Defense in Depth: Foundations for Secure and Resilient IT Enterprises

Christopher J. May
Josh Hammerstein
Jeff Mattson
Kristopher Rush

September 2006

CMU/SEI-2006-HB-003

Carnegie Mellon
Software Engineering Institute

Pittsburgh, PA 15213-3890

Defense-in-Depth: Foundations for Secure and Resilient IT Enterprises

CMU/SEI-2006-HB-003

Christopher J. May
Josh Hammerstein
Jeff Mattson
Kristopher Rush

September 2006

CERT Program

This report was prepared for the

SEI Administrative Agent
ESC/XPK
5 Eglin Street
Hanscom AFB, MA 01731-2100

The ideas and findings in this report should not be construed as an official DoD position. It is published in the interest of scientific and technical information exchange.

FOR THE COMMANDER

Christos Scondras
Chief of Programs, XPK

This work is sponsored by the U.S. Department of Defense. The Software Engineering Institute is a federally funded research and development center sponsored by the U.S. Department of Defense.

This work was created in the performance of Federal Government Contract Number FA8721-05-C-0003 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center. The Government of the United States has a royalty-free government-purpose license to use, duplicate, or disclose the work, in whole or in part and in any manner, and to have or permit others to do so, for government purposes pursuant to the copyright license under the clause at 252.227-7013.

For information about purchasing paper copies of SEI reports, please visit the publications portion of our Web site (http://www.sei.cmu.edu/publications/pubweb.html)

Table of Contents

List of Figures

Acknowledgements

The authors wish to acknowledge the significant talents and contributions of the following individuals:

Julia Allen

David Biber

Claire Dixon

Stephanie Losi

Robin Ruefle

Pamela Williams

Julia Allen reviewed the document while in draft and provided her unmatched insight into enterprise security governance issues. Her numerous recommendations and thoughtful critique helped shape the final document and ensured we considered all relevant areas.

David Biber added all of the unique illustrations and graphics to the document. His talent and originality add clarity and understanding when words are incapable of doing so alone. He also is gifted at injecting humor and subtle analogies that greatly contribute to the document's knowledge transition capabilities.

Claire Dixon painstakingly completed the initial edit of the document and proved skillful in synthesizing the styles of the four authors.

Stephanie Losi was instrumental in bringing this document into its final state. Her amazing blend of technical knowledge and writing talent helped smooth out some of the lingering jagged edges while also incorporating many of her own thoughts on enterprise security.

Robin Ruefle used her vast knowledge of computer security incident handling to greatly improve the Incident Management module. She graciously reviewed the rest of the document and provided many helpful recommendations.

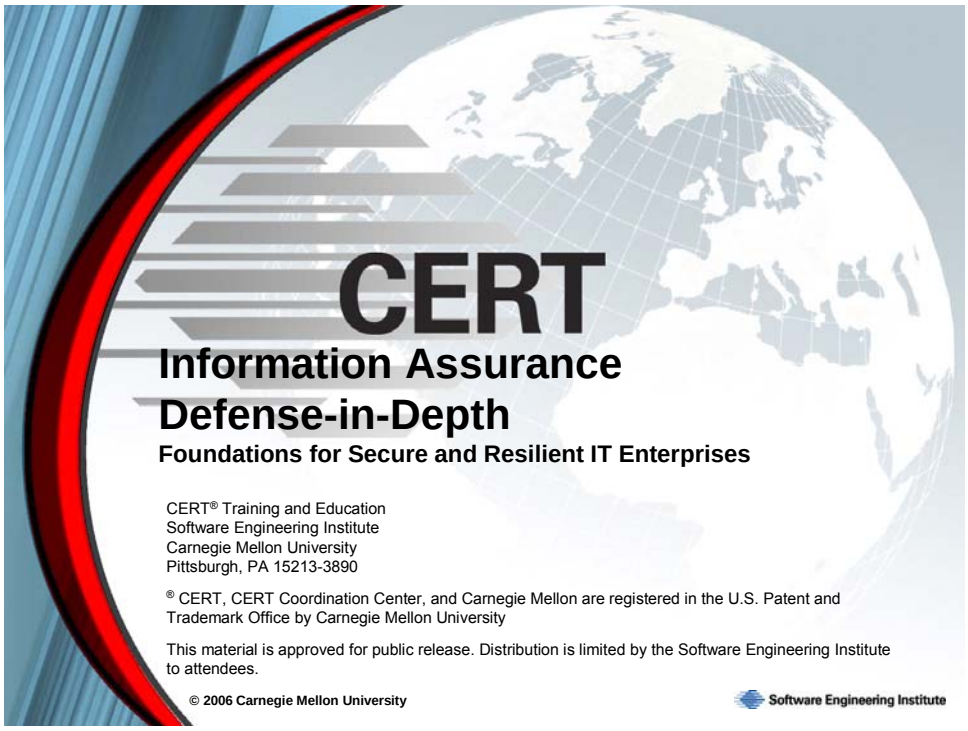
Pamela Williams deftly managed the document's development lifecycle, structure, and formatting. Her in-depth knowledge of CERT training materials and publications ensured the technical language and definitions were universally consistent.

Abstract

The Defense-in-Depth Foundational Curriculum is designed for students, ranging from system administrators to CIOs, who have some technical understanding of information systems and want to delve into how technical assurance issues affect their entire organizations. The course material takes a big-picture view while also reinforcing concepts presented with some details about implementation. Therefore, this course can be a useful pursuit for system administrators and IT security personnel who would like to step up to the management level. It also can provide a refresher for IT managers and executives who want to stay up to date on the latest technological threats facing their enterprises.

The curriculum consists of eight main modules: (1) Compliance Management, (2) Risk Management, (3) Identity Management, (4) Authorization Management, (5) Accountability Management, (6) Availability Management, (7) Configuration Management, and (8) Incident Management. The document also contains an introduction, “Foundations of Information Assurance,” which focuses on how the overarching concepts of confidentiality, integrity, and availability can lead to a comprehensive security strategy.

Module 1: Foundations of Information Assurance



This module introduces the central tenets of Information Assurance (IA) and defines and describes the construct of IA Defense-in-Depth.

Instructional Objectives

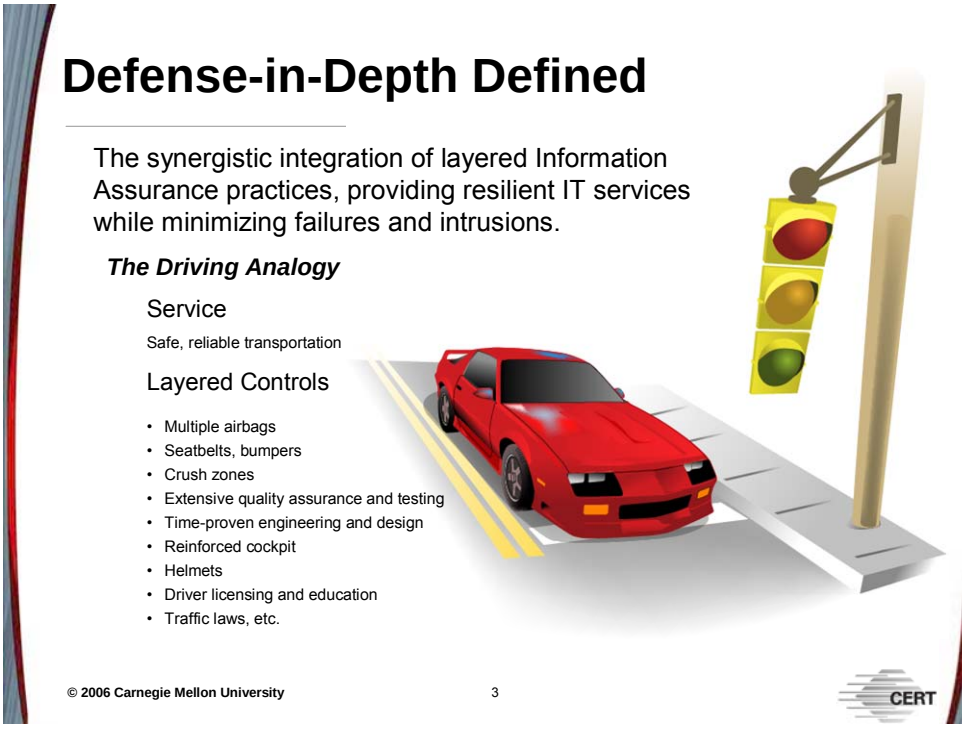
Upon completion of this module, students will be able to

- Define the information assurance concept of Defense-in-Depth
- Define the CIA Triad and related IA terms
- Identify each of the eight high-level components of Defense-in-Depth
- Describe three strategies for achieving Defense-in-Depth

CERT

This instructional module will enable students to complete all of the above learning objectives.

1 Overview of Defense-in-Depth



Defense-in-Depth is an IA construct in which multiple, related, organizational actions and controls are applied to minimize failures and intrusions and their propagation. In essence, it is a multi-pronged protection strategy. When Defense-in-Depth is achieved, reliability and resilience—the ability of IT systems to withstand attacks with minimal impact on services—are also achieved.

Defense-in-Depth can be broken down into component containers—conceptually defined areas that each focus on a particular aspect of the big picture, such as identity management or availability management. These component containers allow IT professionals to more easily understand the larger requirements and thereby identify appropriate actions and controls in the context of their own organizational environment.

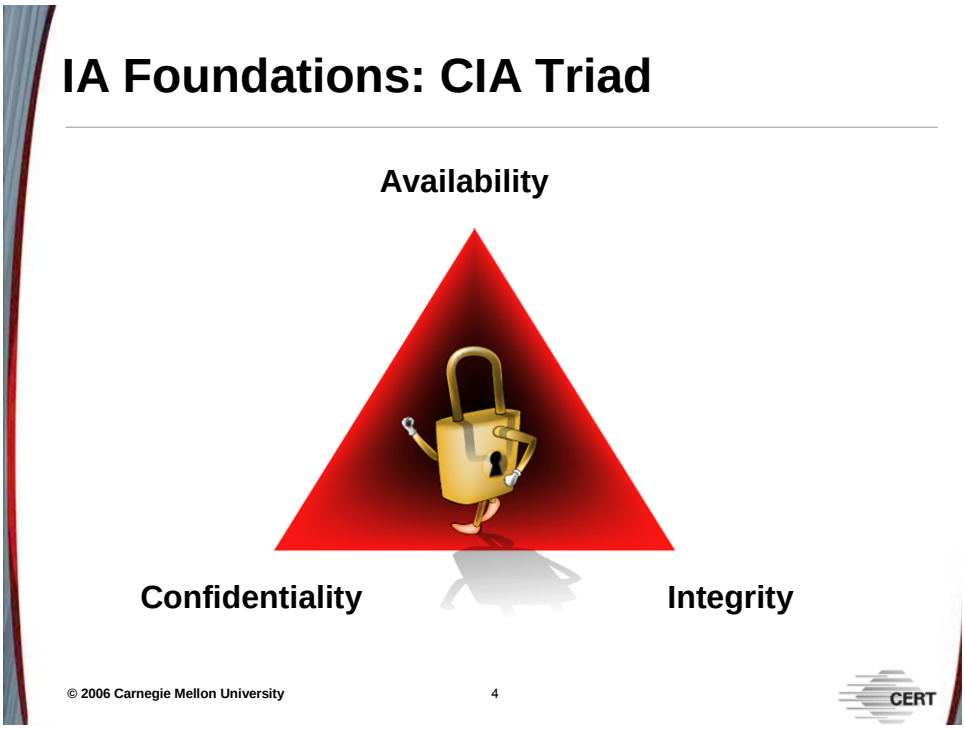
The simple analogy in the slide illustrates how numerous, complementary, protective and responsive controls combine to provide the intended service. If any one of these were removed, it might significantly impact the capability for providing this service. For example, if there were no speed limits on highways and people regularly drove above 100 miles per hour during rush hour, it is unlikely that all of the other controls would prevent injuries in the event of an accident.

Other important definitions follow [ATIS 00]:[1]

Information systems security (INFOSEC and/or ISS) – [The] protection of information systems against unauthorized access to or modification of information, whether in storage, processing, or transit, and against the denial of service to authorized users, including measures necessary to detect, document, and counter threats.

Information assurance – Information operations (IO) that protect and defend information and information systems (IS) by ensuring availability, integrity, authentication, confidentiality, and nonrepudiation. This includes providing for restoration of information systems by incorporating protection, detection, and reaction capabilities.

[1] See http://www.atis.org/tg2k/.

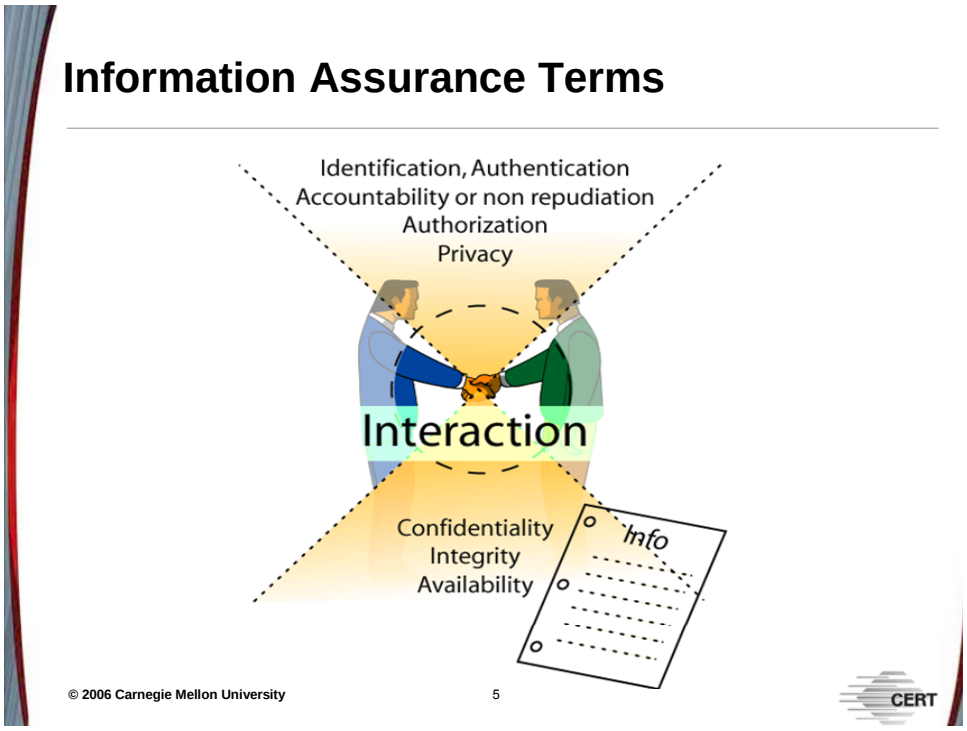


The goal of information security is to sustain and defend three critical security properties of information: *confidentiality*, *integrity*, and *availability*.

Confidentiality refers to assurance that information can be read and interpreted only by persons and processes explicitly authorized to do so. Protecting confidentiality involves implementing procedures and measures to prevent malicious and accidental disclosure of information to unauthorized readers. Information that could be considered confidential is commonly called *sensitive* information. An example would be email content that is protected from being read by anyone other than the intended addressees.

Integrity is the assurance that information remains intact, correct, and authentic. Protecting integrity involves preventing and detecting unauthorized creation, modification, or destruction of information. An example would be implementation of measures to verify that email content was not modified in transit.

Availability refers to assurance that authorized users can access and work with information assets, resources, and systems when needed, with sufficient response and performance. Protecting availability involves measures to sustain accessibility to information in spite of possible interference, including system failures and deliberate attempts to obstruct availability. An example would be protection of access to and throughput of email service.



How is information secured?

The three core properties of Confidentiality, Integrity, and Availability (CIA) serve as the foundation for information security. As information is shared, however, we must concern ourselves with another layer of properties:

- **Identification** – refers to the unique properties of users that separate them from others and the means by which these users claim their identities on a system. Usernames are common means of identification. Identification is tightly linked with authentication.
- **Authentication** – is the process of proving that you are who you say you are—establishing proof of identity. It can be achieved through passwords, smart cards, biometrics, etc.
- **Accountability** – is a system's ability to determine the actions and behavior of a single individual within a system, and to identify that particular individual. It is what binds these actions to users. Audit trails and logs are used for this. This is very tightly linked with nonrepudiation.
- **Nonrepudiation** – is the mechanism that keeps individuals from denying their actions. For example, if a customer places an order and a nonrepudiation security service is not built into the system, the customer could deny ever making that purchase. Nonrepudiation services provide a means of proving that a transaction occurred, whether the transaction consists of an online purchase or an email message that was sent or received. Digital signatures can be used to establish nonrepudiation.

- **Authorization** – refers to the rights and permissions granted to an individual (or process) that enable access to a computer resource. Once a user is authenticated, authorization levels determine the extent of the system rights available to that user.
- **Privacy** – is the level of confidentiality and privacy protection that a user (or process) is given in a system. This is often an important component of security controls. Privacy not only guarantees the fundamental tenet of confidentiality of an organization's data, but also guarantees the data's level of privacy, which is being used by the operator. [Krutz 01].

If any of these higher-layer properties are compromised, you lose CIA as a whole. The key to mitigating this risk is to securely manage the interactions. This can be accomplished through various means, including, but not limited to

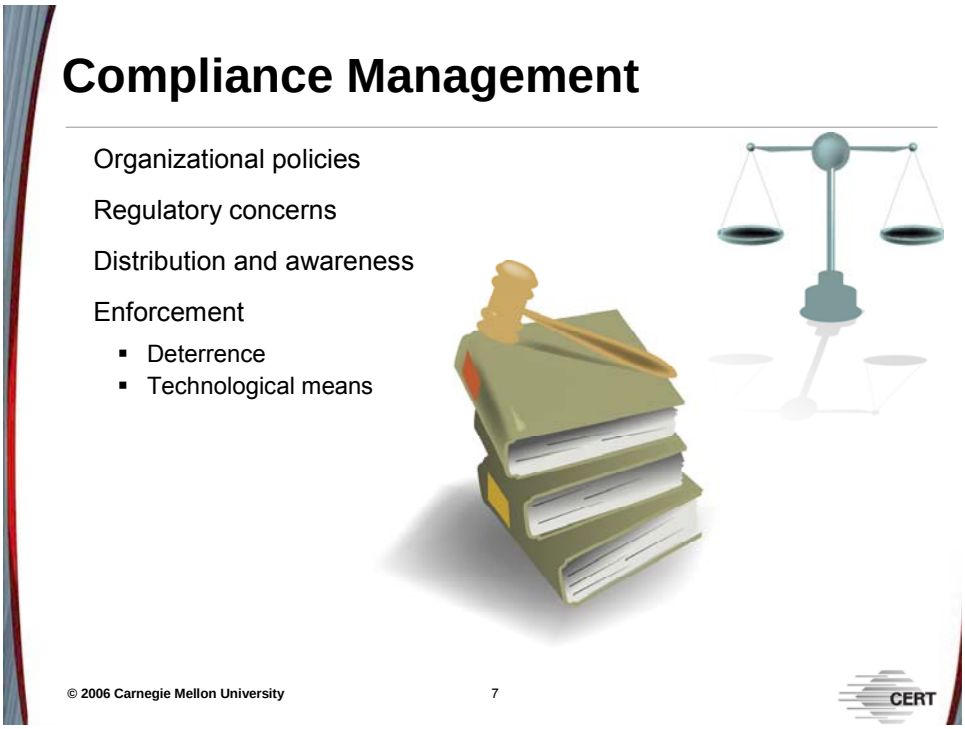
- strong authentication mechanisms (e.g., Kerberos, Radius)
- data encryption (e.g., IPSEC, Encrypting File System, PGP)
- secure/thorough administrative practices (e.g., access controls, permissions/rights, integrity checking systems)
- secure architectural design (e.g., limiting unnecessary services, security perimeters)

These implementations will be covered in detail later in this course.

