can be avoided, and we will briefly go over them now. Defining a maximum hop count will limit the extent of the routing loop. Split horizon also works very well, and what it says is that you should never send information about a route back in the direction from which the information originally came. Poison reverse is a variation of split horizon, whereby router entries are not modified so that they stay consistent with other routers until all routers have had a chance to make the update. Hold-down timers are used with poison reverse and tell routers to hold any changes that might impact routes for a of period time.

Now that you understand some of the issues with distance-vector protocols, let's look at one of the most common distance-vector protocol, RIP.

RIP (Routing Information Protocol)

- RIP is a distance-vector protocol.
- Hop count is used as the metric.
- Maximum hop count is 15.
- Routing updates are every 30 seconds.
- RIP can load balance over multiple paths.

Routing Information Protocol (RIP) is a basic protocol used for routers to exchange routing information, and the details of RIP are specified in RFC 1058. Let's look at some of the key characteristics of RIP:

- RIP is a distance-vector protocol and uses hop count as the metric. This is an important limitation of RIP; the only thing it uses to determine the shortest path is the number of hops or the number of routers a packet has to go through. It does not take into consideration bandwidth. So if I have one route that goes through two routers that are connected via 56k lines and another route that goes through three routers that are all connected via T3s, RIP will only look at hop count and say two is less than three and send the data over the 56k connection.
 - The maximum hop count for RIP is 15; a hop count of 16 is considered unreachable. So with large networks where there are more than 15 routers or 15 possible hops a packet can go through, RIP will not work.
 - RIP works by sending routing updates to all of a router's neighbors every 30 seconds.
 - RIP can also load balance over multiple paths if they are equal in terms of metrics.

As you can see, RIP is not a complicated protocol, but it has limitations because of its simplicity.

Link State

- SPF (shortest path first) algorithm
- Maintains topology information
- Has full knowledge of all routers and how they connect
- All routers have a similar picture of the entire network.

We discussed distance-vector protocols and looked at an example of RIP. As you have seen, they are fairly basic, but are also limited. The second type of routing protocols are link state, and these overcome the limitations of distance-vector protocols but also add in complexity. Link state uses SPF, or the shortest path first algorithm. The way it works is that each router maintains a database that has topology information about the entire network. Each router not only knows about its neighbors, it also knows about all routers that are on the network and how they are connected. Because all routers maintain full knowledge, each router should have a similar picture of the network and therefore similar information.

Now let's look at a common link-state protocol, BGP.

BGP (Border Gateway Protocol)

- Specifies routing between autonomous systems or networks that are very large.
- Is an exterior gateway protocol (EGP)
- Performs three types of routing:
 - Interautonomous system routing
 - Intra-autonomous system routing
 - Pass-through autonomous system routing

Border Gateway Protocol (BGP) is an exterior gateway protocol that determines how routing should be performed between autonomous systems. An autonomous system is a network or groups of networks that are under the control of a single entity. The Internet is composed of a large number of autonomous systems that are interconnected. BGP performs three general types of routing:

- **Interautonomous system routing:** Interautonomous system routing occurs between two or more BGP routers in different autonomous systems. Peer routers in these systems use BGP to maintain a consistent view of the internetwork topology. BGP neighbors communicating between autonomous systems must reside on the same physical network.
- **Intra-autonomous system routing:** Intra-autonomous system routing occurs between two or more BGP routers located within the same autonomous system. Peer routers within the same autonomous system use BGP to maintain a consistent view of the system topology. BGP also is used to determine which router will serve as the connection point for specific external autonomous systems.
- **Pass-through autonomous system routing:** Pass-through autonomous system routing occurs between two or more BGP peer routers that exchange traffic across an autonomous system that does not question and run BGP. In a pass-through autonomous system environment, the BGP traffic did not originate within the autonomous system in is not destined for a node in the autonomous system.

Distance Vector Versus Link State

- Distance vector
 - Has information only on neighbors
 - Simple metric, such as hop count
 - Frequent updates
 - Slow convergence
- Link state
 - View of entire network
 - Calculates shortest path to each router
 - Event-triggered updates
 - Fast convergence

Now that we have covered both distance-vector and link-state protocols, let's summarize this section by taking a brief comparison of the two. From a simplicity standpoint, distance vector is simpler, but it also does not scale as well to larger networks. Distance vector only has information about each of its neighbors, whereas link state protocols have a view of the entire network. Distance vector uses a simple metric such as hop count and does not include critical elements such as bandwidth. Link state calculates the shortest path to each router and looks at various elements such as bandwidth and congestion. A distance vector protocol automatically updates on frequent intervals whether there is a change or not, which results in slow convergence. Link state updates the routing tables only when certain events occur and therefore can converge much quicker.

Remote Access

- Data networking technologies
- Focused on providing remote users with network access
- Protects confidentiality, availability, and integrity
- Restricted address:
 - Authenticates user node (not the user)
 - Determines authorized users based on source IP address
 - Allows access to addresses on approved list

Data Networking Technologies

- Focused on providing remote users with network access
- Protects confidentiality, availability, and integrity
- Restricted address
 - Authenticates user node (not the user)
 - Determines authorized users based on source IP address
 - Allows access to addresses on approved list

Remote Access Security Methods

- Caller ID
 - Compares the phone number of an incoming call to a number on an approved list
- Callback
 - Source user initiates a communication by providing an ID or password
 - Access server hangs up and calls back the user using a listed phone number.
 - Authenticates the user node, not the user

Caller ID and callback are ways to protect remote users and can work even if someone is traveling

Virtual Private Networks (VPNs)

- Data is encrypted at one end of the VPN from cleartext into ciphertext.
- Ciphertext is transmitted over the Internet.
- Data is decrypted at the other end of the VPN from ciphertext back into the original cleartext.

VPNs are a perfect alternative to costly, inflexible private circuits. They give companies the option of setting up virtual circuits across public networks, such as the Internet. Encryption provides the confidentiality needed as the private information flows across the public network. This capability allows VPNs to establish secure communication between different remote organization offices and can be used to establish remote access to internal network resources by employees from their homes or while they travel.

VPN Advantages

- Improved flexibility
 - A VPN "tunnel" over the Internet can be set up rapidly. A frame circuit can take weeks.
 - A good VPN will also support quality of service (QoS).
- Lower cost
 - There are documented cases of a VPN paying for itself in weeks or months.
 - There are also cases in which the hidden costs sink the project!

One of the biggest benefits of VPN technology is its flexibility. If you need a secure channel between two hosts for only a day, or even an hour, a VPN may fit the bill. After you have all the components to establish a VPN, setting one up only requires configuration. This makes the technology far more flexible than private circuits, which must be ordered far in advance of their use and may require additional hardware. This flexibility lends itself to creating new business solutions. For example, it's not cost-effective to wire a T1 for every employee who works from home. It's practical, however, to load software on their laptop and let them connect to the home office via a VPN over the Internet.

There are also some disadvantages to VPNs, the primary of which is performance guarantees. Most private circuits, such as leased-lines or ATM, have an ability to guarantee bandwidth and latency. Similar guarantees have been difficult to achieve with VPNs. TCP/IP, the networking protocol for the Internet, was not designed to provide *quality of service* (QoS) and improvements have been slow in coming. Providing QoS for VPNs is even more difficult because many QoS solutions require the service provider to look into the messages they are passing on to decide whether the message has higher priority than other messages. If the service provider cannot examine the information in a message (because of encryption), it makes it even more difficult to decide which network traffic should get priority.

There are solutions to these problems. *Multiprotocol Label Switching* (MPLS), an alternative over traditional Layer 3 routing, is used to address these problems. It allows forwarding of messages across the Internet without requiring examination of the message contents. MPLS-based VPNs can be purchased from a wide variety of Internet service providers, although they are more expensive than standard IP services.

Modes of Remote Access

- Client-to-site VPN (transport)
 - Example: Laptop dial-up connection to remote access server at HQ
- Site-to-site VPN (tunnel)
 - Example: L.A. office connection to D.C. office location

Modes of VPNs

There are two primary categories of VPNs to consider: client-to-site and site-to-site:

- **Client-to-site VPNs** provide remote access from a remote client, such as a traveling sales rep or telecommuting employee, to the corporate network. Such VPNs are normally established between the client's computer and a gateway device located at the border of the corporate network. The client's computer runs VPN software that allows it to establish the connection to the VPN gateway.
- **Site-to-site VPNs** provide connectivity to networks, such as headquarters and a remote office. In these connections, gateway devices are located in front of both networks. Information needing to flow between the sites is directed to the local gateway, which then encrypts the contents of the message and forwards it to the other site's gateway. The remote site's gateway decrypts the message then sends it on to its final destination.

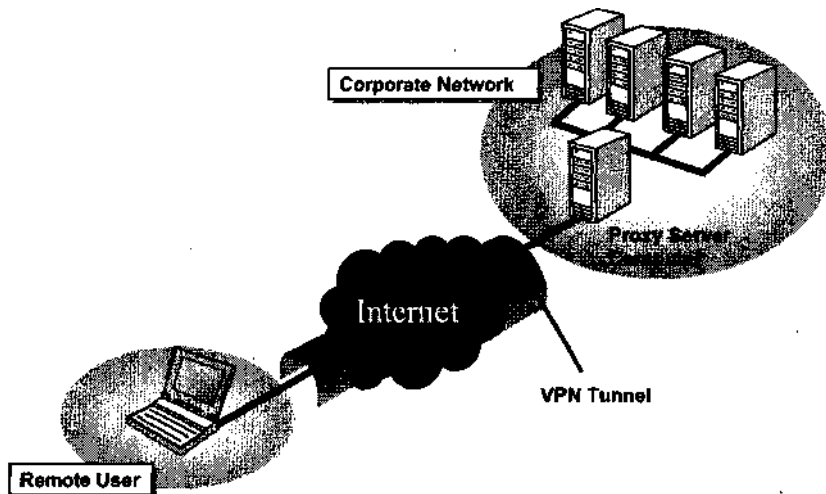
Types of Remote Access

- Dial-up, async, and remote Internet connectivity
- Digital subscriber line (xDSL)
- Integrated Services Digital Network (ISDN)
- Wireless computing, mobile and cellular computing, and personal digital assistants (PDAs)
- Cable modems

There is a third, less common type of VPN, the client-to-client VPN. These VPNs establish a protected link between two specific computers. As such, they could be considered the most secure of the VPN types, because in the client-to-site and site-to-site VPNs, part of the path between the transmitter of a message and the receiver of the message is unencrypted. For instance, in client-to-site VPN, the communication from the client's computer to the VPN gateway is protected, but the message travels unencrypted (and unprotected) from the VPN gateway to the internal corporate server the client is trying to communicate with. If an attacker inserts herself somewhere between the VPN gateway and this server, she would be able to eavesdrop or modify the contents of the message.

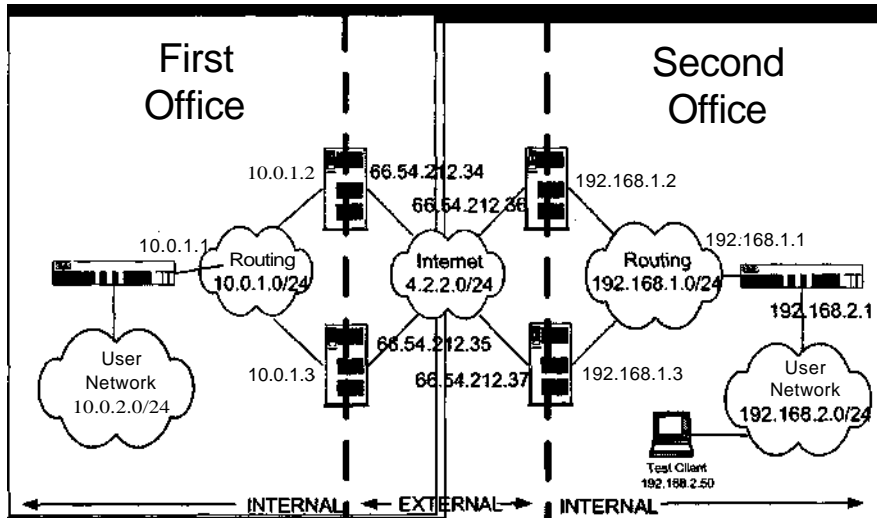
If client-to-client VPNs are more secure, why are they not used more often? The majority of the reason is the configuration required. Each pair of hosts wanting to communicate must be specifically configured to allow the communication. The most important part of this configuration is key installation. Each host must have a separate unique key that it can use to encrypt information to a particular destination host. Because of this, client-to-client VPNs between every two hosts would quickly become unmanageable as the number of hosts increases, if manual configuration is used.

Remote Access (Generic)



This slide shows how remote access works and is set up for clients. This slide depicts what was written in the previous slide. It shows a home user connecting to a corporate network via the Internet and setting up a secure channel. This mode of operation is often called *transport mode*.

Remote Access (Site-to-Site Redundant)



This slide shows how remote access works and is set up for servers. In this slide, two offices are setting up a secure connection over the Internet. This mode of operation is often called *tunnel mode*.

Remote-Access Security Management

Securing enterprise and telecommuting remote connectivity:

- **Securing external connections (such as VPNs, SSL, SSH, and so forth)**
- **Remote-access authentication systems (such as RADIUS and TACACS)**
- **Remote-node authentication protocols (such as PAP and CHAP)**

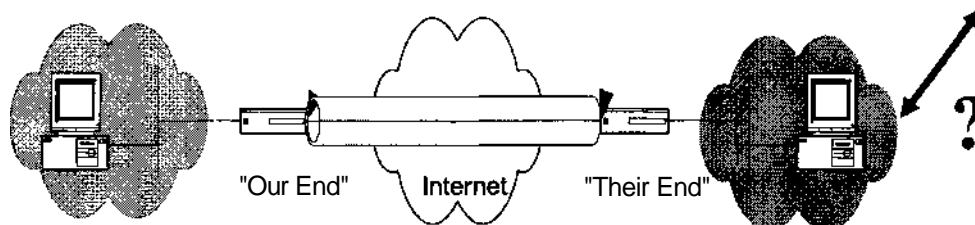
When it comes to security, there is no such thing as perfect security. Every security measure that we deploy has potential weaknesses or security implications that can be attacked. That is why we always want to deploy our security devices in a defense-in-depth architecture. This is where we have many devices working together to protect our network.

Remote access is no exception. There are also security issues associated with remote access:

- The first security issue is with securing external connections (such as VPNs, SSL, SSH, and so forth). If the endpoints cannot be secured, they can easily be attacked and used to gain access to a network.
- The second issue is remote-access authentication systems (such as RADIUS and TACACS). The authentication is only as good as the strength of the passwords and the integrity of the systems that store those passwords.
- The third issue is remote-node authentication protocols (such as PAP and CHAP).

Security Implications

- Bypassing firewalls, IDSs, virus scanners, and web filters
- Trusting the "other end"



Security Implications

Many sites assume that because they have established a VPN, they are secure. This is a bad assumption, because VPNs bring their own special security concerns into your network. One frequent error made with VPNs is to overly trust the other side of a VPN connection.

With site-to-site VPNs, it is common to see the VPN connection allowed into the network without applying any security restrictions to it. This might be appropriate if the other side of the VPN belongs to the same organization and is controlled by the same security policies and procedures. If the other side of the connection is another organization, such as a business partner, however, access through the VPN should be restricted. Most VPN gateways include firewall capabilities that allow them to limit network traffic across the VPN. It is a best practice to restrict this traffic to the minimum necessary to fulfill the business need of the connection.

Another potential security problem VPNs introduce is caused by the encryption VPNs use to protect the messages they exchange. As mentioned before, this encryption prevents an attacker from eavesdropping, but it also prevents intrusion detection systems and antivirus tools from examining the packets for malicious or inappropriate content. This reduces or eliminates the effectiveness of these security tools.

Last, client-to-site VPNs suffer from the trusted client problem. Many organizations have strict rules about the type of software allowed on corporate computers. Part of the reason for these controls is that unauthorized software may contain security vulnerabilities. When allowing employees to use a VPN to access the corporate network, the organization may not be in the same position to dictate a tight configuration. In fact, most home computers are insecurely configured. If an attacker discovers the home computer and takes it over, the attacker may be able to use that access to the computer to leverage access to the corporate network over the employee's VPN connection. For this reason, it is a good idea to recommend, or better yet, enforce the use of a personal firewall product and antivirus software prior to allowing remote users to access client-to-site VPNs.

Now that you understand how encryption can be used to protect communications over a network, it's time to introduce some concrete examples of technology that implements these ideas. The first is IPSec, the current industry standard for setting up VPNs.

IPSec Overview

- Issued by IETF as an open standard (RFC 2401), thus promoting multivendor interoperability
- Enables encrypted communication between users and devices
- Implemented transparently into network infrastructure
- Scales from small to very large networks
- Commonly implemented (most VPN devices and clients are IPSec-compliant)

IPSec

IP Security (IPSec) is an IETF standard for establishing virtual private networks. It is slowly replacing proprietary VPN protocols and becoming the industry standard. Many products on the market now support IPSec natively, such as Checkpoint Firewall-1, Cisco routers, and Windows XP.

Like the application-level and transport-level techniques previously discussed, IPSec provides data integrity, confidentiality, and authentication. IPSec also offers sophisticated replay attack prevention.

Note

Attackers use replay attacks by copying a message as it goes across the network, then by retransmitting the copy to the destination. Even if the attacker cannot read the encrypted message, he can cause undesired results. For example, if the message is a request to transfer \$1000, the replay might be able to cause an additional transfer, making the total transferred \$2000. IPSec includes specific mechanisms to detect and prevent replay.

Types of IPSec Headers

Authentication Header (AH)

- Data integrity: No modification of data in transit
- Origin authentication: Identifies where data originated

Encapsulated Security Payload (ESP)

- Data integrity: No modification of data in transit
- Origin authentication: Identifies where data originated
- Confidentiality: All data encrypted



The Protocols of IPSec

IPSec is actually a collection of protocols used singly or together to implement its various network security services. Primarily, IPSec is composed of the *Authentication Header (AH)* protocol, the *Encapsulated Security Payload (ESP)* protocol, and the *Internet Key Exchange (IKE)* protocol. To understand how IPSec works, let's examine the capabilities offered by each of these protocols.

Authentication Header (AH)

AH provides message integrity, anti-replay, and source authentication. It works by adding authentication information into each IP packet. To see how this works, we need to understand some of the information that goes into an IP packet.

IP packets are composed of many pieces of information, each of which is important. One of the most important, from a security standpoint, is the Source IP field. The Source IP field is used to tell the recipient who sent the message. In a normal network conversation, the computer that sends a message uses its own IP address as the source address. This is important to the security of the system because many firewall systems use source IP addresses to determine whether a message should be allowed into a network. If an attacker can choose to lie about his IP address, he can potentially use an address that the firewall does allow in, fooling the firewall into accepting a message that it should have denied. Without AH, there is nothing to prevent an attacker from lying about the source or any other field inside the packet.

To prevent this, AH adds a keyed hash of the message to the packet. This hash is referred to as the *integrity check value (ICV)*. In the ICV computation, AH includes every field that does not change during its trip from source to destination. This includes the source address, destination address, length, and the data. This information is inserted into the packet after the regular IP header, but before the data.

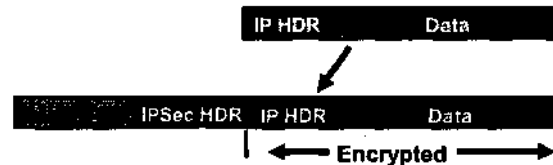
To verify that the packet has not been tampered with, the recipient recomputes the ICV. If any of the hashed fields, including the source address, have been changed, even by a bit, the hash will be different and the integrity check will fail. This provides both integrity checking and authentication. The integrity is guaranteed because the hash must match the message.

Types of IPSec Modes

- **Tunnel mode:** Applied to an IP tunnel

- Outer IP header specifies IPSec processing destination.
- Inner IP header specifies ultimate packet destination.

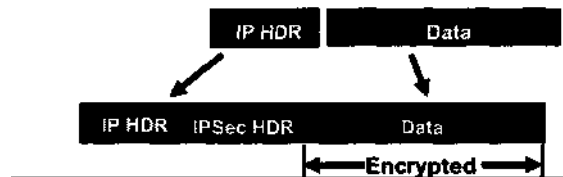
Tunnel mode



- **Transport mode:** Between two hosts

- Header after IP header, before TCP/UDP header

Transport mode



Modes of IPSec

Both AH and ESP can operate in two modes: transport mode or tunnel mode. Transport mode is used to protect a conversation between two specific hosts on a network. For example, two hosts using ESP in transport mode would establish a client-to-client style VPN. Up to now, all IPSec examples have been based upon transport mode. Tunnel mode is used to establish site-to-site and client-to site VPNs. Let's take a look at how tunnel mode differs from transport mode for both AH and ESP.