Defense-in-Depth Components

CERT

This is an overview slide. Each of these eight components will be introduced in this module and then covered in much greater detail in following instructional modules.



1.1 Compliance Management

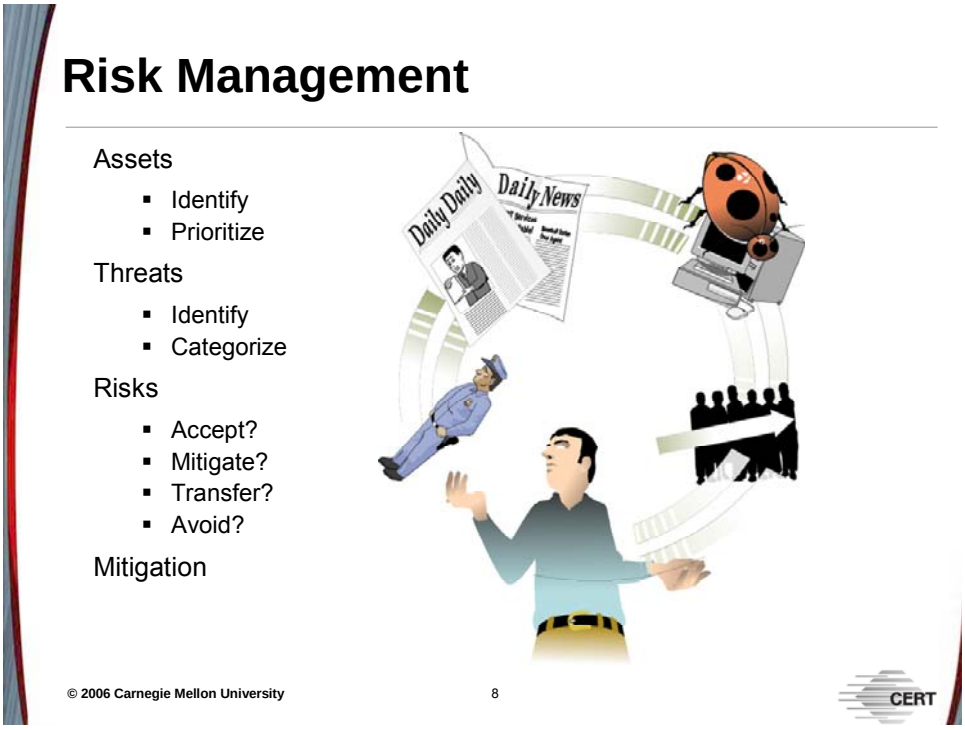
Compliance management, along with risk management, is one of the most important components of Defense-in-Depth. It serves as the bedrock on which many other components build. Assessing the state of an organization's compliance management can be the most telling and illustrative indicator of its overall IA posture and level of IA maturity.

Organizational IA policy provides guidance to users and administrators of information technology services. These policies should be written clearly enough to be easily understood by non-technical users and generally enough that they don't have to be changed as the underlying technology changes. Policies should cover not only day-to-day conduct, but also what to do and who to notify if a breach or attack occurs.

Implementing effective methods for advertising and distributing policy to users is critical for awareness and compliance, as is ensuring users have read and agreed to follow policies. Where possible, senior management should advocate and sign off on IA policies. An example of such a policy is Carnegie Mellon University's computing policy, which can be found at http://www.cmu.edu/policies/documents/Computing.htm.

Numerous enacted laws at all levels have become prevalent in the area of Information Security and Privacy. Examples in the United States are the Health Insurance Portability and Accountability Act of 1996, the Gramm-Leach-Bliley Act of 1999, and the Sarbanes-Oxley Act of 2002. All of these laws enact compliance standards to which affected organizations must adhere.

Policies should also define consequences for individuals who fail to abide by the rules. Such consequences can serve as deterrents and, in some cases, help protect users from themselves. Where feasible, technology should be used to enforce the policies. An example of this would be access controls that help maintain the confidentiality of private organizational information (e.g., human resources or finance data).



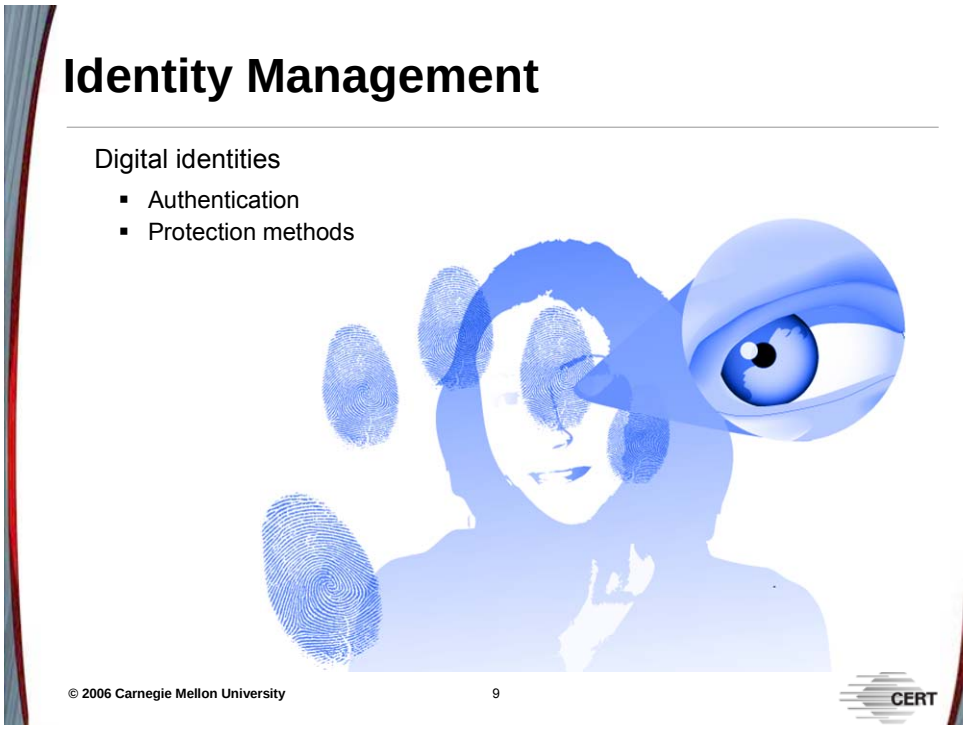
1.2 Risk Management

Any discussion of risk requires a discussion of assets. An *asset* is anything of value to an organization. A *critical asset* is an asset that is vital to an organization's operations, reputation, or future growth.

Risk management is the process of identifying risks to assets and deciding how (or if) to manage those risks. Broken down into its component steps, risk management involves

- The identification and prioritization of assets
- The identification and categorization of threats (risks) to those assets
- The prioritization of risks – which risks are high-priority versus low-priority? Will a risk be accepted, mitigated, transferred, or avoided?
- The mitigation of risks through security and other measures

Remember that security is only one approach to mitigating risk. There are also many others, such as insurance and contracts. Security is part of operational risk management, not the whole solution.



1.3 Identity Management

Digital identities enable computers to distinguish users from one another and allow computers to provide granular degrees of service and access to resources. Digital identities are implemented in various ways, such as biometrics, one-time passwords, digital certificates, and smart cards; however, the standard username/password combination is a ubiquitous example.

Computer systems must have some means of validating the authenticity of presented digital identities; the IA concept of authentication attempts to solve this problem. For example, when you attempt to access a Web-based email service like Yahoo! Mail, you are presented with a standard login page. If you already have an account, you simply type in your username and password. If these credentials match Yahoo!'s stored credentials for your account, you are authenticated and your Inbox will be displayed.

Personal digital identities have become critical to most people's lives—even if they don't realize it. Many people use the same username and password for all of their online shopping, banking, and other everyday Internet services. If this digital identity is stolen by a crafty, malicious individual, the victim could suffer devastating financial loss as well as the emotional distress of losing his or her own identity.

Some basic practices can help prevent theft of your digital identity. Here are a few examples:

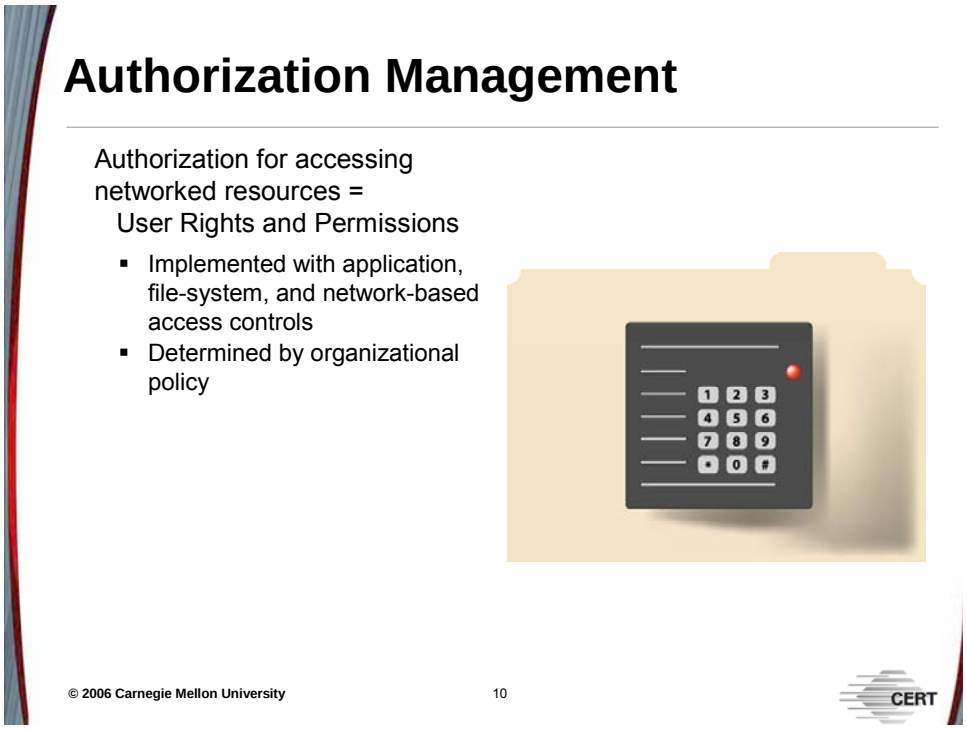
- Use different usernames and passwords for each critical online service account, and make sure your passwords are not easily guessed or from a dictionary. A good practice is to use the concept of a passphrase versus a password. Instead of a standard password (a pet's

name like “tiger”), try stringing together words in an easy-to-remember phrase like “MyCat’sNameIsTiger!” Capitalizing the first letter of each word and including standard punctuation marks help make your password resistant to cracking attacks.

- Do not share the credentials of your digital identity with anyone, including help desk personnel, email requests from supposed friends, etc. Be wary of such requests, and remember that no legitimate IT service personnel will ever ask you directly for your password.
- Do not write down your credentials or store them on your computer in a file that is not secured. If you must record them on your computer, use a free password storage program that will securely encrypt your passwords (i.e., Password Safe http://passwordsafe.sourceforge.net/).

Like individuals, organizations also must take care to safeguard their identity, especially as the incidence of phishing attacks increases. Identity management may involve registering all forms of the company domain name (.com, .net, .biz, etc.) and keeping close tabs on similar URLs and unauthorized use of brand names, logos, and watermarks.

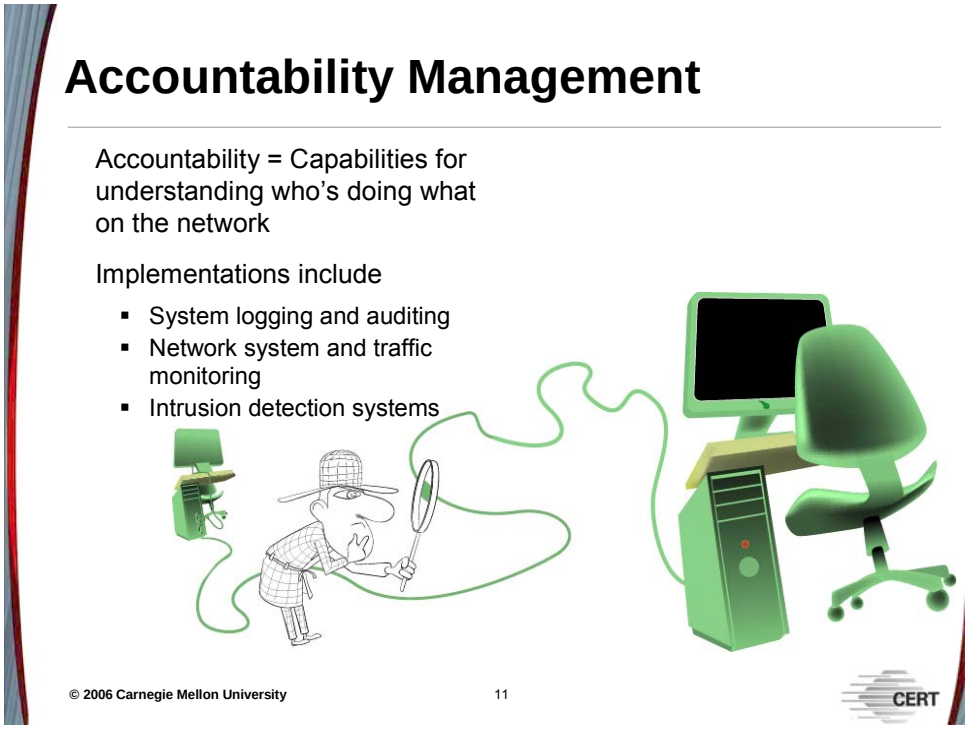
Organizations also can communicate proactively with customers so that customers understand how the organization will and will not contact them. This can help customers spot and avoid fraud attempts.



1.4 Authorization Management

Authorization management deals with user rights and permissions. It basically answers the questions "Who can do what on a computer system or network?" and "When and where can they do it?" Organizational policy is what decides all of this.

For example, Joe works in Human Resources for a small company. Organizational policy says that HR data must be kept confidential and should only be accessible to HR staff and management. Joe keeps private personnel information like employee salary, benefits, and so forth in a password-protected spreadsheet on his computer that is also backed up in his private home directory on the company's file server. While this doesn't follow industrial-strength and recommended security practices, it is in keeping with company policy. A more explicit policy that stipulates more granular and specific controls would likely provide better privacy protection capabilities. For example, such a policy might detail specifically what constitutes sensitive information, how to categorize it, and how to protect various types of information. Authorization management amounts to all of the technological and organizational controls that enable the enforcement of policy.

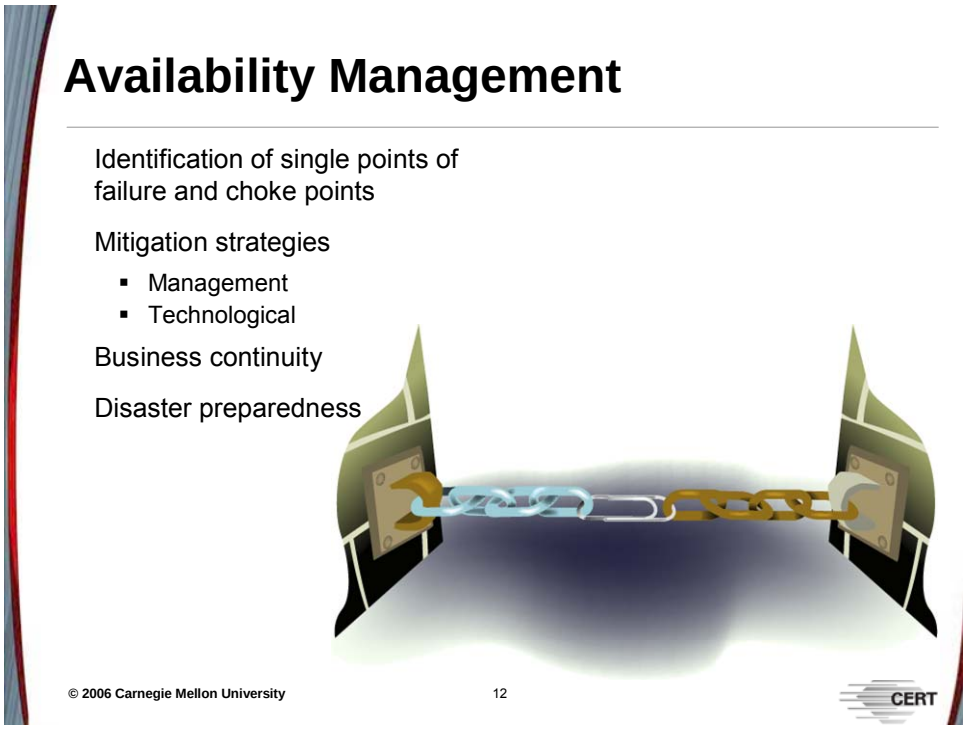


1.5 Accountability Management

Accountability management allows IT staff to know what's happening on their computer systems and networks. It's implemented in many different ways but most commonly by

- configuring systems to log interesting system activities, such as user login attempts
- inspecting network utilization to detect types of traffic traversing the network and their volume
- automated monitoring of systems for service outages
- implementing intrusion detection systems to alert administrators to suspicious activity on computer systems and networks

Even when the IT staff implements these kinds of technologies, it takes a significant amount of human time and resources to review the collected data. Providing just the right amount of accountability management capability often becomes a delicate balancing act.