How Tunneling Works

Tunnel mode, as the name implies, sets up virtual tunnels between gateways. Tunnel mode works by accepting an entire IP packet, which is then packaged inside an IPSec packet. This new IPSec packet is not addressed to the destination of the packet it is carrying. Instead, its destination address is the address of the gateway system at the other side of the tunnel. When the destination gateway receives a tunnel packet, it unpackages it to get out the original packet. This packet is then routed onward to the host listed in its Destination field. From this original packet's point of view, the trip across the tunnel represents just one hop, regardless of how many intermediate routers may have actually existed between the two gateways.

Network Availability

Network availability refers to the area of the Telecommunications, Network and Internet Security domain that directly affects the InfoSec tenet of availability.

- RAID
- Backup functionality

Network availability refers to the area of the Telecommunications, Network, and Internet Security domain that directly affects the InfoSec tenet of availability. If the network is not available when people need it, it does not matter whether the confidentiality and integrity of the information is preserved; it is useless if authorized people cannot access it.

To maintain the availability of your network, you must take a proactive measure and a reactive measure. The proactive measure is RAID, which allows your system to potentially continue operating if one of the disks fails. Backup functionality, a reactive measure, enables you to recover if your entire system fails. The backups can be used to recover all of your data to a new system.

Redundant Array of Inexpensive Disks (RAID)

- Provides fault tolerance and protection against file server hard disk crashes
- Improves system performance by caching and distributing disk reads from multiple disks.
- Can be implemented either as a hardware or a software solution
- Failure Resistant Disk Systems (FRDS) and Failure Resistant Disk Systems Plus (FRDS+) are defined by the RAID Advisory Board.

Redundant Array of Inexpensive Disks (RAID) is a method used to provide fault tolerance if one of the hard drives crashes on your system. RAID will protect against only a hard drive failure, not failures in any other hardware components. Because the hard drive is where your data is stored, however, this is usually a critical component to protect, and RAID helps eliminate the chance of data loss across your system.

RAID also improves system performance by caching and distributing disk reads from multiple disks. It can also be implemented either as a hardware or software solution.

Not all RAID solutions work the same way or provide the same level of protection, so we look at different types over the next several slides.

RAIDO

RAID level 0:

- Creates one large disk by using several disks
- Separates the data into multiple units and stores it on multiple disks by using a process called *striping*
- Stripes data across all disks (but provides no redundancy) by using all the available drive space to create the maximum usable data volume size and to increase read/write performance

RAID level 0 is often referred to as *striping*. Even though this is a RAID level, this level provides no redundancy or protection of your data; that is why it is level 0. RAID level 0 creates one large disk by using several disks. It then separates the data into multiple units and stores it on multiple disks by using a process called striping.

This method provides increased performance by maximizing the usable space to store data and also increases read and write performance; however, it provides no protection against data loss.

RAID 1

RAID level 1:

- Commonly called *mirroring*
- Duplicates the data from one disk or set of disks to another disk or set of disks
- Often implemented by a one-for-one disk ratio
- Each drive is mirrored to an equal drive partner, which is continually updated with current data.
- If one drive fails, the system automatically gets the data from the other drive.

RAID level 1 is often called *mirroring* because it mirrors the data from one disk to another. Essentially it creates a duplicate copy of your data across two disks. This is the first level that provides protection against data loss, but it performs this protection in a straightforward manner.

To protect your data using this method, there is a one-to-one relationship between active disks and backup disks. If you have three disks that you want to implement RAID 1 on, you need a total of six disks to do this. This is the case because each drive is mirrored to an equal drive partner, which is continually being updated with current data. If one drive fails, the system automatically gets the data from the other drive.

RAID 2

RAID level 2:

- Consists of bit-interleaved data on multiple disks.
- Parity information is created using a hamming code, which detects errors and establishes the part of the drive in error.
- Defines a specific disk drive system with 39 disks: 32 disks of user storage and 7 disks of error-recovery coding.

RAID level 2 provides protection of data by interleaving the data at a bit level across multiple disks. This is not a general method of protecting your data, however, but a specific method in which a certain number of disks are required across the system. To implement RAID level 2, you need a total of 39 disks. Of these, 32 disks are used for storage of data, and 7 disks are used for error recovery of that data. The error checking is done through parity information created using a hamming code, which detects errors and establishes which part of which drive is in error.

Because this method is performed at a bit level, it is not as efficient as other methods.

RAID 3 and 4

RAID levels 3 and 4:

- Data is striped across several drives.
- Parity check bit is written to a dedicated parity drive.
- Level 3 is implemented at the byte level, and level 4 at the block level.
- If a hard disk fails, the data can be reconstructed by using the bit information on the parity drive.
- Spare drives can be used to replace crashed drives.

RAID levels 3 and 4 operate and protect the data in a similar manner; the only difference is that level 3 works at the byte level and level 4 operates at the block level. Each approach has advantages and disadvantages. The smaller the unit, the more granular errors can be tracked and the less amount of data has to be replicated. However, the smaller the unit, the less efficient it works because more information has to be tracked.

Using the approach, the data is striped across several drives, and a parity check bit is written to a dedicated parity drive.

These levels do not require a set number of drives like RAID level 2. Essentially, if a hard disk fails, the data can be reconstructed by using the bit information on the parity drive, and spare drives can be used to replace crashed drives.

RAID 5

RAID level 5:

- It stripes the data and the parity information at the block level across all the drives in the set.
- Parity information is written to the next available drive rather than a dedicated drive using an interleave parity.
- RAID 5 allows for more flexibility in the implementation and increases fault tolerance because the parity drive is not a single point of failure now.
- Disk reads and writes are performed concurrently, increasing performance over levels 3 and 4.

RAID level 5 is often called *interleave parity* and builds upon some of the prior methods discussed. This method does not use dedicated drives for data and dedicated drives for error information as other methods do. This method interleaves both the data and the error information (or parity information) across all the drives at the block level.

This method allows for more flexibility in the implementation and increases fault tolerance because the parity drive is not a single point of failure. However, this method is not as easy to implement and can be more complex during the configuration.

RAID 7

RAID level 7:

- Variation on RAID 5, but the array functions as single virtual disk in hardware
- Sometimes simulated by software running over a RAID level 5 hardware implementation
- Allows the drive array to continue to operate if any disk or any path to any disk fails
- Also provides parity protection

RAID level 7 is often called *single virtual disk* and is considered an enhancement to RAID 5. This method is similar to RAID level 5, except that the array functions to create an abstract level, so that from the user perspective, there is a single virtual disk. This method can be implemented in either hardware or software, but hardware is more expensive (also much faster) than software.

This method allows the drive array to continue to operate if any disk or any path to any disk fails. This is the case because the drives are not viewed as separate components but as a single, virtual component.

RAID (Other)

- Some RAID types combine the features of several RAID levels:
 - Level 10 is created by combining level 0 (striping) with level 1 (mirroring).
 - Level 6 is created by combining level 1 (mirroring) with level 5 (interleave).
- These are not defined standards:
 - RAID/7

Many hybrid RAID configurations are available based on specific needs. There are too many to list exhaustively here, but it is important to point out that if none of the methods discussed fits your need, there is probably a hybrid level that will. If there isn't, you can do what those who came before you did: Come up with your own.

Some of the more popular combinations are as follows:

- Level 10 is created by combining level 0 (striping) with level 1 (mirroring).
- Level 6 is created by combining level 1 (mirroring) with level 5 (interleave).

It is important to note that these are not defined standards like the other methods.

RAID Summary

RAID Level	Description
0	Striping
1	Mirroring
2	Code parity
3	Byte-level parity
4	Block-level parity
5	Interleave parity
6	Second independent parity
7	Single virtual disk

This slide summarizes the common RAID levels. RAID 0 is often called *striping* because it stripes the data across the disk, providing more efficient read and write operations (but no protection of data). RAID 1 is often called *mirroring* because data is mirrored across a second disk and has a one-to-one relationship between the primary and mirrored disk. RAID 2 is called *code parity*. RAID 3 is called *byte-level parity*. RAID 3 is called *block-level parity*. RAID 5 is called *interleave parity*. RAID 6 is called *second independent parity*. RAID 7 is called *single virtual disk*.

As you saw in the previous slides, each level has different advantages and disadvantages to its approach.

High Availability

- Electronic vaulting
 - Batch process of electronically transferring backup data to off-site location
- Remote journaling
 - Electronically transmitting data to off-site storage location as data is obtained and written to local storage (not in batch mode)

Electronic vaulting: Batch process of electronically transferring backup data to an off-site location.

Remote journaling: Electronically transmitting data to an off-site storage location as data is obtained and written to local storage (not in batch mode).

High Availability (2)

- Database shadowing
 - Similar to remote journaling
 - Stores data to multiple servers for redundancy
- Redundant servers
 - Uses RAID level 1 mirroring on two servers

Database shadowing

- Similar to remote journaling
- Stores data to multiple servers for redundancy

Redundant servers

- Uses raid level 1 mirroring on two servers

Server Fault-Tolerant Systems: Redundant Servers

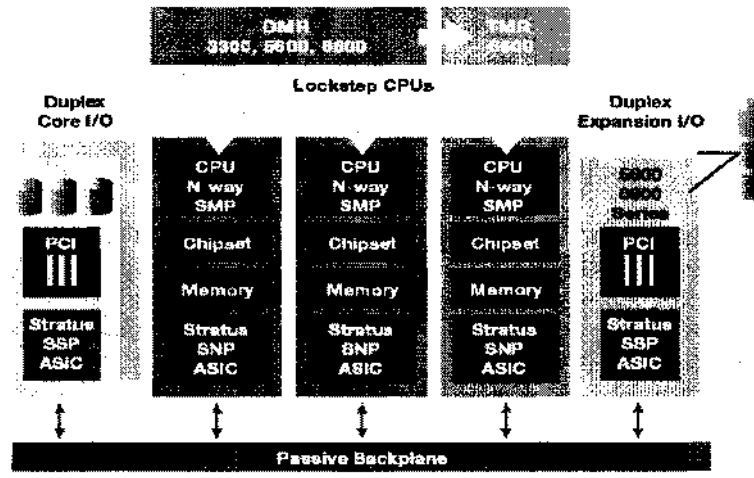
- Redundant server implementation takes the concept of RAID 1 (mirroring) and applies it to a pair of servers.
- The primary server mirrors its data to a secondary server, thus enabling the primary to "roll over" to the secondary in case of primary server failure. (The secondary server steps in and takes over for the primary server.)
- Rollover can be hot or warm (the rollover may or may not be transparent to the user, depending upon the vendor's implementation of this redundancy).

RAID is good level of protection if a hard drive fails, but what if other components on your system fail? Even though the focus of RAID is on the protection of your data if a network card fails and even though it does not impact your data, the data is not accessible and will cause a denial-of-service attack against your data. All the hardware components of a network are critical in terms of allowing authorized users access to your data whenever they need it.

Server fault tolerance focuses on a higher level than RAID, looking at ways to protect all aspects of your servers by clustering many servers together. If a component on a server fails because there are redundant servers, one of the other servers can take over while the problem is fixed.

Redundant servers create a form of protection that essentially implements RAID 1 or mirroring across servers. For each server, there is a backup server that can take over if it fails. In this mode, the primary server mirrors its data to a secondary server, thus enabling the primary to "roll over" to the secondary in case of primary server failure. (The secondary server steps in and takes over for the primary server.) Rollover can be hot or warm (for instance, the rollover may or may not be transparent to the user, depending upon the vendor's implementation of this redundancy).

Example of Redundant Servers



This slide shows how you can set up redundant servers. Essentially, it is a straightforward configuration in which you have several identical servers on your network. The key aspect of having redundant servers is how the data actually gets replicated across the systems. Having an identical hardware platform to take over if a hardware component fails on your main server is important, but more important is how the data replicated across those systems.

In an ideal mode, it is done automatically so that there is no down time, and it is transparent to users. In a less automated mode, however, the backup server would sit offline and if the primary system fails, the data is restored from tape and the new system is put back online. This type of system requires some downtime, but less downtime than if there weren't a redundant system available.

Server Fault-Tolerant Systems: Server Clustering

Server clustering:

- Server clustering is group of independent servers managed as a single system, providing for higher availability, easier manageability, and greater scalability.
- The cluster looks like a single server from the user's point of view.
- If any server in the cluster crashes, processing continues transparently.
- Server clustering is similar to redundant servers, except that all the servers in the cluster are online and take part in processing service requests.
- The cluster acts as an single entity and balances the traffic load to improve performance.

Server fault tolerance focuses on a higher level than RAID, looking at ways to protect all aspects of your servers by clustering many servers together. If a component on a server fails because there are redundant servers, one of the other servers can take over while the problem is fixed.

Some of the key characteristics of server clustering is that it consists of a group of independent servers managed as a single system, providing for higher availability, easier manageability, and greater scalability. It is similar to redundant servers, except that all the servers in the cluster are online and take part in processing service requests. The cluster acts as a single entity and balances the traffic load to improve performance.

Backup Concepts: Full Backup

- Makes a complete backup of every file on the server every time it's run
- Primarily run when time and tape space permits and is used for system archive or baselined tape sets

A full backup is the most comprehensive type of backup and requires the fewest number of tapes to restore your data. The problem with this type of backup is that it takes a considerable amount of time to back up all of your data. A full backup backs up all of your data and therefore needs only one tape to restore your data if the system crashes. A full backup should be performed at least once a week.

A full backup each day is ideal; there is usually not enough time to do a full backup each day, however, because backups must be performed during company off-hours.

Backup Concepts:

Incremental Backup

- Copies only files that have recently been added or changed that day and ignores any other backup set
- Usually resets the archive bit on the files after they have been backed up
- Used if time and tape space is at an extreme premium, but has inherent vulnerabilities

An incremental backup is the most efficient type of backup because it backs up the least amount of data each day. The drawback to this type of backup is that it requires the most number of tapes to restore your data.

When performing an incremental backup, only the data that has changed since the last backup is backed up. Usually this is done through the use of an archive bit. After a full backup is performed, the archive bit is cleared. Any time a file is modified or created, the archive bit is set. An incremental backup backs up only those files that have the archive bit set and then clears the archive bit.

Suppose, for example, that a full backup is performed Friday night; all of your data is backed up. An incremental backup on Monday backs up only data that has changed since Friday. An incremental backup on Tuesday backs up only data that has changed since Monday. An incremental backup on Wednesday backs up only data that has changed since Tuesday. Now if your system crashes on Thursday, you just need the full backup from Friday plus the incremental backups from Monday, Tuesday, and Wednesday.

Backup Concepts:

Differential Backup

- Copies only files that have changed since a full backup was last performed.
- Backup is "additive" because the time and tape space required for each night's backup grows during the week as it copies the day's changed files and the previous days' changed files up to the last full backup.
- File's archive bit is not reset until the next full backup.

A differential backup is a third way to back up your data across your network and is a cross between a full backup and an incremental backup. With a full backup, all your data is backed up on a daily basis. This is ideal, but it requires considerable resources. An incremental backup backs up all data that has changed since the last backup and takes considerably fewer resources to perform. The problem with an incremental is that when you need to restore your data, you need many tapes.

A differential backup takes advantage of the fact that during the course of normal operation only a small percent of your files actually change (and therefore, doing a full backup each day is inefficient). Therefore, each time a differential backup is performed, it backs up all data that has changed since the last full backup. The advantage is that now if your system has to be restored, it requires only two tapes: the last full backup and the last differential backup.

Suppose, for example, you perform a full backup on Friday night; all of your data is backed up. If a differential backup is performed on Monday, all data that has changed since Friday is backed up. If a differential backup is performed on Tuesday, all data that has changed since Friday is backed up. If a differential backup is performed on Wednesday, all data that has changed since Friday is backed up. Now if your system crashes on Thursday, only the full backup from Friday and the differential backup from Wednesday is needed to restore your data.

Summary

- Defense and attacks:
 - Intrusion detection and response
 - Firewalls
 - Methods of attack
- Networking methodologies:
 - Topologies
 - Transmission
 - Hardware
 - Theory and design

This section covered a broad range of topics that are critical to network security. To provide proper defenses, you need to understand the components at your disposal that you can use to implement security. The common devices used today are firewalls and intrusion detection systems. In this section, you learned how these integrate to provide a robust intrusion prevention and detection process across your organization. To create this type of system, you need to understand the various attacks and how they work so that you can build proper defense methods to protect against them.

Understanding how to design and configure a network is also critical. We first looked at various topologies from both a logical and physical standpoint and the relationships between them. We covered the critical protocols comprising the TCP/IP protocol suite and showed how they interoperate together. This section ended with a look at how to maintain redundancy across your network.

3. Information Security and Risk Management

10 Domains of Knowledge

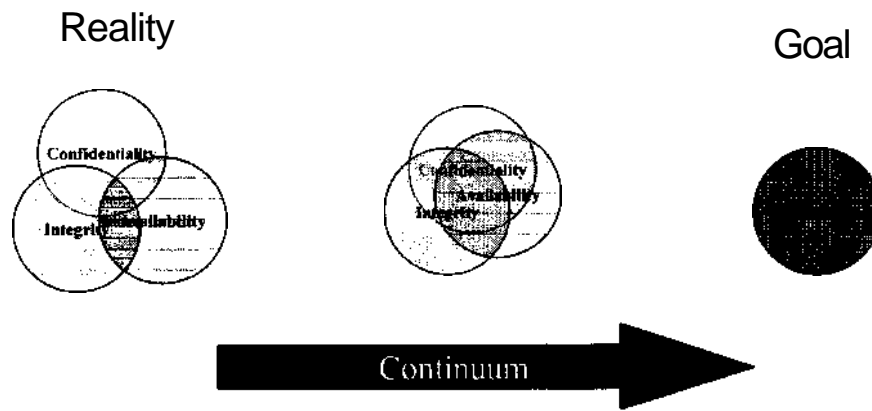
This section covers Domain 3, the Security Management Practices domain.

3. Information Security and Risk Management

- Overview of network security
- Critical components
- General security principles

Security technologies should be implemented according to policies set by upper management. However, technologies should not be installed before an analysis of how well they protect the assets ' and information that is most valuable is completed. Implementing technologies without regard to the organization's needs for confidentiality, integrity, and availability, and without regard to the business itself, is poor management.

Security Objectives



Prioritize across all three areas of security.

The three main objectives of security are confidentiality, integrity, and availability. These form the CIA triad. Many organizations focus on one area more than the others. For example, intelligence agencies are concerned with confidentiality. Financial institutions are focused on accuracy or integrity, and e-business sites emphasize availability. As security professionals, we need to integrate the three elements of the CIA triad together to achieve Defense-in-Depth. The key is to achieve a proper balance of the three. Maximizing availability can sometimes compromise confidentiality. Implementing strong integrity measures, such as error checking, may have an impact on availability if throughput is affected. Requirements for all three categories must be carefully weighed before technologies are implemented.

Concepts and Terminology

- Confidentiality \neq disclosure
- Integrity \neq alteration
- Availability \neq destruction
- Identification
- Authentication
- Authorization
- Accountability

(In these cases, "not equals" means "logical opposite/0

Confidentiality, integrity, and availability can also be expressed as disclosure, alteration, and destruction (D.A.D.). Before we move on, let's quickly define some basic terms used in this chapter. Following are some terms you should know:

- *Confidentiality* — Prevents the intentional or unintentional unauthorized disclosure of data.
- *Integrity* — Deals with the accuracy of information, ensures modifications are not made to data by unauthorized personnel or processes, and ensures that data is internally and externally consistent.
- *Availability* — Ensures the reliable and timely access to data or computing resources by the appropriate personnel.
- *Identification* — Is the means by which users claim their identities to a system. Most commonly used for access, identification is necessary for authentication and authorization. Authentication establishes, tests, or reconciles a user's identity.
- *Accountability* — Is the system's capability to determine the actions and behavior of a single individual within a system. Accountability shows that a particular individual performed a particular action. Audit trails and logs support this concept.
- *Authorization* — Deals with the rights and permissions granted to an individual (or process) that enable access to a computer.