1.6 Availability Management

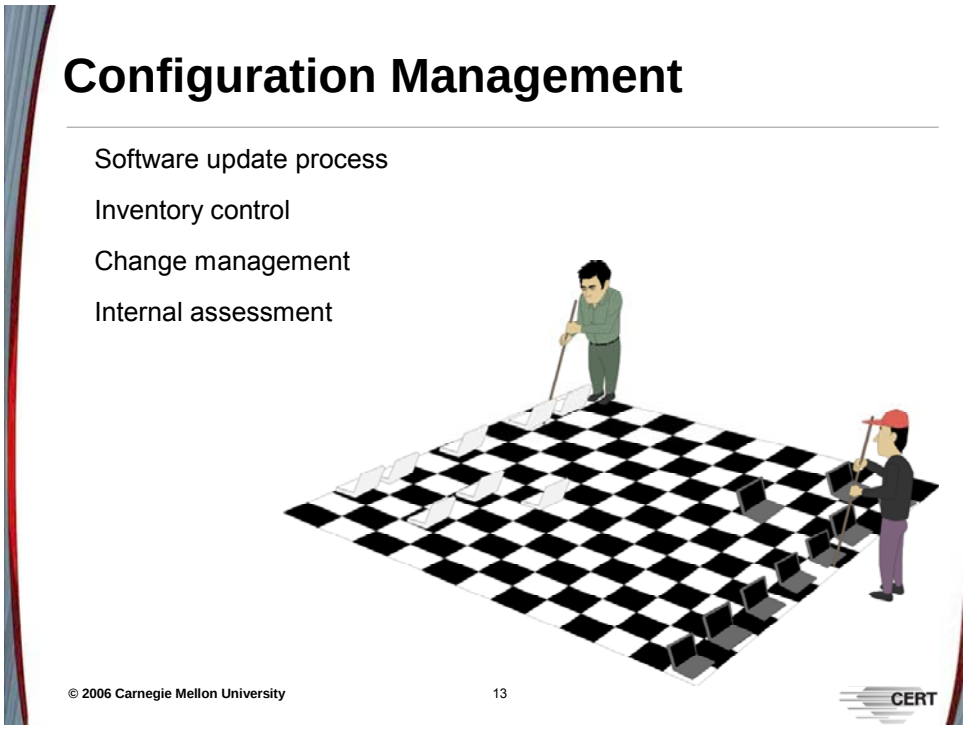
Of the three parts of the CIA triad, none is more important to end users than availability. Users often assume IT is handling the confidentiality and integrity issues, but service availability doesn't lend itself to assumptions—it's either there or it isn't. If it isn't, business continuity is at risk. We'll discuss the critical link between IT continuity and business continuity in more detail later in this course.

For now, suffice to say that system administrators and network managers should attempt to identify potential single points of failure (SPOF) within their network topologies. Critical network infrastructure devices like Ethernet switches and routers, as well as core servers like Domain Name System (DNS) servers, should be studied for their potential for failure and for how their outage would affect the organization. This goes back to risk management—determining to what degree risk must be mitigated to achieve uptime and availability requirements. Redundant or fault-tolerant systems can help mitigate the risk of potential service/network outages.

The same principle applies to key IT staff. If a company has only one trained and qualified Cisco router/switch administrator, then it is risking productivity by allowing a SPOF to exist in personnel. Cross-training and job rotation are mitigation strategies for this problem.

In today's world it is wise for companies to plan for disasters. For example, it's feasible that a company might cease to exist as a result of a major fire in its datacenter. To prevent this, aggressive mitigation strategies should be put into place. Shared sites or mutual-aid agreements with partner companies can help in solving these issues. Practicing business

continuity strategies and planned disaster response is just as important as establishing such a plan, and this should be factored into a company's disaster readiness preparations.



1.7 Configuration Management

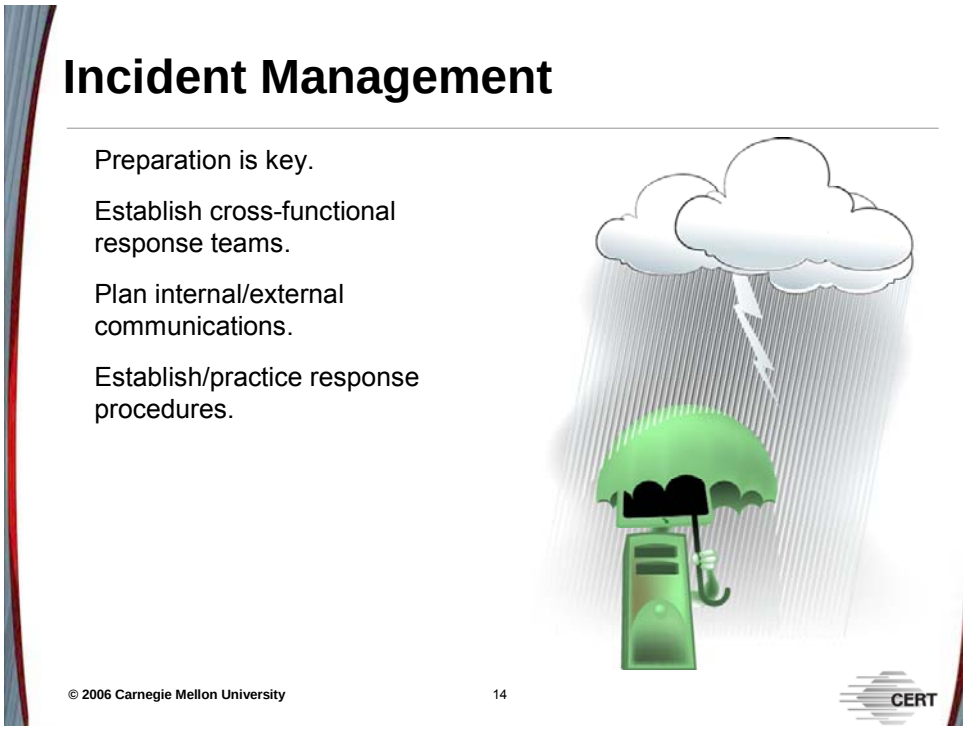
Configuration management amounts to the proactive day-to-day techniques for ensuring that the IT mission is running smoothly and that its present state is well understood by staff members. Specifically, configuration management consists of the software update process, inventory control, change management, and internal, ongoing assessment.

Software update management is one of those day-to-day maintenance activities that on the surface seems relatively easy; however, it can be very complicated in even moderately sized IT departments. Ideally, updates should be available locally (for fast deployment) and then be tested and rolled out to applicable systems the same day that a patch is released. This is difficult to achieve even in small networks; however, vendors like Microsoft and Apple and open-source systems like Linux are dedicating greater resources to this problem and are making it easier with new and more mature technologies. For example, Microsoft recently released its free Windows Server Update Services (WSUS) software that allows system administrators to manage granular patch deployment on Windows networks [Microsoft 06]. You can read more about WSUS at http://www.microsoft.com/windowsserversystem /updateservices/default.mspx. On the other hand, network infrastructure devices like routers and switches are more difficult to update; typically, an entirely new and updated operating system image must be loaded to replace the vulnerable one.

Effective management of the IT system inventory is itself a major endeavor. Many organizations use barcodes and infrared scanners to ease this burden. Inventory life-cycle management should also be documented as part of organizational policy. All configuration

changes to core servers and systems should be carefully documented and recorded; this may help in future troubleshooting and maintenance.

Internal assessment of the security state of the network is an important and often ignored best practice. Tools such as Microsoft's Baseline Security Analyzer, as well as open-source tools such as Nessus, can help administrators with these activities. Some of these tools will be discussed in more detail later.



1.8 Incident Management

Security related events will happen, period. Even the best information security infrastructure cannot guarantee that intrusions or other malicious acts will not happen. When computer security incidents do occur, it is critical for an organization to have an effective means of responding. The speed with which an organization can recognize, analyze, and respond to an incident can limit the damage done and lower the cost of recovery.

Organizations require a multilayered approach to secure and protect their critical assets and infrastructures. This multilayered strategy requires that not only technical issues, but also organizational and procedural approaches be in place to manage computer security incidents. The goal is for enterprises to stay resilient in the face of risks, attacks, and other threats to business continuity.

Incident management is a vital part of that, and it starts with planning. Plans should document the set of steps to be taken in the face of various threats and attacks and should be established for events like computer viruses, denial of service, unauthorized information disclosure or data theft, equipment failures, and dependent service (air conditioning, water-supply, power, etc.) outages. Plans should be well known to all organizational employees and end-users and should be tested periodically to ensure they are effective.

Additionally, incident response capabilities should be developed by identifying and training specific personnel to handle various response actions. Organizations can approach this step formally or informally. In a more formalized structure, a standing computer security incident response team (CSIRT) can be established whose main job is to detect and react to threats

and risks. An informal structure, on the other hand, might consist of a team whose members have other job duties but are called together when an incident occurs. In either case, the team will work to coordinate the analysis and resolution of computer security incidents.

For example, if a company's network is penetrated and compromised by a "hacker," a response team could

- investigate the incident by utilizing effective forensic collection and analysis methods, if team members have been trained in forensics
- determine the scope and impact of the malicious activity, including what was done and which assets and data were damaged or compromised
- identify response or mitigation strategies and coordinate their implementation to contain or eradicate the intrusion and to recover the affected systems
- implement communication plans for dealing with the public, law enforcement, and the media
- implement pre-planned user awareness methods to help prevent further compromise and damage
- document lessons learned and implement new plans for remediation of weaknesses discovered

Incident management is more than just an IT function. It is an organization's responsibility to holistically prepare for security events. Doing so will increase the organization's resiliency.

The takeaway message here is that incident management is not just the application of technology to resolve computer security events. It is the development of a plan of action and a set of processes that are consistent, repeatable, of high quality, measurable, and understood within the constituency. To be successful, this plan should

- integrate into the existing processes and organizational structures so that it enables rather than hinders critical business functions
- strengthen and improve the capability of the constituency to effectively manage security events, thereby keeping intact the availability, integrity, and confidentiality of an organization's systems and critical assets
- support, complement, and link to any existing business continuity or disaster recovery plans, where and when appropriate
- support, complement, and provide input into existing business and IT policies that affect the security of an organization's infrastructure
- implement a command and control structure, clearly defining responsibilities and accountability for decisions and actions
- be part of an overall strategy to protect and secure critical business functions and assets
- include the establishment of processes for
    - notification and communication

- analysis and response
- collaboration and coordination
- maintenance and tracking of records

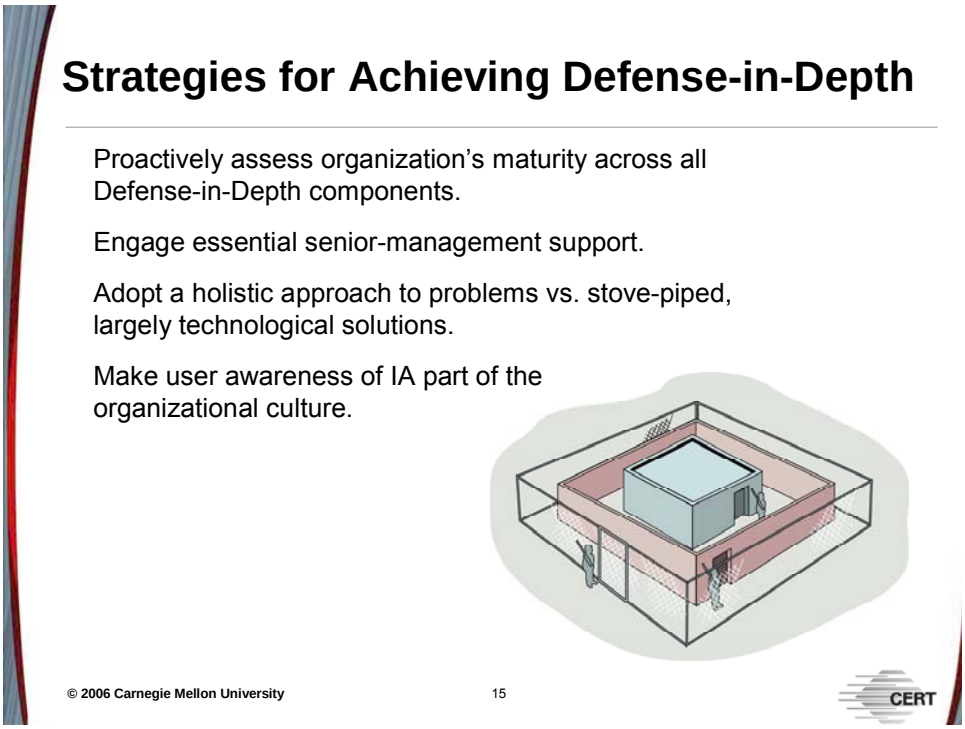
Security-related IT events will happen, period. Systems will crash and equipment will fail. Incident management assumes this and its principle response to this is preparation. Response plans should be established for common events like computer viruses, equipment failures, and dependent service (air conditioning, water-supply, power, etc.) outages. Again, practicing these planned responses is important to their success under fire.

Establishing response teams within an organization is often beneficial. These are groups of identified individuals from across the organization who train in the area of incident management.

For example, in the event a company experiences a penetration and compromise of its network, resources from a "hacker" response team could

- implement communication plans for dealing with the public, law-enforcement, media, and so forth
- implement pre-planned user-awareness methods to help prevent further compromise and damage
- investigate the incident by utilizing effective forensic collection and analysis methods
- document lessons learned and implement new plans for remediation of weaknesses discovered

Incident management is more than just an IT staff function. It is an organization's responsibility to prepare holistically for security related events; this will further increase the IT department's resiliency.



1.9 Strategies for Achieving Defense-in-Depth

Our big-picture focus means that Defense-in-Depth and its eight component areas should be viewed as they relate to an organization's IT assets. Personnel should continually assess and improve the posture of all aspects of these eight areas.

Senior managers should be made aware of the benefit of promoting a systematic, Defense-in-Depth approach to managing the organization's IT resources and services. The goal is for no aspect of security to be left unexamined. Armed with a systematic framework for thinking about security, managers can feel confident they have not overlooked some vital component or perspective. Ideally, they should vocally support and champion this effort, articulating its importance to the mission of the organization to all members.

Traditionally, information and computer security has been a technology- and tool-centric field. The Defense-in-Depth method of dealing with security-related concerns takes a more holistic approach. It includes management techniques, cross-functional planning, and active user intervention, in addition to technological solutions to improve an organization's security posture.

End users of IT services are often the weakest link when it comes to enforcing security standards. Continual efforts at making them aware of security and information assurance best practices will pay off significantly and will save costs and resources as well.

Review Questions

1. Define IA Defense-in-Depth.
2. Define *integrity*, as it relates to IA.
3. Define *authentication*, as it relates to IA.
4. List the eight components of IA Defense-in-Depth.
5. Name two strategies for achieving IA Defense-in-Depth.

Module 2: Compliance Management

This module discusses the role of organizational policy in compliance management, policy development, and law and regulation in information security.

Instructional Objectives

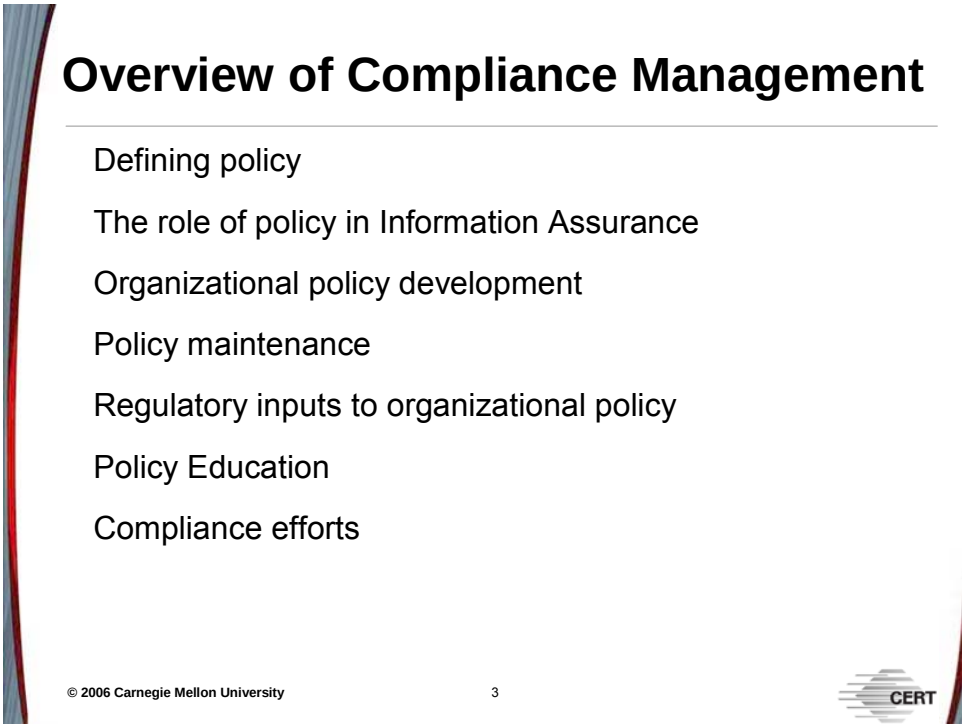
Upon completion of this module, students will be able to

- Explain the role of IT policy
- Define compliance
- List the characteristics of a good policy
- Identify the key security components of IA regulation
- Explore internal issues of policy compliance

CERT

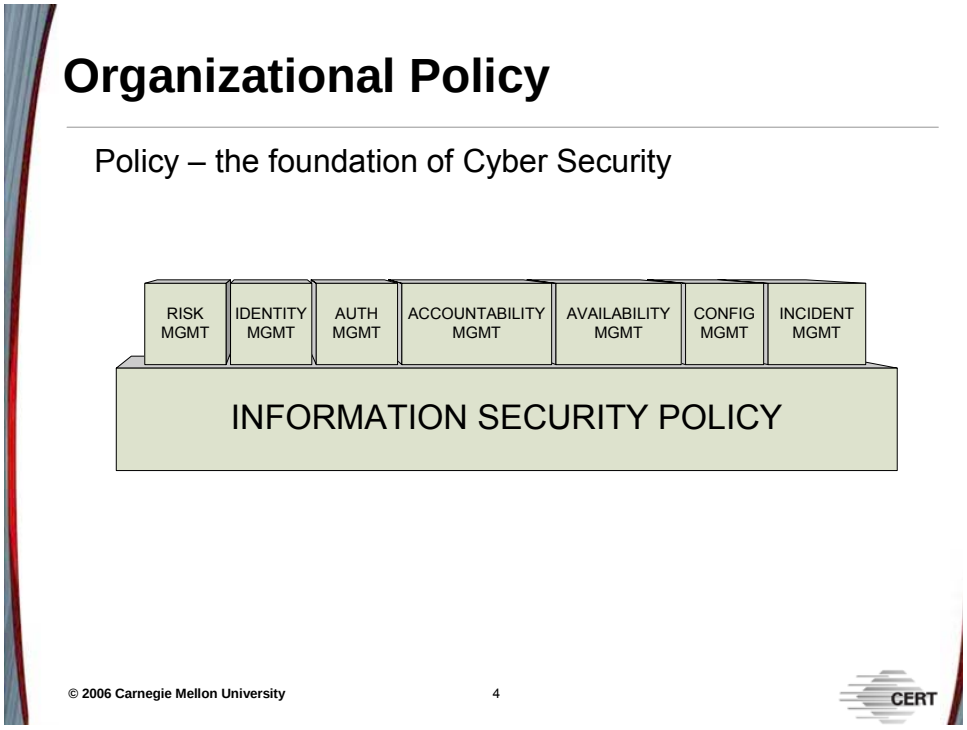
This instructional module will enable students to complete all of the above learning objectives.

2 Overview of Compliance Management



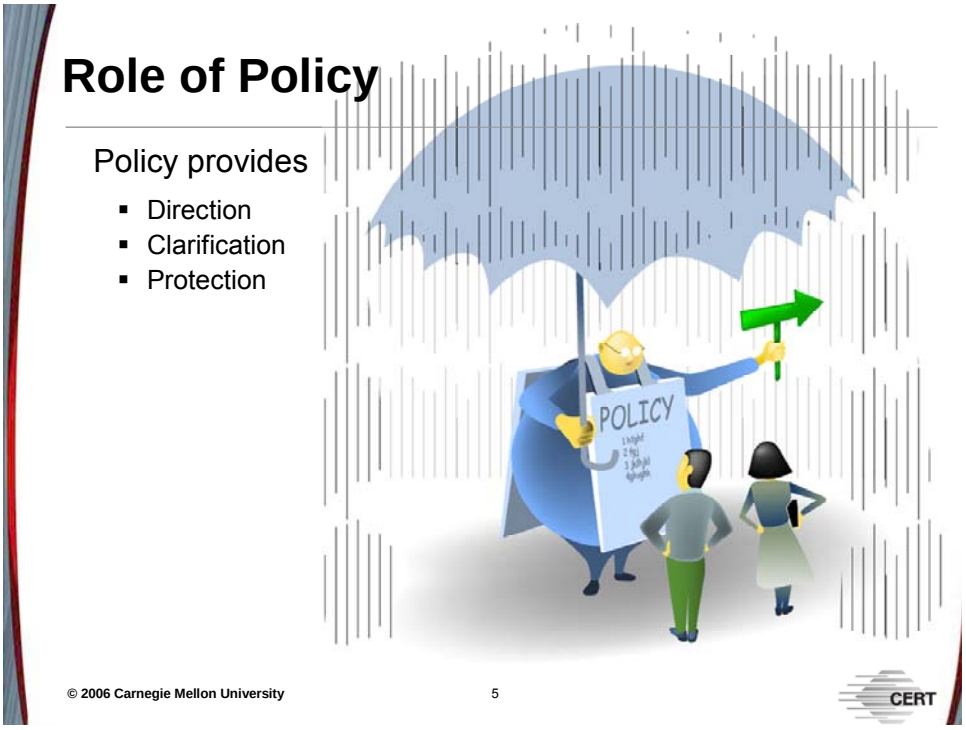
Information security management begins with compliance management—the enforcement of regulations and standards relevant to an activity. We've already noted that compliance management is the most crucial component of Defense-in-Depth and forms the basis for all other components. In turn, policy forms the basis for compliance management. It provides guidance to help users and administrators adhere to the numerous laws that set and govern IA standards.

In this module, we explain policy's primary role in the information assurance arena. We then present methods of policy development and describe how the process unfolds within an organization. Finally, after examining a policy template, we discuss the realm of law and the regulation of information security.



2.1 Defining Policy

What is policy? It is the set of rules by which we operate. As such, it provides direction, clarification, and protection. Different organizational environments describe these rules with different vocabularies. Rather than make distinctions between standards, regulations, and procedures, we will use the term *policy* to denote the whole set of rules that influence an organization.



2.1.1 Role of Policy

Direction – Policy translates the overall goals of the organization into practical specifics, or rules. These rules direct organization members toward fulfilling the organizational mission and prevent them from working at cross-purposes. For example, adherence to policy should prevent a situation in which half a company's members strive to maximize profit and the other half squander resources. Policy sets forth the purpose of an organization's existence and describes ways to act to promote that purpose.

Clarification – Policy can be used to clarify goals. General goals can be framed as more specific tasks or procedures, tailored to the operations of a particular group.

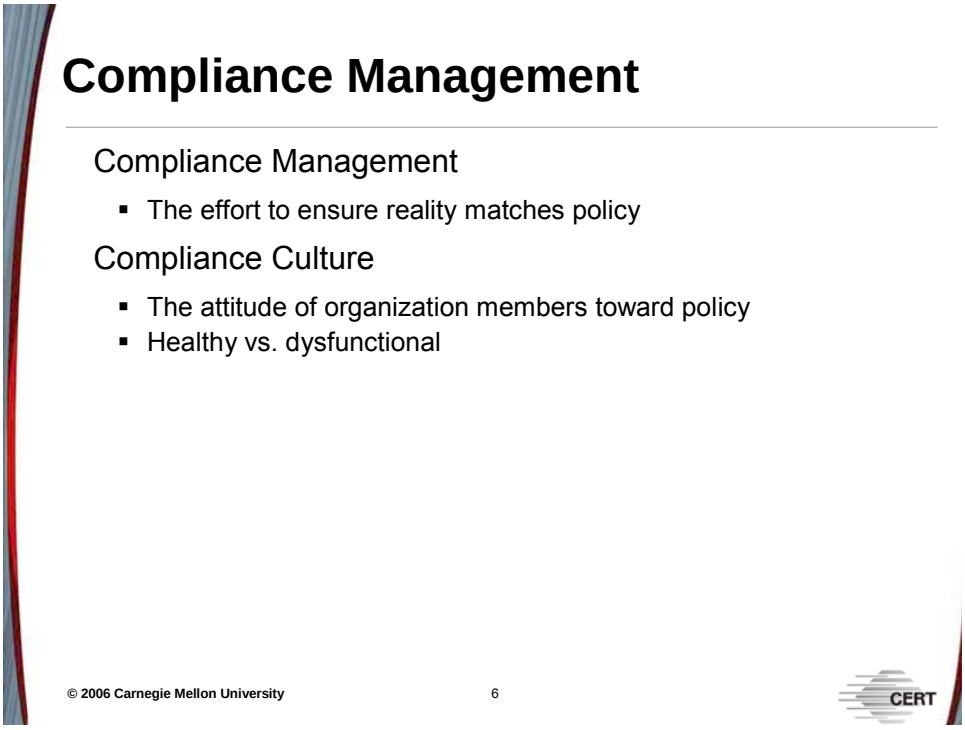
Protection – As organizational law, policy should define potential violations and facilitate enforcement. It is in the organization's best interest to be clear and specific in these definitions. Consider, for example, a situation in which an employee willfully posts defamatory content on his company's Web site. This act can hurt the credibility, and thus profitability, of the organization. The company might terminate him. However, if no clear policy were published about who could post what, with clearly defined consequences for violations, the employee could sue the company for wrongful termination. On the other hand, if the policy forbidding his action were clearly established and generally made known, he would have no argument with any standing.

Indeed, everybody in an organization contributes to the success of its policy. In general terms, managers are charged with keeping personnel in their sphere of influence focused on the organization's goals by ensuring people and systems are in compliance with policy.

Managers at all levels, to varying degrees, are also responsible for creating policy. But other organization members also should take part in the implementation and enforcement of policy. Ideally, this process should involve not only chief officers but feedback from employees who are ultimately affected by the policies.

It is also vital that executives set the tone at the top of the organization by respecting and adhering to policy. Many organizations tend to mirror the attitudes of their executives, so without executive buy-in, policies likely are doomed to fail.

As integral as policy is to the life of an organization, it is often dismissed as unimportant. This is a grave error. All other information security management components draw heavily upon policy; that is why we've chosen to discuss it first.



Compliance

Within this framework of policy, compliance can be described as ensuring reality matches policy. Compliance management is the effort required to make sure what actually happens is in line with what an organization says should be happening in accordance with all applicable laws, regulations, and standards guidelines. Individuals, managers, and executives all have a part to play in fostering compliance.

An organization striving for compliance should systematically develop policy, effectively train the workforce, and promote a positive organizational culture toward policy. Policy can be an effective tool for achieving coordinated, unified progress toward organizational goals.

Culture

The attitudes of all—everyone in management and the workforce—combine to create a culture of policy. This culture is readily apparent to those looking for it and can span the spectrum from a complete lack of concern for the rules to a total focus on the rules. Either position can indicate an unhealthy organization with a dysfunctional culture of policy.

When policy is not enforced but rather disparaged as something with little impact on reality, the culture of policy must be changed. The dysfunction may arise from a lack of attention by organizational policy makers, as when a given policy is irrelevant or out of date. (We've all known "policy" to invoke the connotation of a big book buried on a shelf.) Such inattention indicates a broken policy process and affects the culture negatively. Regular periodic review is the corrective measure, with attention paid to policy's support of organizational goals and

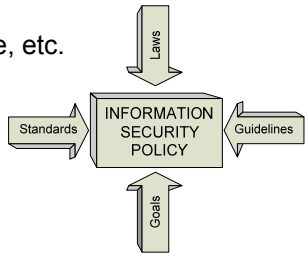
to feedback from the workforce. On the other hand, if a policy itself is current and relevant, lack of adherence might indicate educational or behavioral problems. In such cases, the organization's training or enforcement programs should be bolstered to improve the culture of policy. A major effort must go toward having all members accept the policy as beneficial; otherwise, it may be necessary to spend significant effort on enforcement. Again, the tone at the top is vital in predicting the outcome of any policy initiative.

Policy Development -1

Policy is not developed in a vacuum.

Various influences

- Standards
- Guidelines
- Laws
- Organizational goals: profit, service, etc.

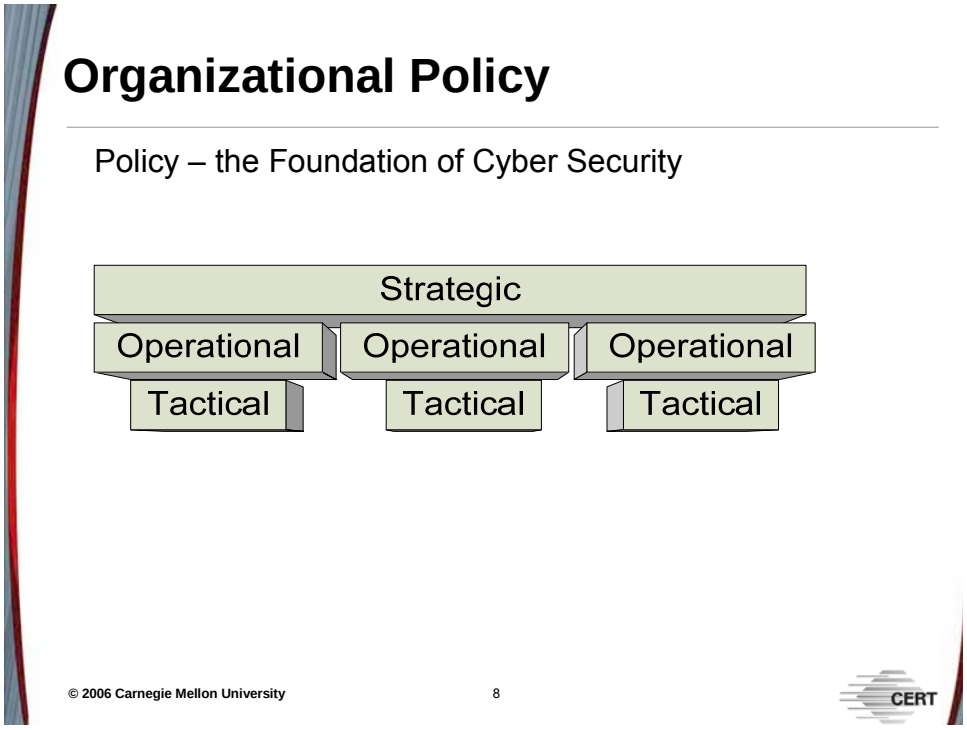


CERT

2.2 Organizational Policy Development

An organization's policies must take into account information from various sources. Standards can shape internal policy. Standards are defined by international bodies like the International Standards Organization (ISO) to ensure global interoperability. Government standards from the National Institute of Standards and Technology (NIST) ensure that regulations are defined to improve system compatibility, levels of reliability and performance, as well as set required securing baselines.

In the end, the mission of the organization and its strategic objectives will drive internal policy development.



2.2.1 Layers

Primarily, policy development is structured around the concept of layers. It originates from general principles and plans at a high level and grows continually more specific for each sub-level of the organization. This structure allows a specificity that is flexible where appropriate. It may be helpful to consider layers of policy at three different levels: strategic, operational, and tactical. These levels are not necessarily fixed and, in fact, may differ among organizations; however, they encompass the conceptual divisions of policy.

Strategic-level policy is the most general and outlines the "what" of the policy–"What is it that we are trying to accomplish?" For example, "What are our security goals as an organization?" These goals may include protecting customer data, safeguarding intellectual property, and maintaining system integrity and trustworthiness. Indeed, strategic-level policy is where one finds stated the overarching goals of an organization and the general methodology and plans for reaching them. Responsibility for strategy and vision rests with the executives of the organization; they set the direction that they want the company to follow.

Operational-level policies, in tandem with operational plans, add some definition to the general, strategic policy and may focus at a fairly general level on the "how" of implementation–"How are we going to accomplish this?" To answer this question, operational managers may evaluate different approaches to the same goal and designate the best one as policy. Specific step-by-step details will be covered further in tactical-level policy.

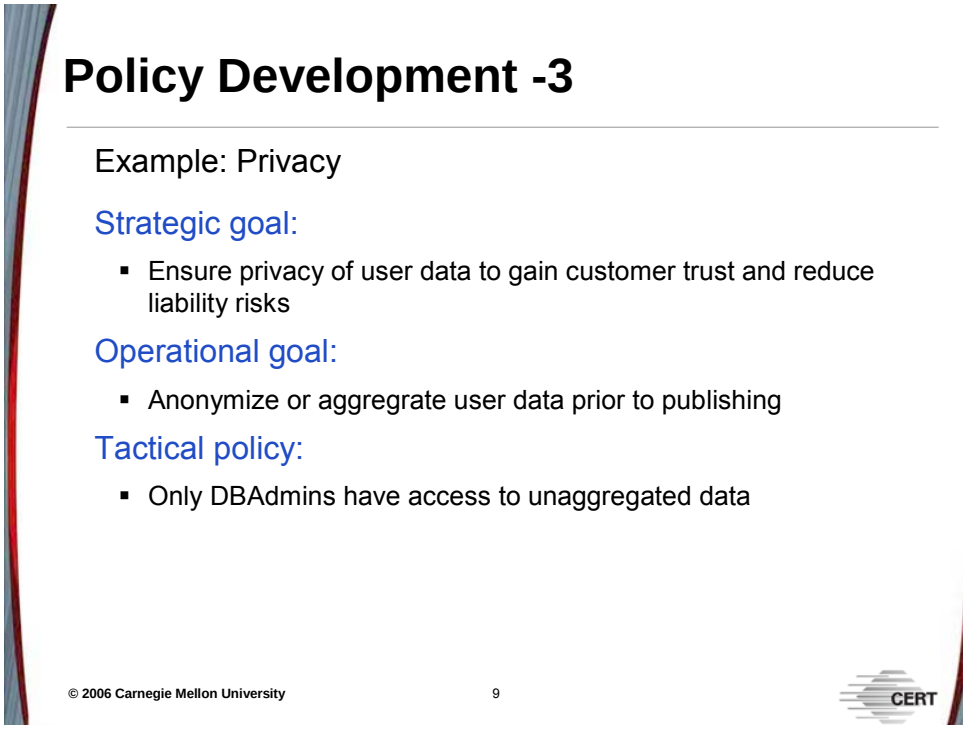
Indeed, tactical-level policy provides the most granular expression of what is supposed to happen. It codifies practices and procedures to ensure operational policies are respected. One is likely to discover the "who, when, what, where, and how" details of policy at this layer.

Regardless of level, all policies should have

- Defined authority and approval or sign-off from management
- Definition of terms
- Defined roles and responsibilities
- Effective date
- Version number and last review date
- Owner and/or maintainer
- Distribution list
- Related training and education
- Defined process for handling violations
- Reference for online storage and access (i.e., URLs, etc.)

When a policy is disconnected from reality, this will generally manifest at the tactical layer. Because certain tasks can be done in many ways, workers may be able to comply with an operational-level policy while disregarding a corresponding tactical-level policy. To maintain a positive policy culture, managers must attend to these discrepancies. Educating workers on the risks of deviating from standard procedure can convince them that the right way to operate is according to policy. Also, managers should consider whether a particular policy is too restrictive or excessively hinders employees in their work.

The example on the following slide illuminates many of the concepts discussed here.



2.2.1.1 Example of the Layered Approach

The issue of user data privacy, although a bit simplistic, provides a small-scale example of the how the layered approach works in developing policy.

First, at the strategic layer, the fundamental goals of the organization must be considered, and the policy's mission and scope must be defined. In most businesses, one goal would be to increase profit. In keeping with this goal, a company might want to 1) nurture user trust and 2) reduce liability risks. Gaining user trust strengthens user loyalty and patronage to the business, which enhances profitability. Profitability can be further promoted by avoiding costly lawsuits—for losing control of sensitive user data, for example. Since the inappropriate release of individual user data could jeopardize this company's main goal of profitability, the company ensures such data stays private as a matter of policy.

The operational layer has the flexibility to refine this policy further. If the stakes are high enough, management might rule that no user data will be released or compromised. However, in the interest of profitability, there might be a compelling reason to utilize the data. If so, the operational policy would need to ensure that individual data would be anonymized or aggregated prior to publishing. That approach would still be compliant with the overall strategy and would further the company's goal of greater profits. Typically, if the strategic policy is written generally enough, there should be a fair amount of latitude for decision making at the operational layer.