Concepts

- Confidentiality
 - Prevents the intentional or unintentional unauthorized disclosure of a message's contents
- Integrity
 - Modifications are not made to data by unauthorized personnel or processes.
 - Unauthorized modifications are not made to data by authorized personnel or processes.
 - Data is internally and externally consistent.
 - Well-formed transactions
 - Ensures data can be changed only by a specific set of instructions

Confidentiality ensures that only approved people and processes have appropriate access to information. Labels are often used to define the level of controls needed to ensure confidentiality. The Federal government commonly uses FOUO (For Official Use Only), Sensitive but Unclassified (SBU), Confidential, Secret, and Top Secret. In the commercial sector, information is sometimes designated as proprietary or as trade secrets, whereby disclosure of the data may damage the company's profits in some way and prevent competitors from gaining an unfair competitive advantage.

Integrity ensures that data has not been intentionally or unintentionally changed. Message hashes, checksums, change control, and auditing are methods for ensuring integrity.

Concepts (2)

- **Availability**
 - Reliable and timely access to data or computing resources by the appropriate personnel
 - Confidentiality, integrity, and availability can also be expressed by listing the opposites, which are disclosure, alteration, and destruction (D.A.D.)
- **Privacy**
 - Confidentiality and protection of personally identifiable information

Availability refers to the ability to access the information whenever it is needed. Availability can be denied by either preventing access to the information or by actually destroying the data. Denial of Service attacks, hostile code, EMI, power outages, or brownouts are just a few examples of threats that can affect availability. It is important to note that natural disasters can lead to unavailability of a network or facility. The important thing to remember is that these events do not have to be intentional. Accidents or unintentional events can cause loss of availability for an organization.

- **Identification**
 - The means in which users claim their identities to a system
 - Most commonly used for access
 - Necessary for authentication and authorization
- **Authentication**
 - Testing or reconciliation of a user's identity
 - Establishing the user's identity

The difference between identification and authentication is the claim to be someone and the proof you are who you say you are. For example, we often use our driver's license as a means of authentication in every day life.

User IDs are a common form of identification. Schemes for assigning user IDs should be logical, but not necessarily easily guessed by those outside the organization. In many organizations, user IDs are paired with passwords, so if outsiders can figure out the user ID, they can gain access to your systems. Passwords, as mentioned, are a method of authentication or verifying a user's identity.

There are four categories of authentication:

1. Something you have (such as a token, smart card, or badge)
2. Something you are (biometrics: fingerprints, retina scans, voice, palm scans, hand geometry, and so on)
3. Something you know (passwords or phrases, for example)
4. A place you are (such as GPS)

Using two of these three categories is known as two-factor authentication. Using more than one of these three categories strengthens the level of security. When using biometrics, consider the Crossover Error Rate (CER) or the percentage of False Rejection Rate compared to the False Acceptance Rate. Ideally, false rejections and acceptances should be low.

Concepts (4)

- **Accountability**
 - A system's capability to determine the actions and behavior of a single individual within a system
 - To identify that particular individual
 - Audit trails and logs support accountability.
- **Authorization**
 - The rights and permissions granted to an individual (or process), which enable access to a computer

Accountability makes you responsible for your actions. Enabling auditing is not enough to ensure accountability. Someone must actually review the logs and identify violations. Violations must be reported to the appropriate authorities, and those authorities must then enforce the organization's rules.

Authorization determines which users or groups of users should have access to what group of information. This is usually based on the services and data a person requires to do his job or his "need to know." Users should not have the same level of access to the network as system administrators.

Data Classification

- Classification
 - Top secret
 - Highest level of information classification
 - Unauthorized disclosure can cause exceptionally grave damage to national security.
 - Secret
 - Unauthorized disclosure can cause serious damage to national security.
 - Confidential
 - Unauthorized disclosure can cause damage to national security.
 - Sensitive, but unclassified (SBU)
 - Unclassified, but disclosure does not cause damage to national security
 - Unclassified
 - Information designated as neither sensitive nor classified
 - Public release does not violate confidentiality.
- Commercial terms: public, official use only, internal use only, and company proprietary

We classify data with differing levels of sensitivity. Why do we put labels on our data? You can't protect all of it, so some data requires more protection than others.

- Subject label = Object label
- Permission is still required (need to know).

The reality is that there isn't an organization that has sufficient resources to protect all information with the rigor that the most sensitive information requires. Not all information requires the protection needed for nuclear weapon designs or war plans. Consequently, so that appropriate protections can be applied based on the sensitivity of the information and on the potential impact of loss, organizations often classify data using various levels. Loss might be measured in terms of confidentiality (what we usually think of regarding government or corporate secrets), but it can also be measured in terms of integrity or availability.

Governments and their militaries, such as the U.S. Department of Defense (DoD), started the phenomenon of labeling data to apply higher levels of protection to data that was so sensitive that if it were leaked, it could harm their countries' national security. Subsequently, this has also become commonplace in the corporate world. A quick listing of the DoD and Federal levels follows:

- **Top Secret:** The highest levels of protection are given to this data; it is critical to protect the data.
- **Secret:** This data is important, and its release could harm national security.
- **Confidential:** This is important, and if released, it may be detrimental to national security.
- **Sensitive But Unclassified (SBU):** This is sensitive information and should not be released (such as social security numbers).
- **Unclassified:** This is information that should not be released, but the nation would not be harmed if it were released

Corporations label their data, too. It is extremely difficult to protect all data in a company; however, some data definitely needs special protection and be identified as such. Perhaps you manufacture closed-source software; that source code might need special protection because its release could directly affect revenue. It might damage the morale of company employees if everyone learned the salaries of their co-workers. Do the employees earn the same amount of money?

Data Classification Criteria

- Value
- Age
- Useful life
- Personal association — Personally associated with specific individuals or is addressed by a privacy law

How is data classified?

- Value: What is the information worth to the company? What if it is lost or compromised?
- Age: How current is the information? Does your organization need data that is 5 years old? Is real-time information more important to your organization than information received last week?
- Useful life: At what point is data in your systems no longer worth protecting? We know hardware can become obsolete, but how often do we continue protecting outdated information?
- Personal association: Examples include medical records, case files, and personnel files.

Data Classification Process

1. Identify the administrator/custodian.
2. Specify the criteria for how the information will be classified and labeled.
3. Classify the data by its owner who is subject to review by a supervisor.
4. Specify and document exceptions to the classification policy.
5. Specify the controls that will be applied to each classification level.
6. Specify the termination procedures for declassifying the information or for transferring custody of the information to another entity.
7. Create an enterprise-awareness program about the classification controls.

The process of creating a classification scheme is shown in the previous slide. In addition to the administrator, other specific roles and responsibilities should also be identified. Information should not only be classified and labeled, but prioritized according to the organization's needs. Each step should be clearly and accurately documented.

Distribution of Classified Information

- Court order. Classified information might need to be disclosed to comply with a court order.
 - FOIA
- Government contracts
- Senior-level approval
 - WithNDA

Court orders or other legal mandates, such as FOIA requests (Freedom of Information Act), can require release of information that would otherwise remain protected.

Management can approve distribution of classified information outside of the organization, possibly in conjunction with a non-disclosure agreement. Such documents prevent organizations that work together from capitalizing on information they share for mutual benefit.

Data Classification Roles

- Data owner
- Custodian
- Application owner
- Manager
- User
- Administrator
- Analyst
- Auditors

The following are key data classification roles:

- Data owner
- Custodian
- Application owner
- Manager
- User
- Administrator
- Analyst
- Auditors

Data Classification Responsibility

Senior management is *ultimately responsible* for the success of an organization:

- Responsible for establishment of an organization's computer security program and goals
- Priorities to support the mission of the organization

Someone in your organization must be responsible for data classification. The person responsible is not necessarily the one who implements it or manages it; he is the one who dictates the requirements and makes sure they get done.

As with any security policy, senior management is responsible for implementing an effective and appropriate data-classification program. They must provide adequate funding and manpower to implement, maintain, and enforce the program policy when needed. Management should also oversee an audit program and receive periodic reports of violations.

Data Classification Roles: Owner

- An executive or manager of an organization
- Responsible for the asset of information that must be protected
- Has the final corporate responsibility of data protection

Most of us have heard the expression that "the captain goes down with the ship." This reflects the ultimate responsibility the captain has for the safety and operation of everyone and everything aboard that ship.

Management has the same sort of responsibility toward the business and its stakeholders. Management must take measures to adequately protect information and networks from significant threats.

Management can delegate authority to an assigned department or specific individuals for daily operations, but management is still accountable for mishandling of data. Even though this role is usually performed by someone higher up in the organization, it does not have to be done by management.

Data Classification Roles: Custodian

- Running regular backups and routinely testing the validity (integrity) of the backup
- Performing data restoration from the backups when necessary
- Maintaining those retained records in accordance with the established information-classification policy

The custodian conducts any activities regarding the maintenance of the data. A database or system administrator may be assigned the role of custodian. In addition to overseeing the backups, a custodian might be required to administer the classification scheme. The custodian is the person who provides the hands-on management of the data as dictated by the data owner. It is important to remember that the custodian is not the person who makes critical decisions; he simply implements the decisions about the data that the owner determines.

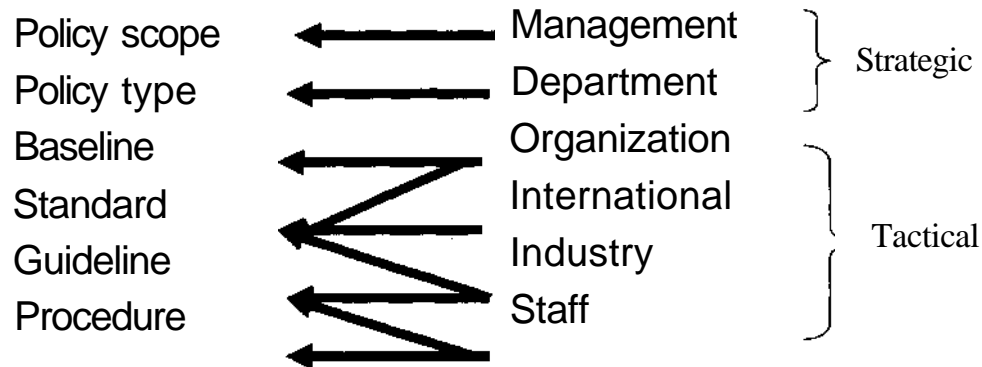
Data Classification Roles: User

Anyone (such as an operator, employee, or external party) who routinely uses the information as part of his job

Users are responsible for proper use of data and files. They should take adequate measures to protect the data, such as strong passwords, locking workstations when they aren't in use, using proper classification labels, and so on. Users are normally granted read access. Depending upon their job function, they might also have limited write, execute, and delete permissions.