At the tactical layer, the actual procedures used to protect data are defined. At this point, the policy describes who is supposed to do what. This level of detail often provides specifics on

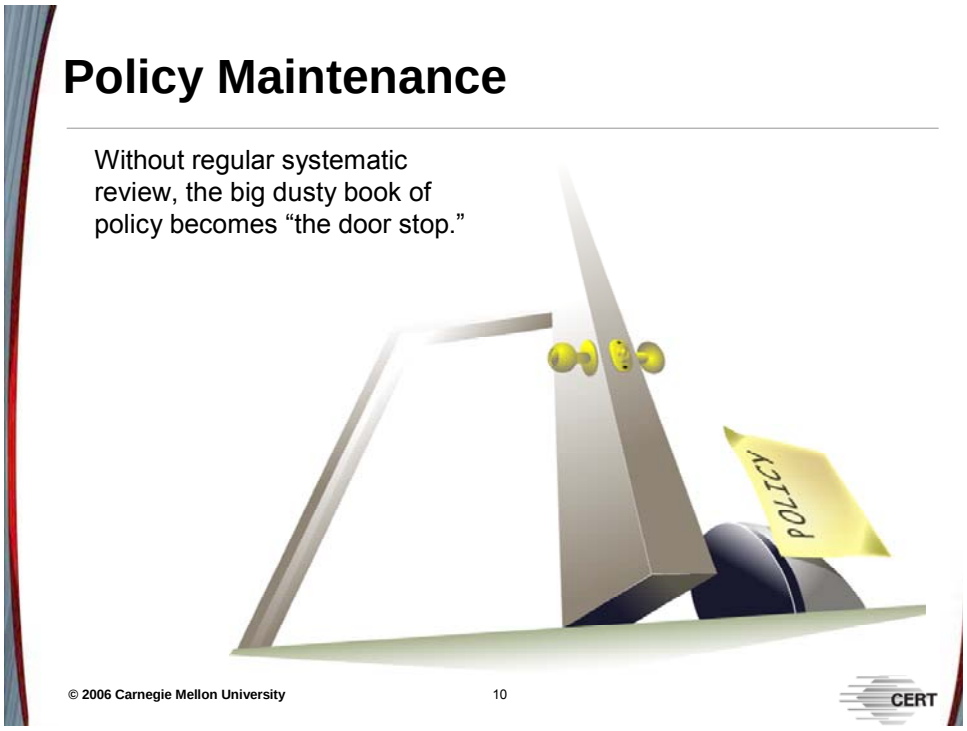
"when" and "how." The policy in the current example might order database administrators to deny access to certain data to all users except those who have been explicitly authorized. It also likely would mandate some form of internal auditing so that compliance could be monitored. If aggregation routines were used to remove specific individual information, those routines would require periodic review to ensure they continued to accomplish the task. Sometimes, tactical policy will not dictate how to follow certain rules and thus will provide leeway in how different people accomplish the same task. When there is a need for standardization, however, the policy may provide a level of detail that ensures everybody performing the task follows the same procedure.

2.2.1.2 Flow

As discussed earlier, it is important to write policy with layers in mind. This approach allows policy to grow clearly from a general concept to specific tasks. It also ensures subordinate policies support and reinforce higher-level goals, rather than being isolated and discordant. Throughout the policy writing process, input from those who will be implementing the policy must be solicited and seriously considered. This "bottom up" validation of policy has the important effect of ensuring realism in the written policy.

2.2.1.3 Risk

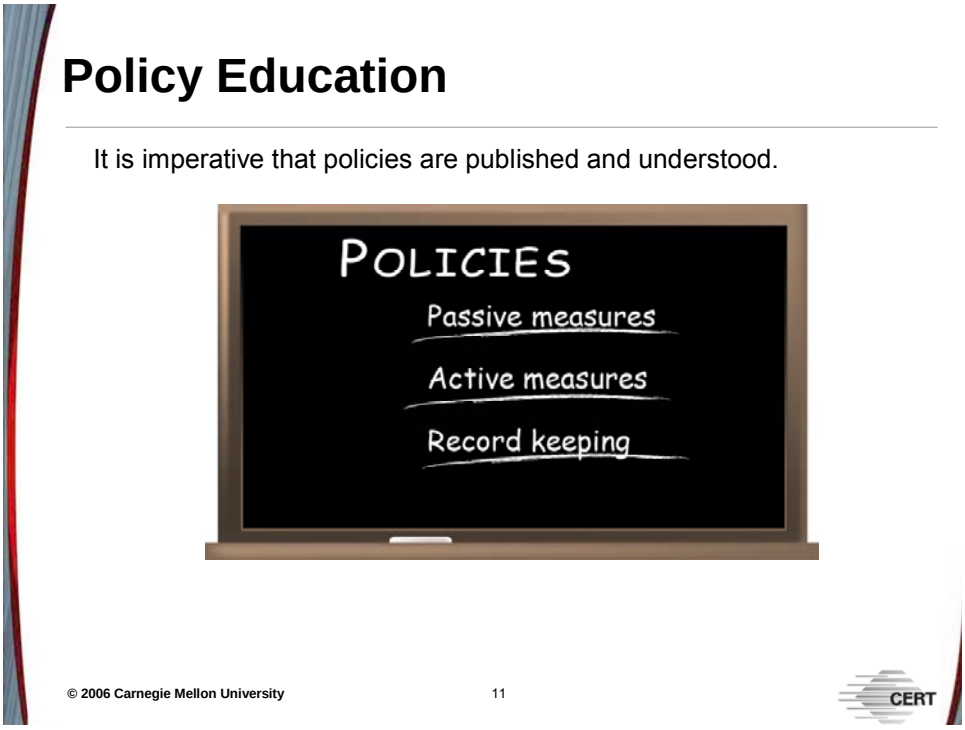
A critical aspect of developing policy is the study of risk. It is not sufficient to define only what an organization wants to accomplish. A further step must be taken to identify any obstacles that might impede progress toward those goals. This process is commonly known as *risk assessment*: an analysis of potentially adverse circumstances and the negative impact they might have on reaching success. Generally, risk assessment methods involve categorically listing possible negative events and assigning them some indication of likelihood. After threats are determined, procedures and plans are created to reduce the likelihood of those threats and/or to reduce the negative effects of those threats should they occur. The key outputs of this process are mitigation measures. These risk mitigation plans are distilled into policy at various levels so that an organization can prepare for threats during normal operations. Risk analysis is an important tool for developing sound policy, and the risk management process will be discussed in more detail in the next component lesson.



2.3 Maintenance

Once policies are formalized, it is imperative that a company maintain a schedule of policy review. Most organizations are dynamic in that their business goals—or the emphasis among those goals—can shift from time to time. Published policies must continue to match higher-layer goals. When a higher-layer policy changes, whether formally or informally, and subordinate policies are not adjusted to reflect those changes, the written policies lose credibility. When people start changing their tasks and procedures to achieve new goals, the underlying policy therefore must be updated or it will become outdated. It is when policy documents become static or unchanging that they become those “dusty books on the shelf.”

In the information technology field, these dynamics can be even more pronounced. As software or hardware changes, new procedures often become necessary. Organizations must adapt by adopting mechanisms for policy and procedure review. They may implement a periodic policy review or a less structured change review. A periodic review entails examining policies on a regular basis and adjusting them based on changes to upper-layer policies or on feedback from in-the-trenches employees. Meanwhile, a change review implies that a set of policies is only reviewed when a higher-layer policy changes. Any review process should take the “top down–bottom up” approach into consideration, striving to both support high-level policy and consider feedback from subordinate employees.



2.4 Policy Education

Policy efforts do not stop after sound policy is developed or reviewed. It takes considerable effort to disseminate the policy information accurately to those to whom it applies. For people to adhere to policies, they obviously must first know and understand them. A traditional method of disseminating human-resources policies is through a policy bulletin board. No doubt, all employees have seen such boards indicating their rights as employees of their organizations. Many of these notices are produced by the U.S. Department of Labor in accordance with federal laws that mandate public posting of certain rights. Information about the Fair Labor Standards Act and the Occupational Safety and Health Act is a common sight in many workplaces. The same bulletin-board method can be helpful in disseminating IT policy detail. IT policy also can be published electronically, via an intranet Web server or downloadable documents. It is important to provide a bullet-point summary through this medium while also providing the more detailed text of the policy itself. This approach allows employees to quickly comprehend the main emphasis of the policy and gives them direction for further investigation if necessary.

Such passive education techniques are low-cost ways of publishing policy. However, to offer legal protection for individuals and the company, it must be verifiable that the policies were not only known but understood. Therefore, managers may conduct more active training through classes or small group counseling, online learning, and employee orientation training. Classes can be taught by managers themselves or by company policy educators and can provide some verification of employee understanding, especially when complemented by a quiz or test that reviews the policies covered. Yearly refresher courses can provide

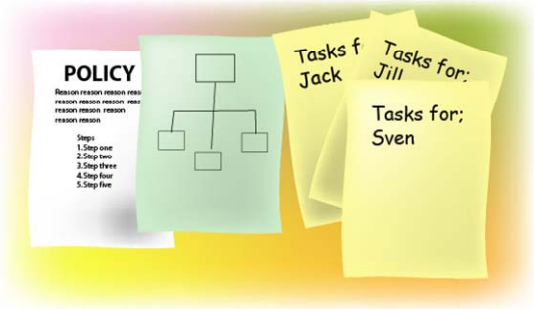
assurance that employees remain up-to-date with organizational policies. Ultimately, it may be desirable to record employee signatures to acknowledge understanding of policies governing their positions. In the event that legal action against an employee becomes necessary, management then will have substantial justification with which to prosecute. The employee will have a much more difficult time pleading ignorance if he or she has signed a statement that claims understanding.

Sample Policy -1

Develop policy from general to specific.

Distill goals into actions.

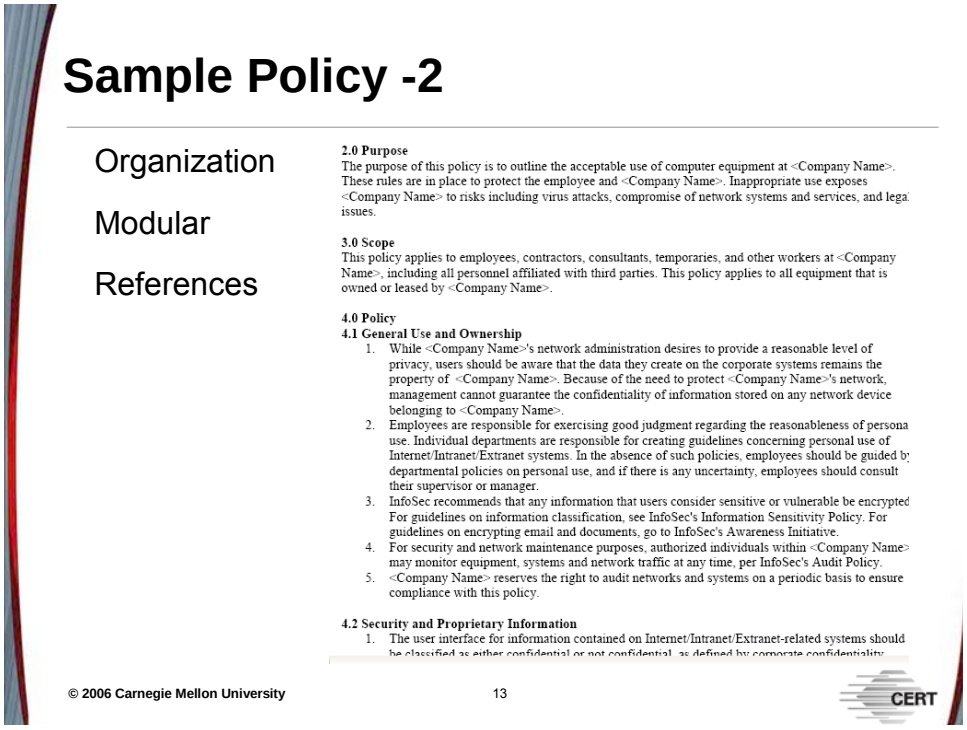
Assign responsibility.



CERT

2.4.1 Policy Sample

A wide range of IT policy template resources are available. The SANS Institute offers free templates online for many information security topics that are relevant in today's workplace, such as acceptable use, acceptable encryption, information sensitivity, network monitoring, and remote access, among others.



2.4.1.1 Organization

The SANS Institute policy templates generally follow the pattern described below:

- Purpose – states the goal of the policy
- Scope – outlines to whom the policy applies
- Policy – describes the rules in effect for this policy topic
- Enforcement – publishes the penalty for noncompliance
- Revision History – details any changes that have been made to the policy

When necessary, the policy document can include an overview to provide background for the purpose section. Also, a paragraph may be reserved for any definitions that are not common knowledge or that need to be clarified.

Other policy resources that you may want to review include

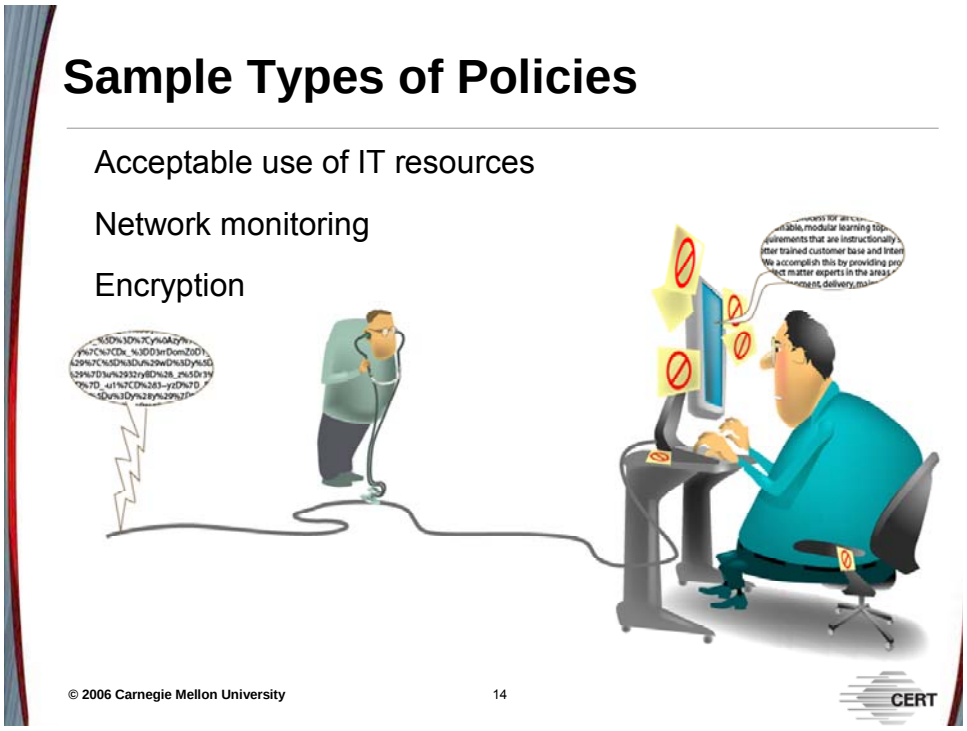
- Information Security Policies Made Easy, Version 8, PentaSafe, 2001 (ISBN 1881585077) by Charles Cresson Wood
- Site Security Handbook (RFC2196)
  http://www.ietf.org/rfc/rfc2196.txt
- EDUCAUSE/Cornell Institute for Computer Policy and Law
  http://www.educause.edu/icpl/

2.4.1.2 Modular

It is a common practice to write policy using a modular methodology. To prevent confusion, a specific policy should be distilled into rules pertaining to the single topic being addressed. Education efforts are also aided by this process because the body of policy can be indexed and easily searched. Rather than scanning through policy text and trying to determine which rules apply to which topics, a glance at an index or table of contents provides quick access to the particular policy being sought.

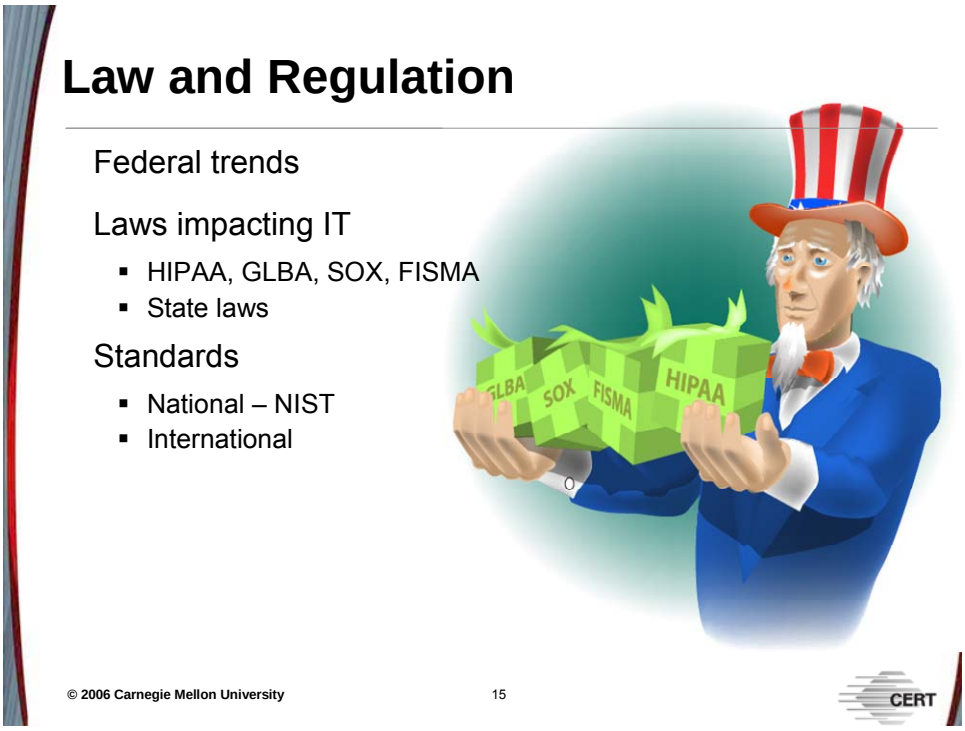
2.4.1.3 References

Some policies build on other policies. Therefore, it is important to include references to all related policies, from high-level to low-level. Maintaining good references while writing policies has two positive effects. First, when a low-level policy specifically references the higher-layer policy that prompted its existence, it is difficult to create contradictions between the two layers. Second, the compliance management process is simplified by the ability to trace high-level policy requirements to specific rules implemented in lower-level policies.



Sample Policy Summary

To be effective, policy requires executive support, managerial oversight, and workforce acceptance. An inclusive development process and, more importantly, a substantial review process including quality assurance can help organizations meet these requirements. This creates a positive policy culture in an organization.

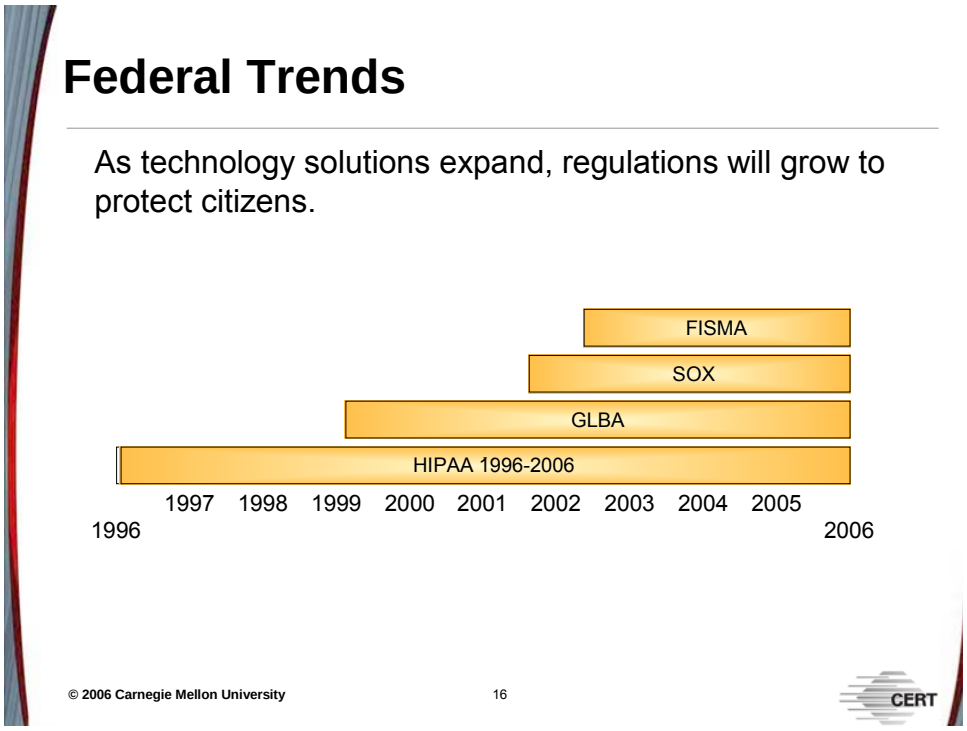


2.5 Regulatory Inputs to Organizational Policy

2.5.1 Law and Regulation

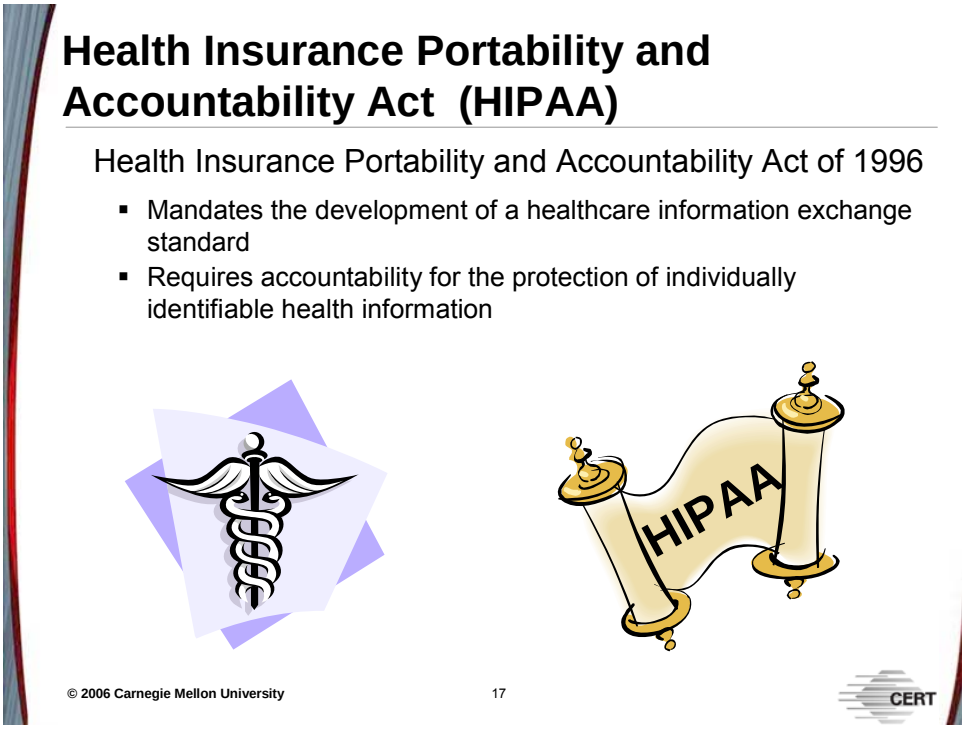
Policy is not developed in a vacuum. Policy is developed through a process and with many inputs into the result. An unwritten policy in nearly every organization is to conduct operations in accordance with applicable legal requirements. Therefore, one of the most important considerations informing an organization's internal policy is law. Law, however, rarely provides the specifics necessary to comply with its intent. Generally, supporting documents are created to expand the requirements of law into detailed actions. These documents are published as standards or guidelines. Standards are also generated by groups of organizations with common interests. In these cases, standards are accepted and enforced because of their benefit to the whole group.

Note that while the laws and regulations discussed in this course tend to be U.S.-based, the general concepts of managing compliance with legislation are applicable to all.



2.5.2 Regulatory Trends

Due to the interconnectedness of the information infrastructure, even when one organization is properly secured, threats can arise through less secure neighbors or partners. Therefore, as business, government, and society have increased their dependence on technology, organizations and governments have created laws and regulations for those working with technology to provide information security. Specifically, standards are being created that mandate information security policy and that require accountability for compliance with those policies. This auditing of policy adoption and implementation helps ensure greater attention to security through industry, government, and even the world. After all, to expand the reach of sound information security policy beyond mandated laws, several industry groups have merged best practices into standards to be followed by all members of the group.



2.5.3 Law

2.5.3.1 Health Insurance Portability and Accountability Act

The Health Insurance Portability and Accountability Act (HIPAA) of 1996 was enacted to improve the U.S. healthcare industry. The administration and financing of this industry has traditionally been labyrinthine. The process of submitting claims and getting money routed to the proper places is challenging, especially because there historically have been different data requirements among various entities. Before the law's enactment, it was also difficult to switch healthcare providers or health insurance companies, since the proprietary data formats forced a considerable switching cost. In the early 1990s, the healthcare industry faced increasing administrative costs at high rates, and it was widely acknowledged that the inefficiency was unacceptable for future needs. Thus, two related goals of HIPAA are to simplify the administration of the nation's healthcare system and to improve its efficiency and effectiveness.

Title II of the Act enacts the law for "Administrative Simplification." Responsibility for the oversight and development of regulatory standards was assigned to the Department of Health and Human Services. The legislation covers four general topics and requires that for each of these areas standards be developed and enforced:

1. Standards for Electronic Transactions – to facilitate data exchange by unifying data format structures and procedures

2. Unique Identifiers Standards – because the health information flow becomes portable, different entities must be identified consistently. These standards dictate a common method for constructing new identifiers that do not conflict with any other entities' identifiers.
3. Security Rule – mandates sound information security processes
4. Privacy Rule – describes lawful and unlawful disclosure of health information. This rule deals with privacy as an overarching concept, rather than with technological privacy only. It has more to do with purposeful disclosure of information than with technological means of protecting privacy.

The significant aspect of HIPAA with respect to information security policy is the Security Rule. The Department of Health and Human Services published the final Security Rule in April 2003. "Covered Entities," those healthcare institutions affected by the law, were given 24 months to comply. The goal of the Security Rule is to convince organizations that sound policy related to information security is essential to successful operations. The rule requires organizations to designate the person responsible for security, as well as to implement a policy process including risk assessment, compliance management, policy review, and audits. It also mandates some measure of personnel screening, access management, and security education and training. To facilitate readiness when security problems arise, the rule requires organizations to have formulated incident response mechanisms and to have implemented an operations continuity plan.

In addition to these administrative safeguards, the rule also categorizes physical and technological safeguards. Physical safeguards cushion the impact of infrastructure decisions on security. Technological safeguards offer specific principles for handling electronic information, without mandating a solution dependent on a specific technology. An interesting aspect is the emphasis on technological auditing solutions. The rule calls for some degree of continuous logging and reporting. Enough information must be kept to track what was done by whom.

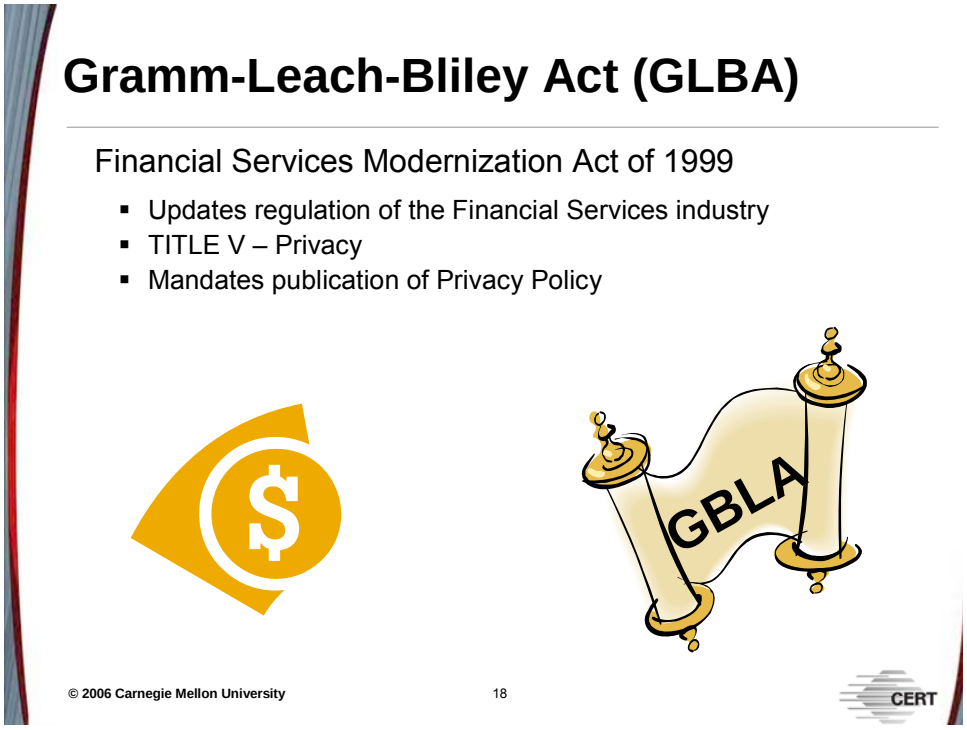
HIPAA imposes fines and imprisonment for those who fail to comply with its standards and requirements; however, the language leaves latitude for grace toward those who are making efforts. Outright penalties seem directed toward those who are actively rebelling against the standards. Further, the act outlines a graduated increase in severity for wrongful disclosure of individually identifiable health information. Sanctions range from $50,000 to $250,000 in fines and 1 to 10 years of imprisonment.

The following outline of the administrative safeguards contained in the Security Rule provides a concise formulation of security policy. The entire Act can be read online at http://www.legalarchiver.org/hipaa.htm.

Sub-part C: Security Standards for the Protection of Electronic Protected Health Information

§164.308 – Administrative safeguards

- security management process: implement policies and procedures to prevent, detect, contain, and correct security violations
  - risk analysis
  - risk management
  - sanction policy
  - information systems activity review
- assigned security responsibility
- workforce security
- information access management
- security awareness and training
  - security reminders
  - protection from malicious software
  - login monitoring
  - password management
- security incident procedures
- contingency plan



2.5.3.2 Gramm-Leach-Bliley Act

The Financial Services Modernization Act of 1999, otherwise known as the Gramm-Leach-Bliley Act (GLBA), updated federal regulation of the financial industry. The act was written into law to make obsolete the Glass-Steagall Act, which prevented banks from offering investment, commercial banking, and insurance services all together. The GLBA allows banks to consolidate these services but mandates that they protect the privacy of their consumers by safeguarding nonpublic consumer information, informing consumers of the institution's privacy policy, and allowing consumers to prevent their information from being shared with unaffiliated parties. Financial institutions are also required to securely store, transmit, and dispose of electronic data as a matter of consumer privacy and to prevent theft of information

The law applies not only to companies traditionally recognized as banks, but also to any company that deals with consumer financial information. This less traditional "banking" role includes brokering, tax preparation, money transfer, debt collection, credit counseling, and other areas. With the GLBA, legislators recognized the networked effects of financial-sector business.

Even among traditional banks, after all, financial information is often traded. It is an accepted practice for banks to sell loans to other organizations to collect. With the sale of the loan, a consumer's private financial data transfers to an organization with unknown credibility. In an effort to ensure these types of data exchange were conducted in a manner

that protects consumer interests, the GLBA created a Privacy Rule, a Safeguard Rule, and provisions against pretext.

Different types of financial institutions are accountable to one of several different regulatory agencies: federal banking agencies, the Securities and Exchange Commission, the Commodity Futures Trading Commission, and state insurance authorities. Those that do not fall under the jurisdiction of one of these organizations are regulated by the Federal Trade Commission (FTC).

Privacy

The Privacy Rule is an effort to ensure financial institutions clearly communicate their information sharing practices with customers. Specifically, this rule requires banks and other financial institutions to provide customers with a written copy of their practices with respect to maintaining individual privacy. This privacy notice should clearly state what information is kept, which third parties it is shared with, and how the company protects it. In this way, the act provides some measure of control to customers. Also, banks generally collect more data than is essential, and the Privacy Rule mandates that they must provide customers with an option to omit unnecessary data from information sharing and selling practices.

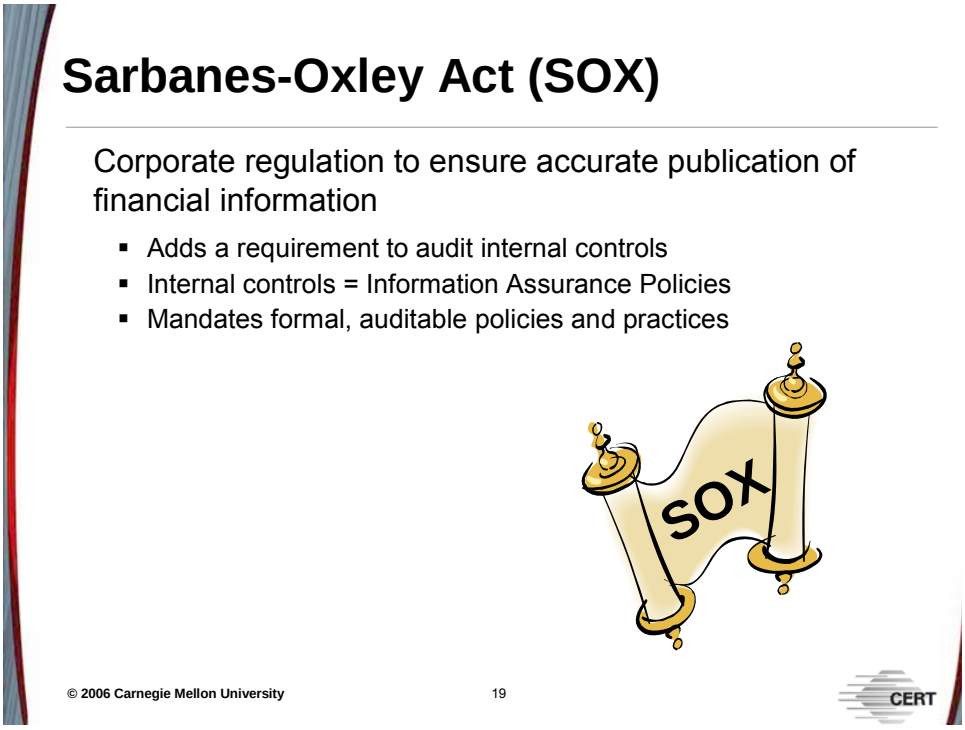
Safeguard

The Safeguard Rule formalizes accountability and responsibility for information security programs. Under this rule, financial institutions are required to develop, implement, and maintain written information security programs that contain comprehensive administrative, technical, and physical safeguards. As part of its program, each financial institution must do all of the following:

- Designate an employee or employees to coordinate its information security program.
- Identify reasonably foreseeable internal and external risks to the security, confidentiality, and integrity of customer information that could result in the unauthorized disclosure, misuse, alteration, destruction, or other compromise of information, and assess the sufficiency of any safeguards in place to control such risks.
- Design and implement safeguards to control reasonably foreseeable risks, and monitor the effectiveness of these safeguards.
- Take reasonable steps to select and retain service providers that are capable of maintaining appropriate safeguards for customer information, and require those service providers, by contract, to implement and maintain such safeguards.
- Adjust the information security program in light of developments that may materially affect the entity's safeguards.

Pretext

The other provisions of the GLBA aim to reduce deceptive marketing or operational practices in the financial industry. They prohibit the practice of obtaining customer information from financial institutions using false pretenses—that is, under pretext. The provisions also prohibit the sharing of account numbers with non-affiliated marketing firms, such as telemarketers, mass mailers, or email marketers. This reduces confusion about who is offering a certain product and makes it more difficult for customers to mistake marketers for "the bank." The Act can be read online at http://ftc.gov/privacy/glbact/glbsub1.htm.



2.5.3.3 Sarbanes Oxley Act

In the wake of corporate corruption scandals at the end of 2001, most notably involving Enron and WorldCom, the U.S. Congress initiated and passed legislation to mandate legal accountability in corporate accounting reports. The intent of the Sarbanes-Oxley Act (SOX) of 2002 was to protect the nation's economy by bolstering investor confidence in corporate reports through increased accountability and objective auditing.

Three of the major provisions of SOX are Objective Auditing, Executive Accountability, and Internal Control Assessment.

Objective Auditing

Enron's problems brought to light the practice of using auditors in circumstances where there could be conflicts of interest. The 200-section paragraphs of SOX set basic rules to prevent such conflicts of interest. Specifically, a company cannot use an auditing firm for which an executive has worked recently; further, audits should be conducted by partner teams that exchange primary responsibility every several years. The Act also mandates that the audit is no longer solely the responsibility of management, but rather must be shared with an audit committee among the company board members. The Enron case involved considerable ignorance on the part of the board regarding many management practices, including selecting auditing firms with a conflict of interest.