Users should be properly trained in the organization's security policies and procedures, and they should be held accountable if they fail to adhere to those policies. Acceptable use policies are one way of making sure users know what is expected of them.

Policies, Procedures, Standards, Baselines, and Guidelines



The relationship between policies, procedures, standards, and so on is shown in this slide. The more general or overarching the document, the higher up the chain of command is the authority and responsibility for the document. Each of these categories is addressed in more detail in subsequent sections. It is also important to remember that each document should have consistent and complimentary information in them. The same information should not be repeated between the documents.

Policy Definitions and Issues

- Senior management directives
 - Create a computer security program.
 - Establish the goals of the program.
 - Assign responsibilities.
 - Consistent with other existing directives, laws, and an organization's mission.
 - Should be integrated with other organizational policies (such as personnel, privacy, appropriate use of equipment, and facilities)
- Senior management responsibility
- Policy is mandatory.

Policies define the strategic goals for the organization. Upper management defines the scope, distributes, and maintains the policies. These policies must be well defined and unambiguous because adherence to the policies is mandatory. The policy is based on the direction of senior management. Senior management dictates the high-level parameters of direction and focus. However, senior management does not do the work. It assigns the responsibilities to someone in the organization to perform or manage. Managers then assign the responsibilities to others who perform the work.

Defining a Policy

- Policies direct the accomplishment of objectives
 - Program policy
 - Issue-specific policy
 - System-specific policy

An effective and realistic security policy is the key to effective and achievable security.

Defining Security Policy

A *policy* is a *guideline* or *directive* that indicates a conscious decision to follow a path toward a specified objective. Often a policy can institute, empower resources, or direct action by providing procedures or actions to be carried out. With that in mind, this book attempts to provide guidance toward the goal of *developing a Basic Security Policy* for an organization, or even better, defining the existing one. The policy itself should be effective and realistic and have achievable security goals.

This section shows what a policy looks like in general or typical content within policies and common types of policies. It then focuses on the content that is specific to a security policy. This section is about documentation. It is critical to write down in a clear and concise manner what is expected of everyone in the organization when it comes to security. It is also helpful to inform people about what is expected of them, what the organization does, and what others in various roles within the organization do.

Defining a Policy (2)

- What makes up a policy?
 - Purpose
 - Related documents
 - Cancellation
 - Background
 - Scope
 - Policy statement
 - Action
 - Responsibility

Content of a Policy

Almost every security-related class mentions the necessity of basing procedures on a good security policy. We must understand what is meant by *policy* because there are many conflicting definitions.

What does a policy look like; what kind of content does it have? Because we go into this in much more detail in Step 1 of Evaluating Security Policy, we are brief here. (Do not worry; we provide much more detail later.) A policy typically includes the following content:

- **Purpose:** Explains the reason for the policy.
- **Related documents:** Lists any documents (or other policy) that affect the contents of this policy.
- **Cancellation:** Identifies any existing policy that is cancelled when this policy becomes effective.
- **Background:** Provides amplifying information on the need for the policy.
- **Scope:** States the range of coverage for the policy (to whom or what does the policy apply).
- **Policy statement:** Identifies the actual guiding principles or what is to be done.
- **Action:** Specifies what actions are necessary and when they are to be accomplished.
- **Responsibility:** States who is responsible for what.
- **Ownership:** Identifies who sponsored the policy and from whom it derives its authority; also defines who can change the policy.

Levels of Policy

Recognize that policies can exist on different levels:

- Enterprise-wide/corporate policy
- Division-wide policy
- Local policy
- Issue-specific policy
- Procedures and checklists

Policies at Different Levels

A policy can exist on different levels within an organization. Unless you are at the top of the organizational hierarchy, it is likely that a part of the organization above your level issues a policy that you are expected to implement. A common hierarchy for policy in an organization looks like this:

- **Enterprise-wide or Corporate Policy:** Consists of documents from the highest level (perhaps national or world-wide) within the organization that provide a general direction to be implemented at lower levels in the enterprise.
- **Division-wide Policy:** Typically consists of an amplification of enterprise-wide policy and implementation guidance. This level might apply to a particular region of a national or multinational organization.
- **Local Policy:** Contains information specific to the local organization or corporate element.
- **Issue-Specific Policy:** Contains information related to specific issues—that is, firewall or anti-virus policy.
- **Security Procedures and Checklists:** Consists of local Standard Operating Procedures (SOPs), aligned with and perhaps derived from security policy.

Security policy can exist on some levels and not on others. Documents interact and support one another and generally contain many of the same elements. In a typical organization, policy written to implement higher-level directives might not waive any of the requirements or conditions stipulated at a higher level. Security policy must always be in accordance with local, state, and federal computer crime laws and other applicable government statutes, such as U.S. export regulations.

Now that you understand the policy hierarchy, you can collect policy documents that are available at several levels in the organization. A security policy usually exists (and is enforced to some extent) *even if it is not written down*.

Checkpoint: Procedure Guidance

- **Policies** address the who, what, and why.
- **Procedures** address the how, where, and when.

Policy Worksheet

What do you do when some work does not seem to be covered by an organizational policy? Procedures are derived from policies; if you can characterize the procedures you follow (and you should be able to do that easily), then you can derive the parent policy. This is true even if it has not yet been written and signed. By walking through the who, what, when, where, and why, the parent policy is derived from an understanding of the procedure.

In your organization, what procedures can you list for which you need to document the policy? Make notes on the who, what, when, where, and why of your procedures. You will be able to derive any missing policies based on these notes.

Standard Definitions and Issues

- Organizational
- Specifies uniform use of specific technologies or parameters
- Compulsory
- Usually refers to specific hardware and software

Standards are applied to the organization as a whole. As with policies, these are mandatory. Standards are more specific than the overarching policies. They provide additional definition to the policies and tailor them to specific technologies. Unlike a policy, a standard does not state what is expected of a user from an organizational security stance. Instead, a standard specifies a certain way something should be done or a certain brand or type of equipment that must be used. A simple example of a standard is that all computers purchased must be a certain model and from a certain vendor.

Guideline Definitions and Issues

- Suggestions
- Assists users, systems personnel, and others in effectively securing a system
- Helps ensure that specific security measures are not overlooked
- Applies to security measures that might be implemented in more than one way
- Not compulsory

Guidelines, unlike standards and policies, are not mandatory. Best practices are examples of guidelines that many organizations try to achieve; however, there is not a penalty if guidelines are not met. A guideline is more like a recommendation of the way that something should be done; however, people can choose whether they want to follow it or not. A best practice might start off as a guideline, and if analysis shows that there is a great benefit to following this guideline from either a security or efficiency standpoint, the guideline might become a standard, which would then make the guideline mandatory to follow.

Procedure Definitions and Issues

- Detailed steps to be followed by users, system operations personnel, or others to accomplish a specific task (preparing new user accounts and assigning privileges)
- Mandatory

A procedure is a step-by-step document that is used for operations. Procedures can be daily operations, such as nightly backups or infrequent operations, such as recovering from a disaster. These steps must be clear, complete, and concise. They should also be reviewed on a regular basis to ensure they are accurate and that no changes have occurred.

Baseline Definitions and Issues

- Baselines are similar to standards.
- After a consistent set of baselines has been created:
 - The security architecture can be designed.
 - Standards can be developed.
- If adopted by the organization, baselines are compulsory.

A baseline definition is essentially a more specific implementation of a standard. A baseline definition usually gets into specific technical details of how a system should be configured from either a software or hardware standpoint. After these documents have been thoroughly tested, they become mandatory for someone to enforce. Usually a baseline starts off as a guideline until it has been properly modified to meet the needs of the organization. Hardening rules for setting up a new server is an example of something that starts off as a guideline and quickly turns into a baseline.

Documentation Review

- Policy: All servers must be properly hardened.
- Standard: Administrators must use Windows 2003 as the base operating system.
- Baseline: The specific settings for Windows 2003 should match those in the CIS security template.
- Procedures: The template should be applied when a system is built.
- Guidelines: To ease the application of templates, local GPOs can be used to roll out the changes.

The following are the key documents an organization must have:

- Policy: All servers must be properly hardened.
- Standard: Administrators must use Windows 2003 as the base operating system.
- Baseline: The specific settings for Windows 2003 should match those in the CIS security template.
- Procedures: The template should be applied when a system is built.
- Guidelines: To ease the application of templates, local GPOs can be used to roll out the changes.

Objective of Security Controls

To determine the impact a threat may have on an organization and the likelihood that the threat can occur:

- The process that analyzes the threat scenario and produces a representative value of the estimated potential loss is called risk analysis (RA).

As stated in the beginning of this chapter, security objectives are confidentiality, integrity, and availability. To ensure these goals are met, you must first identify the threats to the systems and the impact those threats might have on your business. Risk analysis is the process by which threats and their impacts are assessed. It is important to remember that threats help you focus in on the weaknesses that exist across an organization, and risk reduction focuses on the specific steps you need to take to reduce those risks across the organization.

Goal of Security Controls

IMPACT
VALUE OF
THREAT

LIKELIHOOD

A MATRIX WITH MORE THAN FOUR SUBDIVISIONS CAN BE USED FOR MORE
DETAILED CATEGORIZATION OF THREATS AND IMPACTS

The goal of security controls is to lower the probability of an adverse event from occurring and to reduce the impact on the business. Lowering the impact enables the organization to continue operations with minimal effect on the essential business functions. Likelihood refers to the probability that an event will actually occur.

Threats Defined

- Activity that represents possible danger
- Comes in different forms and from different sources
- You cannot protect against all threats.
- Protects against the ones that are most likely or most worrisome based on:
 - Business goals
 - Validated data
 - Industry best practice

In security discussions, you will hear a lot about *threats*. Threats—from an information security perspective—are activities that represent possible danger to your information. Danger can be thought of as anything that would negatively affect the confidentiality, integrity, or availability of systems or services. Thus, if risk is the potential for loss or harm, threats can be thought of as the agents of risk.