Executive Accountability

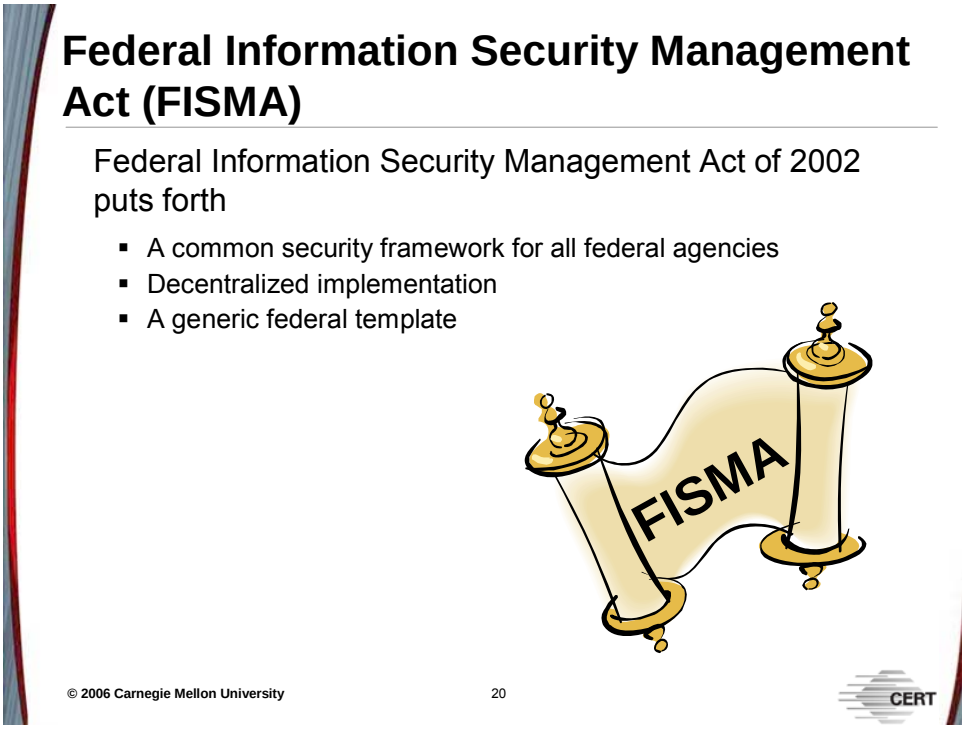
Along with improving the objectivity of corporate financial audits, SOX also assigns legal accountability to chief executives for the accuracy of their company's financial statements. Many corporations are complex structures composed of various entities. It was not uncommon for Enron and WorldCom executives to admit they did not necessarily know how some of those entities' financial statements were generated. CEOs and CFOs are now required to sign their company's financial reports, indicating their accuracy and validity, and to assume personal liability for those reports' accuracy and validity.

Internal Control Assessment

Most importantly to the discussion of information security, the 400-series paragraphs of SOX build a mechanism by which executives can attest to the effectiveness of the internal controls by which financial data is kept accurate. This is a form of information management. For executives to have true confidence in the financial statements, they need to be sure that data is not tampered with or misplaced, but rather is accurately aggregated from its many sources. Investors need the same confidence; hence, SOX mandates objective auditing of these internal controls as well. The Act can be read online at http://www.legalarchiver.org/soa.htm.

Essentially, this means that publicly traded companies must have policies in place to secure their information systems and must prove those policies' effectiveness through an external audit process.

At issue is the integrity of financial data that is used within corporations and presented to the investment community at large. Remember that integrity is one of the pillars of the information assurance CIA triad.



2.5.3.4 Federal Information Security Management Act

In 2002, the Congress of the United States passed the e-Government Act, which mandated improved efficiency of government operations through increased reliance on technological solutions. One portion of this broad-scoped legislation, Title III, created the Federal Information Security Management Act (FISMA). In essence, legislators realized that increased dependence on technology required an increased management effort. They created FISMA to ensure government agencies would take proper precautions with respect to information security in the process of expanding and integrating their information systems.

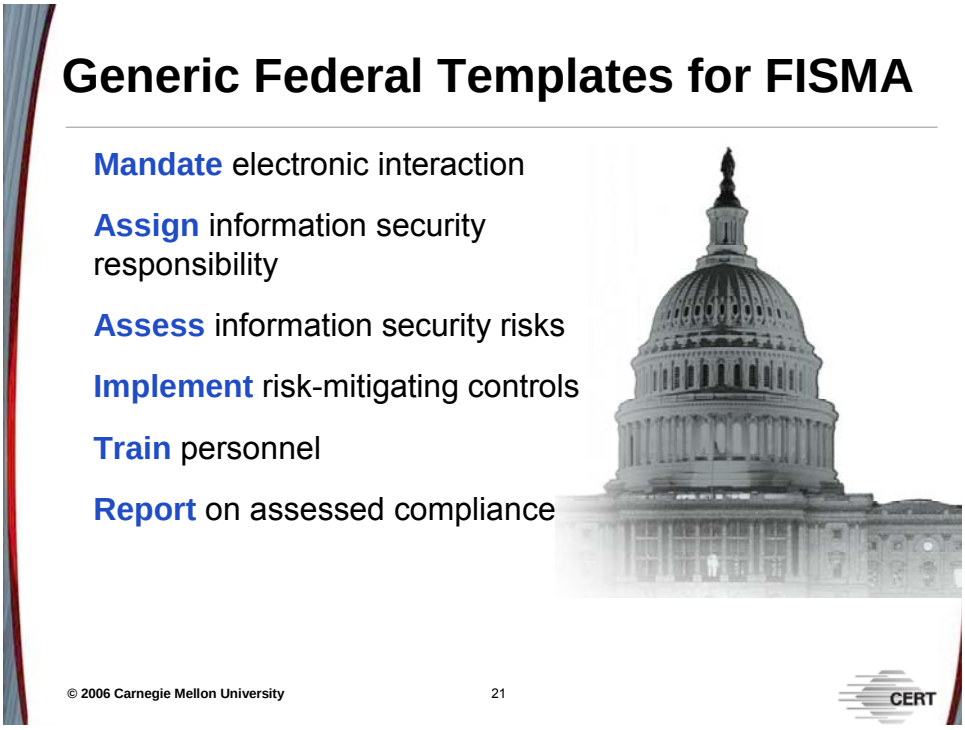
The stated purposes of the legislation are as follows:

- security framework – to provide a comprehensive framework for ensuring the effectiveness of information security controls for resources that support Federal operations and assets, including minimum standards for risk categories
- improved oversight – to provide effective government-wide management and oversight of information security risks related to the government's highly networked computing environment, including coordination with partners
- decentralized operations – to allow agencies to develop their own solutions, with freedom to consider commercial products

FISMA emphasizes risk management and policies to reduce discovered risks. In a networked environment, a partner's security risks must be considered part of an organization's own risks. Because coordination and cooperation are required to increase the security of

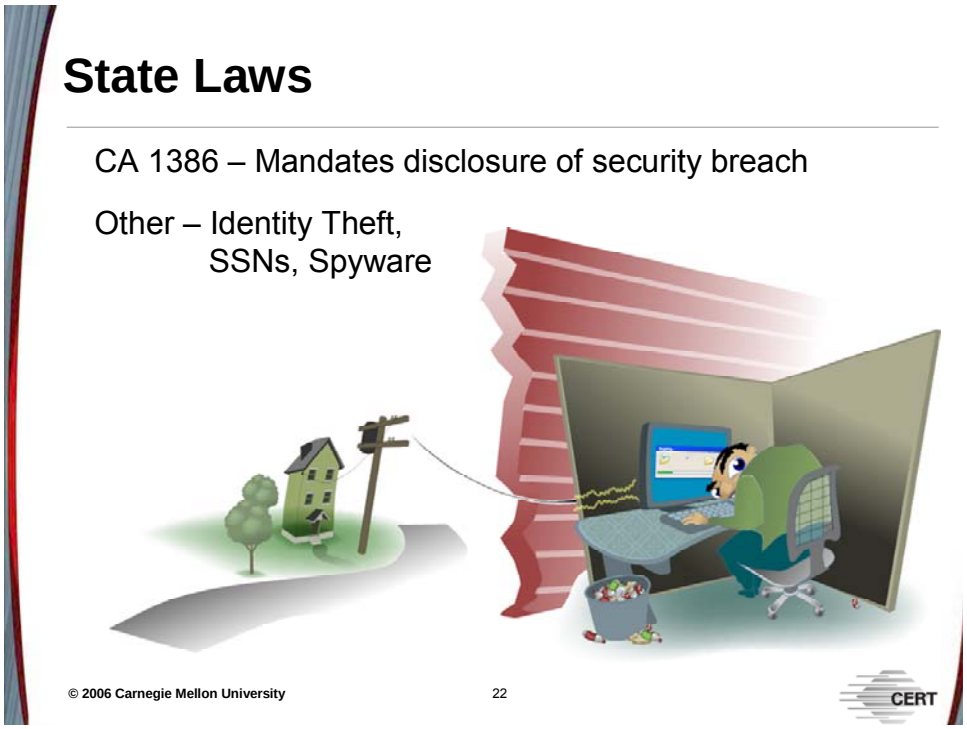
integrated information systems, FISMA is an effort to ensure all parties are working toward the same security goals.

Congress communicated the high priority of this Act by assigning the responsibility for information security efforts to the agency heads. They are responsible for developing, documenting, and implementing agency-wide information security programs that assess risk, craft policy, train personnel, measure status, respond to incidents, report incidents when required to US-CERT and other organizations, and ensure continuity of operations. Agencies are then required to report the status of their programs to congressional committees annually. You can read this Act online at http://csrc.nist.gov/policies/FISMA-final.pdf.



2.5.3.5 FISMA Objectives

The above slide provides the six high-level objectives for implementing FISMA across the federal government.



2.5.3.6 State Laws

Over the last five years, governments at the state level have been following the federal government's lead on information security measures. Most effort at this level is designed to protect citizens from negligent or malicious information disclosure. This type of legislation has been prompted by the ease with which attackers have been able to carry out identity thefts. One simple precaution is to move away from a single identification number, such as the social security number. States are also beginning to address the problems of spyware and spamming. Some states have enacted legislation that imposes penalties for surreptitiously installing software on people's computers to collect personal information for business use.

A notable event regarding information security policy was California's enactment in 2003 of the California Security Breach Information Act[2], a bill mandating disclosure of information compromise [CASB 02]. According to the law, if unauthorized people gain access to personal information held by a company, the company must inform California residents whose data was accessed in the security breach. This is a significant development. Normally, there is little incentive to notify the public about security breaches, as doing so tends to tarnish a company's reputation, and there may also be little incentive to secure information systems with costly controls. By enacting a disclosure law, California forced companies to take the security of private information seriously. Such a law provides incentive for companies to develop and implement their own formal information security policies to preclude disclosure of "lost" data. Several states have followed California's lead with respect to this legislation, indicating a trend toward increasing accountability of

[2] See http://www.legalarchiver.org/sb1386.htm.

information security procedures to state governments. As of January 2006, 23 states had enacted similar disclosure laws.

The burden of notification in response to a data security breach often falls to an IT employee who is familiar with the details of the compromise. If a non-IT employee shoulders the task of customer notification, he or she should consult privacy and security officials to determine what information was compromised and which customers were affected.

In the future, it is also possible that a similar law could be enacted at the U.S. federal level. Already, certain types of organizations, such as critical infrastructure providers, are required to report certain types of computer security incidents to the federal government.

And many international laws also exist, including

- Canada's Personal Information Protection and Electronic Document Act
- The European Union's Privacy Directive
- Japan's Personal Data Protect Act
- The U.K.'s Data Protection Act
- Basel II Guidelines

2.5.4 Standards

Although U.S. law often mandates standards compliance, it rarely gets into the details of standards requirements. Traditionally, it empowers an executive agency to draft and publish a standard to meet the intent of the law.

Of course, not all standards are spurred by Congress. Many standards are created by groups because it is in the best interest of involved organizations to work in cooperation with the larger group. Two examples of standard-setting entities are the International Standards Organization (ISO) and the National Institute of Standards and Technology (NIST).

2.5.4.1 ISO 17799

The International Standards Organization is a critical entity for coordinating operational efforts among different countries. Specifically, the ISO often publishes standards that help differing cultures operate within the same set of rules. This can be helpful to both governments and businesses.

The ISO has published Standard 177993[3] to provide a sound approach to information security management. ISO 17799 actively promotes sound information security policy as one of its primary tenets. It directs that "management should set a clear policy direction and demonstrate support for, and commitment to, information security through the issue and maintenance of an information security policy across the organization."

The ISO 17799 standard consists of 11 primary sections:

1. Security Policy
2. Organization of Information Security
3. Asset Management
4. Human Resources Security
5. Physical and Environmental Security
6. Communications and Operations Management

[3] See http://www.iso.org/iso/en/prods-services/popstds/informationsecurity.html.

7. Access Control
8. Information Systems Acquisition, Development and Maintenance
9. Information Security Incident Management
10. Business Continuity Management
11. Compliance

Again in this standard, we see the familiar pattern of a systematic policy process: risk assessment, policy development and review, compliance efforts, personnel safeguards, technological security measures, and business continuity plans.

The 17799 standard, in combination with a second standard called BS7799 Part II, is being reclassified as part of a broader standard called ISO 27001. BS7799 Part II is the assessment standard for ISO 17799 compliance.

2.5.4.2 National Institute of Standards and Technology

The National Institute of Standards and Technology (NIST) has for nearly a century produced standards necessary to facilitate economic health in the United States. The institute has been involved with information security standards since computers were created and has produced valuable guidelines and practices for the computer era. The major focus of NIST activities in information technology is on developing tests, measurements, proofs of concept, reference data, and other technical tools to support the development of technology.

As a matter of policy, the NIST encourages the development and use of voluntary industry standards. In a careful effort to prevent unnecessary duplication, NIST only writes standards and guidelines when there are compelling Federal requirements and no existing voluntary industry standards to fulfill them. Therefore, much NIST work involves coordinating with industry and international organizations to develop needed standards.

The NIST provides three publications: Information Technology Laboratory (ITL) Bulletins, Federal Information Processing Standards (FIPS) Publications, and Special Publications. ITL Bulletins are written by the Information Technology Laboratory's Computer Security Division and present thorough treatments of significant IT-related topics. FIPS Publications constitute the body of standards applicable to government agencies. Not all FIPS publications are related to information security, nor are they all government-wide in scope; however, they address standards in a comprehensive, modular format. Special Publications, particularly the 800-series information technology security branch, are a way for the ITL to make various reports about its ongoing operations. Through these publications, the NIST sets

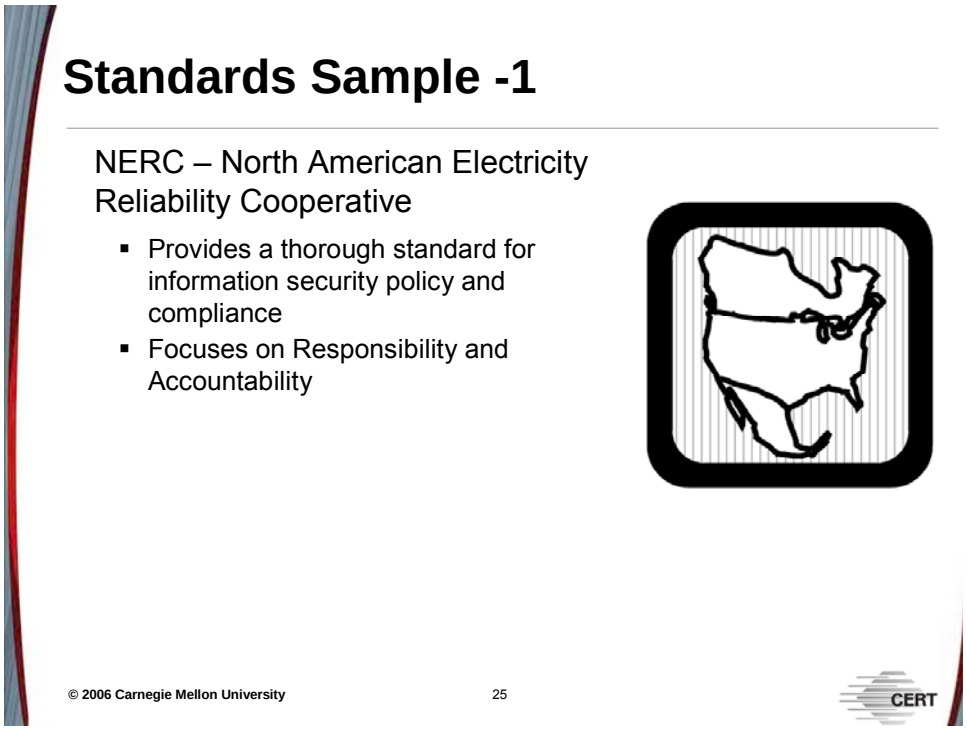
forth guidelines, reports on research, and outlines collaborative efforts with other organizations.

The NIST is often called on to develop standards that support the intent of legislation enacted by Congress. Such is the case with FISMA, the Federal Information Security Management Act. The law requires three documents from the NIST: a standard that defines categories of information in terms of risk; guidelines for assigning information to each category; and a requirement for the minimum risk-mitigating security controls necessary for compliance among government agencies.

The first of these standards has been published as FIPS 199: Standards for Security Categorization of Federal Information and Information Systems. As required by FISMA, federal agencies must develop and report on information system risk assessments. To make cooperation, auditing, and enforcement practical, all of the agencies must have a common vocabulary and terminology. This standard was published for that purpose. It gives all agencies a common starting point for discussing and evaluating their risks.

Further, the Special Publications provide general guidelines that may also benefit agencies affected by FISMA. These supporting documents are helpful not only for those complying with FISMA and FIPS 199, but also for any organization involved with risk and compliance management. They can serve as an excellent starting point for developing policies and procedures.

- **SP 800-14** Generally Accepted Principles and Practices for Securing Information Technology Systems, September 1996.
  http://www.hipaadvisory.com/regs/finalsecurity/nist/800-14.pdf
- **SP 800-16** Information Technology Security Training Requirements: A Role- and Performance-Based Model (supersedes NIST Spec. Pub. 500-172), April 1998.
  http://www.hipaadvisory.com/regs/finalsecurity/nist/800-16.pdf
- **SP 800-26** Security Self Assessment Guide for IT Systems
- **SP 800-30** Risk Management Guide for Information Systems
- **SP 800-33** Underlying Technical Models for Information Technology Security, December 2001
- **SP 800-50** Building an IT Security Awareness and Training Program
- **SP 800-55** Security Metrics Guide for IT Systems
- **SP 800-66** Introductory Resource Guide for Implementing the HIPAA Security Rule



Sample: NERC

Another standard that covers several key information security issues is the North American Energy Reliability Council (NERC) Cyber Security Standard. This document details the standards to which member organizations must conform and can be viewed as a higher-layer policy with which they must comply. Members' organizational policies should reflect and reinforce the components of this standard. In fact, this standard outlines the requirements essential to a member organization's cyber security policy.

Most notably, this standard focuses heavily on responsibility and accountability. It requires a cyber security policy that lists the individuals in an organization who have authorized access to secure information. It also formalizes a compliance monitoring process that outlines the audit tasks necessary to prove compliance. Further, it delineates multiple levels of noncompliance. This granularity is important in a standard with such wide breadth as this one, but it is often also helpful in policies of smaller scope. Compliance is not always a black-and-white, yes-or-no proposition. The audit reports for compliance monitoring allow for these different degrees of compliance that can then correspond to varying sanctions.