Threats can come in many different forms and from many different sources. There are physical threats, such as fires, floods, terrorist activities, and random acts of violence. In addition, there are electronic threats, such as hackers, vandals, and viruses. Your particular set of threats will depend heavily on your situation. This includes what business you are in, who your partners and enemies are, how valuable your information is, how it is stored, maintained, and secured, who has access to it, and a host of other factors.

There are too many variables to protect against all the possible threats to information. To do so would cost too much money, take too much time, and require too much effort. You will need to pick and choose what threats you want to protect against. Start by identifying the threats that are most likely to occur or that are most worrisome to your organization. The way to do this is by identifying three primary areas of threat. The first is based on your *business goals*. If your business is heavily dependent on a patented formula, consider theft of that formula to be a likely threat. If your business is the movement of fund transfers over a network, consider attacks on that network link to be a likely threat. These are two examples of business-based threats.

The second category of threats are based on *validated data*. If your web site is repeatedly hacked through your firewall, consider Internet hackers to be a major threat. If your main competitor always manages to find key confidential information about your business plans, consider corporate espionage a threat. These are examples of threats identified because of validated instances of damage based on those threats. In some ways, these are the most serious because they have already happened and are likely to happen again in the future.

The final category of threats are those that are *widely known* in the security industry. To protect against them is good common sense, which is why we put badge readers and guards in buildings, why we use passwords on computer systems, and why we keep secret information locked in a safe. We might not receive attacks against these, but it is commonly understood that it is foolish not to protect against the attacks.

Vulnerabilities Defined

- Weaknesses that allow threats to happen
- Must be coupled with a threat to have an impact
- Can be prevented (if you know about them)

The third element of the risk spectrum is the notion of *vulnerabilities*. (Remember that the first two elements are *risk* and *threats*.) In security terms, a vulnerability is a weakness in your systems or processes that allows a threat to occur. However, simply having a vulnerability by itself is not a bad thing. It is only when the vulnerability is coupled with a threat that the danger starts to set in. Let's look at an example.

Suppose you like to leave the doors and windows to your house unlocked at night. If you live in the middle of the woods far away from others, this might not be a bad thing. People don't wander around your house; if you are high enough on a hill, you can see someone long before they present a danger to you. Thus, the vulnerability of not having locks exists, but it isn't a threat.

Suppose you move to a big city that has a lot of crime. In fact, this city has the highest burglary rate of any city in the country. If you continue the practice of leaving the doors and windows unlocked, the same vulnerability as you had in the country exists. However, in the city, the threat is that much higher. Thus, the overall danger and risk is much greater.

Vulnerabilities can be reduced or even prevented, provided, of course, that you know about them. The problem is that many vulnerabilities lay hidden, undiscovered until somebody finds out about them. Unfortunately, the "somebody" is usually a bad guy. The bad guys always seem to find out about vulnerabilities long before the good guys.

Relating Risk, Threat, and Vulnerability

$$\text{Risk} = \text{Threat} \times \text{Vulnerability}$$

The few slides discussed risks, threats, and vulnerabilities. The three concepts are extremely interrelated. Their relationship can be found in this simple formula:

$$\text{Risk} = \text{Threat} \times \text{Vulnerability}$$

This formula shows that risk is directly related to the level of threat and vulnerability you, your systems, or your networks face. Following is how the formula works:

- If you have a high threat, but a low vulnerability to the threat, the resulting risk is low. In the example we used previously, if you live in a high crime neighborhood (high threat), but you keep your doors and windows locked (so that there is a low vulnerability), the overall risk is low.
- If you have a high vulnerability to a threat (keeping your doors and windows unlocked), but the threat is minor (living in the woods), the risk is low.
- If, however, there is a high level of threat potential (a high-crime area) and the vulnerability to that threat is high (no locks), the risk factor is high.

Of course this formula is good, but recall that there are no absolutes in security. It is typically impossible to assign numeric values to threats and vulnerabilities, so this formula should be used as an aid to guide your thinking rather than as an absolute mathematical calculation. When you get into discussions and arguments about risks, threats, and vulnerabilities (and yes, you will get into arguments about this stuff), refer back to this basic formula to help guide you in the decision-making process.

The Threat Model

- Threat
- Vulnerability
- Compromise

Vulnerabilities are the exposure points (gateways) by which threats are manifested.

On the bottom of this slide, it says that "vulnerabilities are the exposure points (gateways) by which threats are manifested." Therefore, for a threat model to have any meaning at all, there has to be a threat. Are there people with the capability and inclination to attack—and quite possibly harm—your computer systems and networks? What is the probability of that happening? The probability is high that any non-private address will be targeted several times a year. The most common countermeasure for most organizations is to deploy firewalls or other perimeter devices. These work well to reduce the volume of attacks that originate from the Internet, but they do not protect systems from insiders or attacks, such as macro viruses that are able to pass through firewalls about 99 percent of the time.

There are threats, and there are certainly vulnerabilities, and when a threat is able to connect to its specific vulnerability, the result can easily be system compromise. Again, the most common tactic is to protect systems with perimeter devices, such as firewalls. Firewalls are cost-effective, practical, and highly recommended.

The Three Risk Choices

- Accept the risk as is.
- Mitigate or reduce the risk.
 - Eliminate the risk
- Transfer the risk (insurance model).

It is critical to have an understanding of *risk management* to properly choose and deploy intrusion detection and response assets. To manage a risk, one must be able to assess it. In this section of the course, we cover the basic theory of risk assessment. We also talk about three methods of risk assessment: *qualitative*, *quantitative*, and *knowledge-based* (also known as *best practices*).

Whether or not you explicitly choose, you have exactly three options to choose from: *acceptance*, *mitigation*, and *transference*. The following list describes each of these:

- When you accept a risk, you make no changes in policy or process. This decision means that you judge the risk of a given threat to be inconsequential in the greater scheme of things.
- If you feel the threat is significant and can cause harm to the business or enterprise, you have the option of taking action to protect operations by reducing the risk. Firewalls or system patches are obvious examples of risk mitigation.
- Transferring the risk is sometimes a workable technique. The classic example is to buy insurance. This means that you do not have to fully protect yourself against a catastrophic threat. Instead, for a fee, you pass this risk to a risk broker who ensures you up to a limit against the threat. In this case, a risk broker arranges the risk transfer and an underwriter assumes the risk. A real-world example of this is hacker insurance." The insurance company expects you to use a firewall and patches, but ensures you should these fail.

Risk Management

Concept	Derivation Formula
Exposure Factor (EF)	% of asset loss caused by threat
Single Loss Expectancy (SLE)	Asset Value X Exposure Factor (EF)
Annualized Rate of Occurrence (ARO)	Frequency of threat occurrence per year
Annualized Loss Expectancy (ALE)	Single Loss Expectancy (SLE) X Annualized Rate of Occurrence (ARO)

RISK ANALYSIS FORMULAS

The above slide has key formulas that you must remember when analyzing risk.

Risk Requires Uncertainty

If you have reason to believe there is no uncertainty, there is no risk. For example, jumping out of an airplane two miles up without a parachute isn't risky; it is suicide. For such an action, there is a 1.0 probability you will not survive when you hit the ground and almost 0.0 probability you will survive.

Probability ranges between 0.0 and 1.0, although people often express it as a percentage.

You seldom run into cases in which there is a 100 percent certainty when it comes to security. The only thing you can count on is that if you do not have security, you will eventually have a security problem. The only question is when and how severe will it be.

To assess risk requires an understanding of probability or the likelihood that a particular event will occur. Probability is often estimated based on historical data—either of the organization or the industry as a whole.

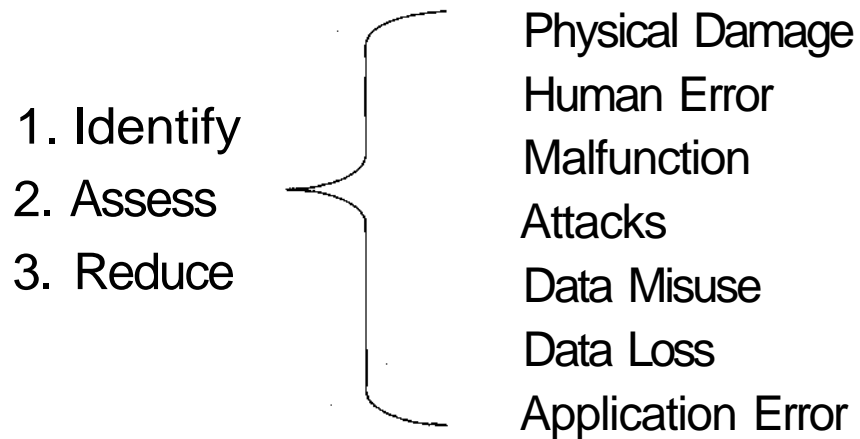
Risk Management Questions

- What could happen (what the threat is)?
- If it happened, how bad could it be (the impact of the threat)?
- How often could it happen (the frequency of the threat, annualized)?
- How reliable are the answers to the previous questions (the recognition of uncertainty)?

Jumping out of an airplane with a parachute involves risk. If you try a James Bond stunt and jump out of an airplane without a chute, you are committing suicide, but you aren't doing anything risky. Risk involves *uncertainty*. Let's relate this to the information assurance world.

If you run a DNS server that has known vulnerabilities and is neither patched nor shielded by the perimeter, it is certainly going to be compromised. It might not happen in a single day, but it will happen over the course of a year. In the same way that gravity is the compelling reason that jumping from a plane without a chute is near-certain death, the continuous probing and poking of exposed systems on the Internet is the compelling reason a box will be compromised. So what? How bad can a compromise be? Well, once an attacker compromises a box, the attacker has the ability to manipulate the addresses associated with the names of the network entities (such as the computers) at the site. These names and addresses are often used to identify which computers are allowed to access other computers—your organization's *trust model*. If you have valuable assets, this might be what happens. The attacker might also create system domains and then hit systems all over the Internet, giving your organization a bad reputation.

Risk Management



Identify the risks. This goes back to knowing the threats to your systems.

Assess the impact. If a threat actually occurs, how badly will your operations be affected? Will you go out of business or will it be a minor annoyance? What is the bottom line in terms of financial, personnel, or asset loss?

Reduce risk by applying mitigation strategies or by transferring the risk, as with insurance. You can also accept the risk, but that does nothing to actually reduce it.

Risk Management Process

- Asset identification
- Threat analysis
- Vulnerability analysis
- Preliminary risk evaluation
- Interim report
- Risk acceptance criteria
- Risk mitigation measures- counter-measures
- ROI analysis
- Final report
- Operation and maintenance

For the intrusion detection risk assessment business case, we follow these steps:

- Asset identification
- Threat analysis
- Vulnerability analysis
- Preliminary risk evaluation
- Interim report
- Risk acceptance criteria
- Risk mitigation measures
- ROI analysis
- Final report
- Operation and maintenance

The process begins with the identification of the threats that organizations face. After these threats are identified, you compare the effectiveness of the threat against the value of the assets it can affect. Then research the known vulnerabilities and evaluate the risks. This evaluation will tell you if there is a significant problem. This information is used as the basis for the interim report to management.

Single Loss Expectancy (SLE) One Shot

- Asset Value x Exposure Factor = SLE
- Exposure Factor: 0 - 100 percent of loss to asset
- Example: Nuclear Bomb / Small Town
(\$90M x 100 percent = \$90M)

How much financial loss are you willing to accept in a single event? It all comes down to money in the end. When considering one shot or *Single Loss Expectancy (SLE)*, consider the value of the information resource asset. For example, a company's top salesman accounts for 25 percent of their \$40 million in revenue, or \$10 million. His client contact list and fee schedule is stored on his laptop and it is not encrypted. If the information fell into the wrong hands, it would be worth at least 10 percent of its value to the competition (\$1 million) and possibly more if the competition could finesse the information. You can calculate a minimum approximate SLE, but there is uncertainty about its maximum value.

Another example is an author who takes a royalty of \$100,000 to write a book. He receives partial payments for each 25 percent he completes of the project. What is the SLE if his hard drive crashes at the 70 percent mark and the data is not recoverable? \$25,000 x 80 percent, or \$20,000 is the SLE, unless the author has been sending chapters in as they are completed.

Annualized Loss Expectancy (ALE) Multi-Hits

- $SLE \times \text{Annualized Rate Occurrence} = \text{Annual Loss Expectancy (ALE)}$
- Annual loss is the frequency with which the threat is expected to occur.
- Example: Web surfing on the job
 - SLE: 1,000 employees, 25 percent waste an hour per week surfing or $\$50/\text{hr} \times 250 \text{ hours} = \$12,500$
 - ALE: Employees do this every week, except when on vacation: $\$12,500 \times 50 = \$625,000$

What happens when the event occurs more than once? You must then calculate the Annualized Loss Expectancy (ALE). The ALE is the annual expected financial loss from a threat. The formula is:

$$\text{Annual Loss Expectancy (ALE)} = \text{Single Loss Expectancy (SLE)} \times \text{Annualized Rate of Occurrence (ARO)}$$

The Annualized Rate of Occurrence (ARO) is the estimated frequency at which a threat is expected to occur. Its value can range from 0 to a large number. Sometimes the ARO is easy to calculate. Other times, it is difficult to compute; in fact, many times, this number represents the uncertainty factor in the risk management calculation.

As a real case scenario, imagine you need to calculate the amount of revenue loss because of your employees' Web surfing during work hours (not work-related, of course). We start by calculating the SLE. For this, we need the asset value and the exposure factor. If 1,000 employees waste 25 percent of their time surfing the Web, and the cost per hour is \$50, then the formula becomes:

$$SLE = \$50/\text{hr} \times 250 \text{ or } \$12,500 \text{ per week}$$

That cost is significant. If you want to calculate the annualized cost, the formula becomes:

$$ALE = \$12,500 \times 52 \text{ weeks or } \$650,000 \text{ per year}$$

ARO and TCO

- **ARO (Annualized Rate of Occurrence)**
 - How often does this problem occur?
 - How does this relate to the overall risk?
- **TCO (Total Cost of Ownership)**
 - What is the total cost of maintaining a security device?
 - Includes installation maintenance and any hidden costs.
- **Safeguard: countermeasure or risk-reducing measure**

ARO is an estimate that should be based on the best available data. What is the probability that your organization will lose Internet access because of a DOS attack? It varies year-to-year and day-to-day. This is one of those times when historical data is used to provide a best guess.

To purchase something, you need to know its cost. *TCO* includes operations and maintenance costs, disposal fees, profits from selling obsolete equipment, lost productivity from unplanned downtime, and so on.

Quantitative Versus Qualitative

- Qualitative is easier to calculate, but its results are more subjective.
- Qualitative is much easier to accomplish.
- Qualitative succeeds at identifying high-risk areas.
- Quantitative is more valuable as a business decision tool because it works in metrics (such as dollars).

The main point between the quantitative and qualitative approach is that qualitative is much easier, and when done well, it can certainly identify the areas that need attention. This is because when an area is marked as high risk, you know you need to look into it.

There is still another approach to risk assessment. This is the *knowledge-based* or *best practices* approach. There is much more upfront work required to implement this, but the results are more accurate and consistent.

Qualitative: Another Risk Assessment Approach

- Banded values: high, medium, low
- Asset value and safeguard cost can be tied to monetary value, but not the rest of the model.
- Very commonly used

For most applications, the best approach is the financial one. The exceptions are critical systems (such as nuclear plant control) and weapon systems. It does take a lot more effort to quantify what the value of things are, and so the qualitative approach is far more popular.

The single biggest problem with the qualitative approach is in the implementation. People tend to mark low risk, even when something is not a low risk. People also mark medium or high for their pet peeves, as opposed to calculating the actual risk.

Best Practice

- No single organization or person is likely to produce best practice.
- Consensus of many organizations and stringent review
- Helps define due care

Is there a resource that can help you define the best secure configuration for your various systems? No single organization or person is likely to produce the best practice recommendation. That is, perhaps an individual has a pretty good idea on how to secure a Windows 2000 system, but is this person the definitive expert on the subject? Are their recommendations the best solution for your specific environment? Even recommendations from a corporation should be looked at closely. A single corporation might not look at the problem (and the solution) from different perspectives. A better approach is to get several organizations to participate in the development of these recommendations. Then, the recommendations include input from diverse industries and from people with diverse experience. There are several samples of consensus-based best practice recommendations. The SANS Research Consensus Projects and the Center for Internet Security are two of the most widely used in the industry.

Threat Vectors

- Outsider attack from network
- Outsider attack from telephone
- Insider attack from local network
- Insider attack from local system
- Attack from malicious code

Take a minute to think about threats that might come from each of these vectors and how they could impact your company. What sort of damage can an attacker from outside the network do? Is there a difference to your business if you get hit from a DOS attack or if a hacker gains control of your web server?

Outsider attacks from telephone include war dialing and social engineering. Are your users trained to know that they cannot give their password out if someone calls up claiming to be from the IT department and asks for the password?

The threat of insider attacks is often ignored; however, think of the damage that could be done if a system administrator finds out he/she is going to be fired and decides to take some revenge? What happens if a normal user discovers he has access to payroll or other confidential files he normally would not see? What damage could the attacker do?

Regarding malicious code, Reuters reported in January, 2004 that computer virus attacks cost global businesses an estimated \$55 billion in damages in 2003, and that number is expected to increase in 2004. Have you updated your antivirus software lately? Don't forget laptops and remote users.

Controlling Your Environment

- Policy: Tells a user what to do
- Training: Provides the skill set
- Awareness: Changes user behavior
- Key threat: Social engineering
 - Manipulation
 - People need to be made "aware" of the dangers.

The following are key principles to remember:

- Policy: Tells a user what to do
- Training: Provides the skill set
- Awareness: Changes user behavior

A key threat: Social engineering is a form of manipulation; employees need to be made "aware" of the dangers.

Security Awareness

- General, collective awareness of an organization's personnel and the importance of security and security controls
- Benefits
 - Reduces unauthorized actions attempted by personnel
 - Increases the effectiveness of protection controls
 - Helps to avoid fraud, waste, and abuse of computing resources

A major security issue is continually justifying the existence of security programs when there has been a lack of incidents. Security officers must actively engage management to ensure that security remains on their radar and continues to gain support. Implementing security controls one year only to fail to update them in future years is a common occurrence, which is why security officers must actively promote security throughout the organization.

Approaches to Security Awareness

- Live, interactive presentations
- Publishing distribution
- Incentives
- Reminders

The items listed in the previous slide promote security. Training and testing must continue on a regular basis to be effective. Reminders can include e-mails, posters, flyers, or formal and informal training. People forget if they are not reminded, so keep security in the forefront and don't let it get pushed to the back burner.

Security Training

- Uses a specific classroom or one-on-one training
- Elements of training
- Training for specific groups or departments
- Advanced INFOSEC training for security practitioners and auditors
- Training for senior managers, functional managers, and business unit managers

As the security professional for your organization, your goal as the great translator might or might not extend to training the staff. Informally, you will deliver the information. You might want to utilize the training department for the more formal activities. For large projects, external organizations achieve a higher rate satisfaction due to the "consultant factor." (When an internal person tells you to do it, you tend to ignore it, whereas when a consultant tells you, you tend to listen.)

Some ways to accomplish this is to use specific classroom or one-on-one training. Elements of training include:

- Implementation
- Targets an audience
- Motivates management and employees
- Maintains program
- Evaluates program
- Training for specific groups or departments

Outsourcing

- Allowing someone else to perform services for your organization
- Typically used in monitoring and coding
- Off shoring is when outsourcing is done over seas

Typically people are your most valued resources. Many organizations have people, but they do not have the head count to accomplish the tasks at hand. Therefore, outsourcing is where manually intensive tasks are given to third parties to perform.

Service Level Agreements

- Service level agreements (SLAs) are an agreement between two parties on how a service will be performed.
- Contractual arrangement
- Provides restitution if one side does not perform.

An SLA or a service level agreement is a contractual arrangement between two parties on how a service is to be delivered.

Summary

- Concepts
- Data classification
- Policies, procedures, standards, baselines, and guidelines
- Security controls
- Risk, threat, and vulnerability
- Risk management
- Risk assessment

This Security Management Practices domain discussed the objectives and concepts of security, especially the CIA triad. We also addressed authentication and identification including the difference between the two. Data classification schemes can vary from organization to organization, and labels used for classification can also vary depending on the nature of the business.

We discussed high-level policies management, which should draft mandatory standards and baselines. We also detailed procedures and suggested guidelines. Additionally, we examined the objective of security controls and how risk management and risk assessment can help meet those objectives by identifying threats and determining the impact of those threats.