Standards Sample -2

2. Levels of Noncompliance

2.1. Level 1:

2.1.1 A senior manager was not designated for ten or more calendar days, but less than thirty calendar days during a calendar year; or,

2.1.2 A written cyber security policy exists, but has not been reviewed in the last three calendar years; or,

2.1.3 Deviations from requirements or written cyber security policy have not been documented within thirty calendar days of the exception; or,

2.1.4 A program to identify and classify information and the processes to protect information associated with Critical Cyber Assets has not been reviewed in the previous full calendar year.

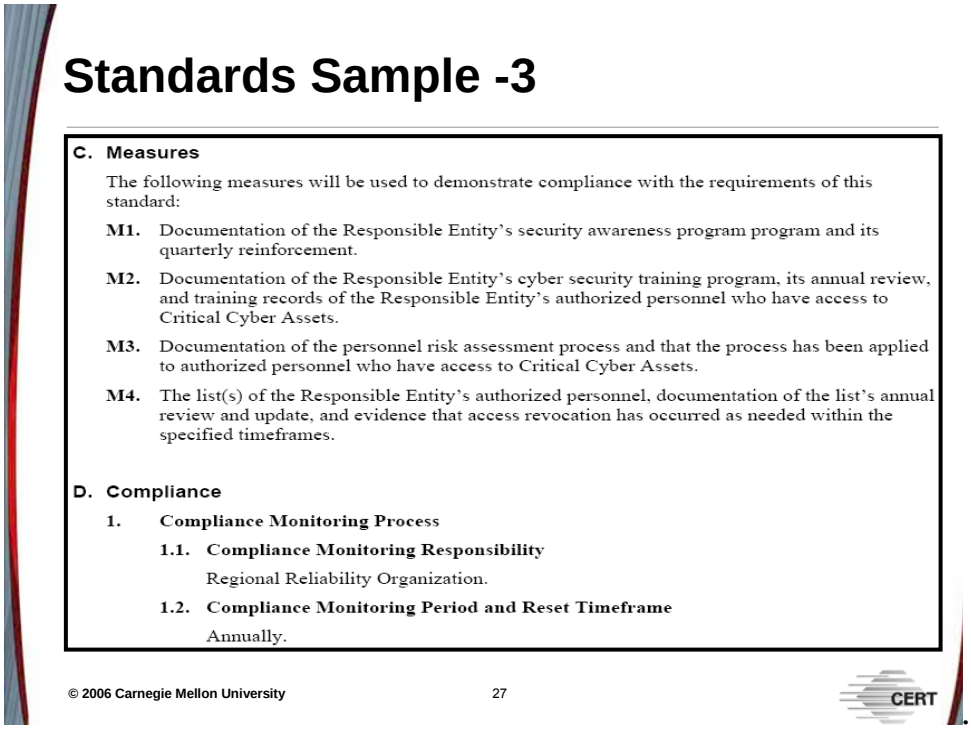
2.2. Level 2:

2.2.1 A senior manager was not designated for thirty or more calendar days, but less than sixty calendar days, during a calendar year; or,

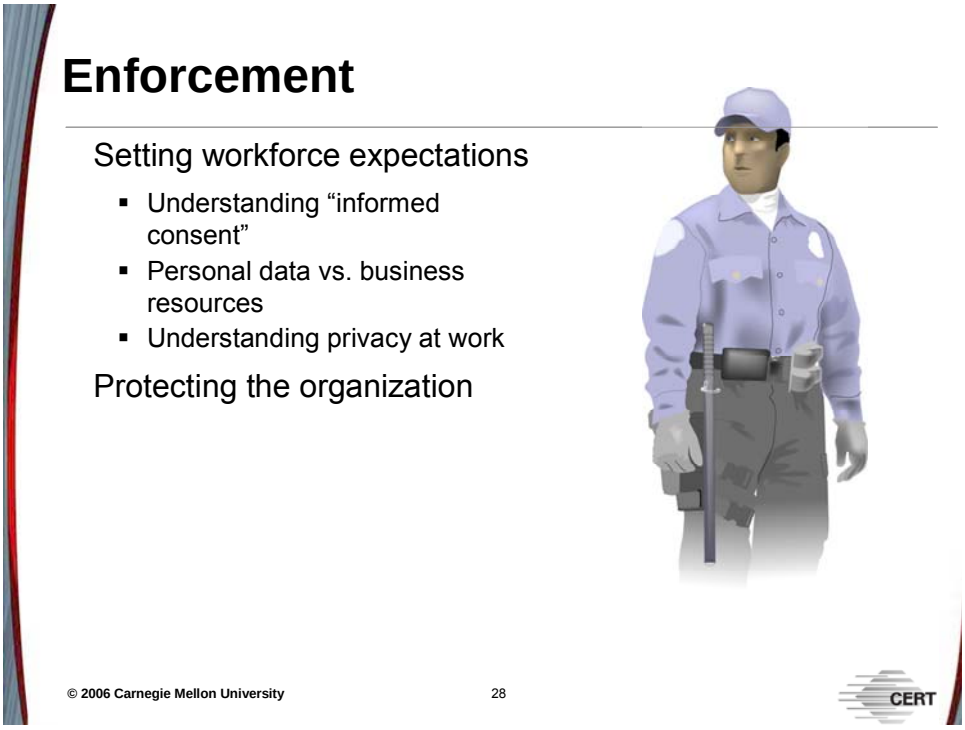
2.2.2 Exceptions to policy are not documented or authorized by the senior manager or delegate(s); or,

2.2.3 Access privileges have not been reviewed within the last calendar year; or,

The above excerpt from the NERC Cyber Security Standard defines multiple levels of noncompliance.



The above excerpt from the NERC Cyber Security Standard defines required documentation regarding individuals responsible for security and formalizes a compliance monitoring process.



2.6 Compliance Efforts

2.6.1 Enforcement

We have noted that compliance management involves making sure reality matches policy. The first step toward compliance, then, is having policies that make sense at every layer. It is counterproductive to spend energy on enforcing rules that do not support the mission of the organization. Therefore, policy development should flow first from the top down, and then from the bottom up.

2.6.1.1 Expectations of Workforce

Informed consent takes place when an employee signs a statement indicating that he or she understands an organization’s governing policies. Its purpose is to attest to the employee’s commitment to those policies.

The practice of requiring informed consent is prevalent in the healthcare industry, where it serves a different purpose. To limit the liability inherent in most medical procedures, medical professionals usually provide detailed counseling on all the possibilities of risk with respect to a given procedure. Armed with all the facts, the patient then takes responsibility for accepting or rejecting the procedure. The patient cannot, after the fact, claim negligence on the part of the doctor if the outcome is undesired but previously published as a possible outcome and acknowledged by the patient’s signature.

In a similar manner, IT managers require the informed consent of those using the organization's IT infrastructure. Users must be given all the facts about what is expected of them, and just as importantly, they must be aware of the consequences of violating those requirements. They acknowledge this awareness via signature. When a user's activity violates the published policies, he or she cannot then claim ignorance of the policy or of the consequences that follow.

Nearly all organizations clarify in their policies the distinction between business resources and individual property. Most often, they clearly specify that business resources are for business uses only. This poses a problem when users add personal information to the computer or use it for personal tasks.

2.6.1.2 Privacy

The expectation of privacy in the workplace is a contentious topic. Information technology opens many possibilities for monitoring, recording, and auditing the use of information systems resources. What a user does with the computer on his or her desk can become readily known to the IT staff and managers. As a matter of policy, organizations should explain what information is gathered and to whom it is reported. It may be that a manager has access to all Web requests made by those whom he supervises. To prevent a suspicious workforce, the policy should indicate the purpose of gathering such information and should provide a clear definition of misuse.

2.6.2 Audits

A comprehensive information security process includes validation of compliance. As noted earlier, the policy life cycle includes regular review to ensure policies make sense in the current environment. In the same manner, compliance must be checked routinely to ensure policies are being implemented effectively.

2.6.2.1 Self-Compliance Check

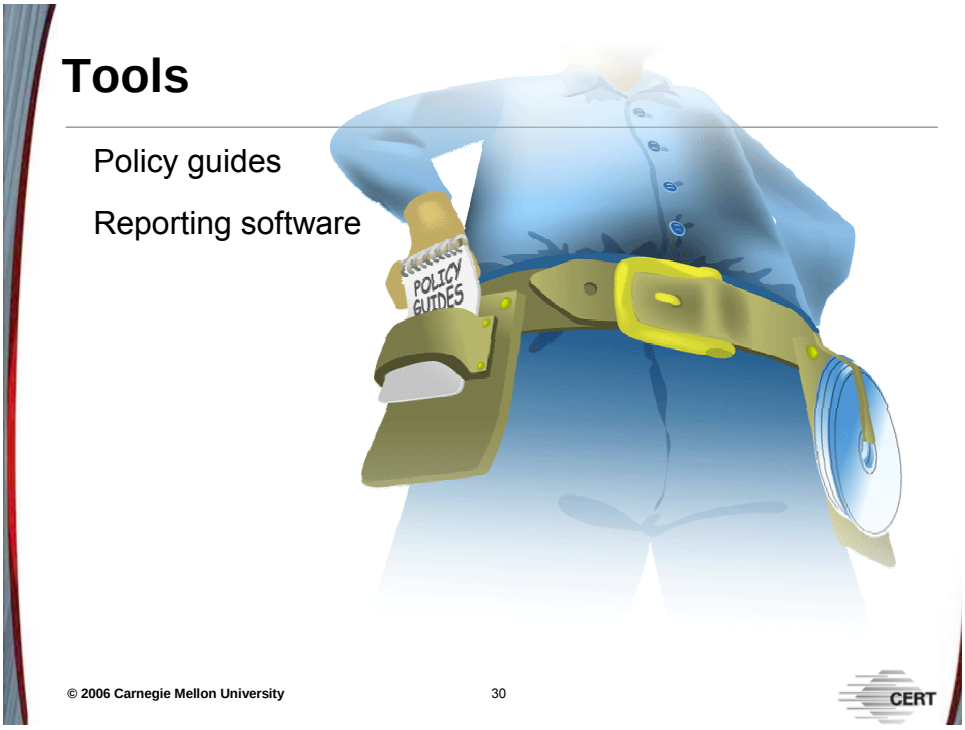
In some organizations, there may be little incentive to conduct a formal compliance audit. Nevertheless, managers should continually watch and check for compliance, because good policy directs employee efforts down the most favorable path for the company. Thus, a company's success can largely be measured by the compliance of its members.

When policy requirements are enumerated clearly in written documentation, the document itself can be used as a compliance checklist.

2.6.2.2 External Audits

Many organizations, in light of the legislation discussed above, must prove compliance to a higher authority. In such cases, an organization's information security policy compliance may be scrutinized by an external, objective entity. As in a financial audit, the external auditors will report to the appropriate authorities on how well the organization adheres to its

own policies and how well its policies reflect the pertinent standards. Having a well-documented policy program is essential to earning a favorable audit report.



2.6.3 Tools

As organizational policy for information security becomes more ubiquitous and comprehensive, the resulting policy-related work volume increases as well. The processes of developing policy, mapping standards to policy, and tracking and reporting compliance each require their own dedicated information system. These tasks also are supported by a growing number of resources, as discussed in the following subsections.

2.6.3.1 Policy Guides

Many resources can aid in policy development. Rarely do organizations have time to start the security policy process from scratch. Such an approach involves first reading and understanding authoritative standards, which often involve complex legal and legislative language. Instead, it is often helpful to get a jump-start with prewritten policy templates. If these are too restrictive or canned, many companies provide services to help lead organizations through risk assessment and policy development to mitigate assessed risks.

2.6.3.2 Reporting Software

Commercial software built for specific standards-compliance tracking and reporting has become increasingly common. This can serve as another effective aid to organizations in their compliance efforts.

Review Questions

1. What act requires public corporations to audit their IT systems?
2. In what way does policy provide “protection?”
3. Why does policy need to be layered?
4. List three components of the policy process.
5. List several characteristics of well-written policy.

Module 3: Risk Management

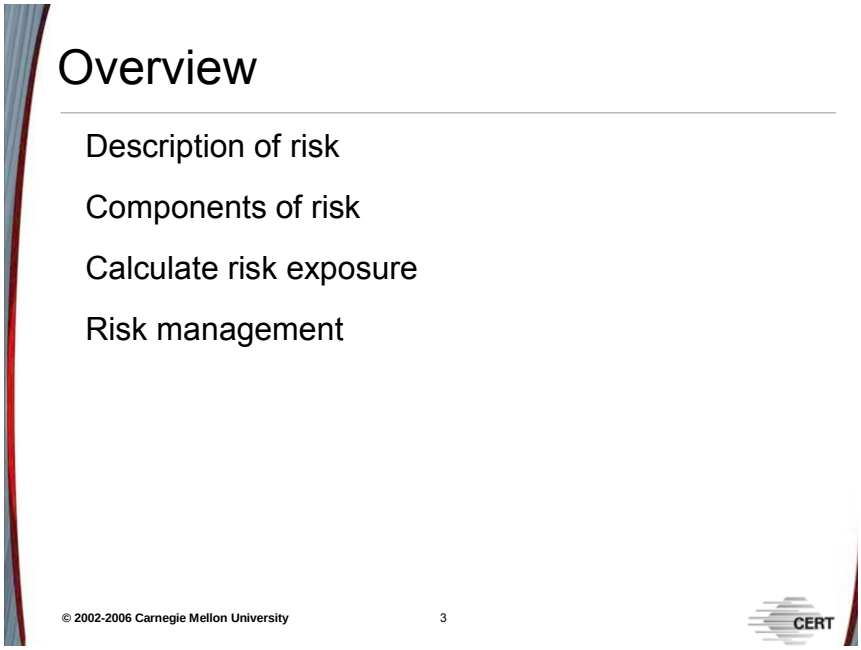
Instructional Objectives

Upon completion of this module, students will be able to

- Define
  - Risk
  - Components of risk
- Calculate risk exposure
- Describe common risk management strategies

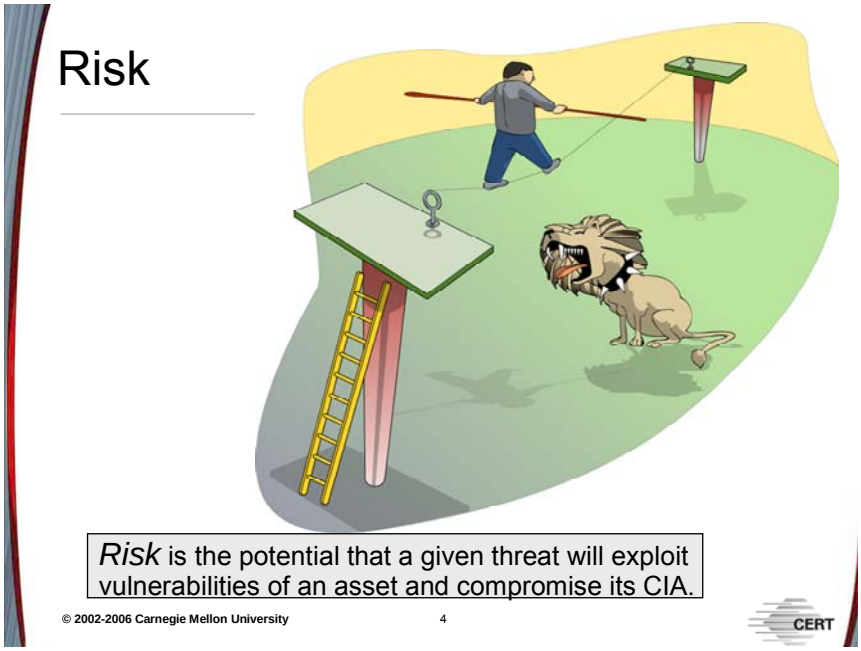
This instructional module will enable students to complete all of the above learning objectives.

3 Overview of Risk Management



The purpose of this module is to familiarize students with risk management. To do this, students first must understand the key attributes of risk and the concepts underlying risk management (including risk analysis assessment). One of these key attributes is the valuation and determination of assets. In most organizations, this process consists of identifying and prioritizing assets based on their value, cost, or importance. This module focuses on organizations that select and prioritize assets based on the assets' importance or relevance to fulfilling the mission and objectives of the organization.

This module also introduces concepts such as risk, risk impact, risk attributes, assets, asset categories, risk analysis, and risk management. Additionally, this section examines the application of risk management. This examination includes a description of risk assessment and analysis activities, comprehension of the impact of risk events, and recognition of mitigation strategies for managing and reducing risk.



3.1 Description of Risk

Risk is the possibility of suffering harm or loss. With respect to information and computer data, risk is the potential that a given threat will exploit vulnerabilities to compromise an asset's confidentiality, integrity, or availability.

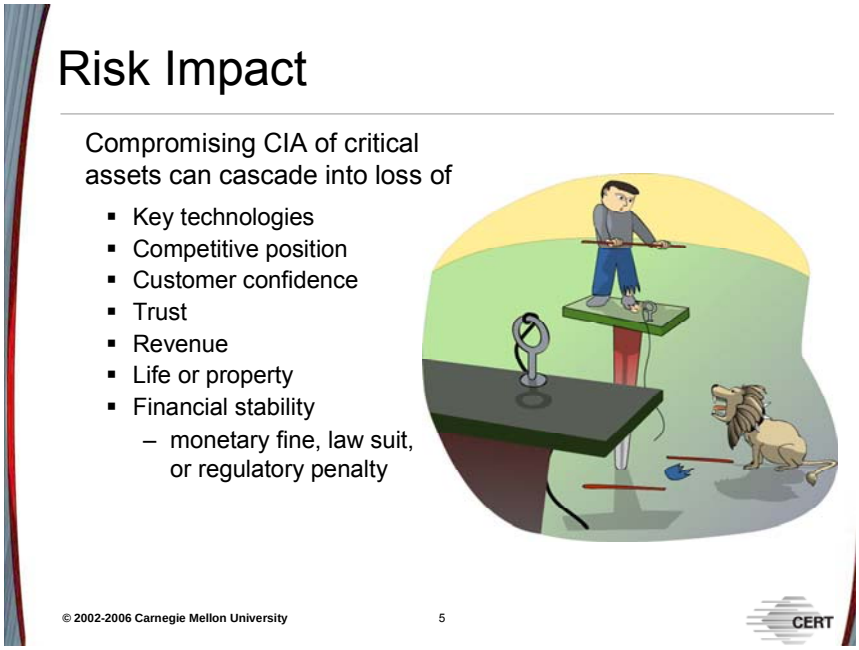
Before risks can be managed, they must be identified. One way to identify potential risk is to list the components of risk in an asset-driven scenario and gauge the risk's plausibility.

Example of Risk

For example, let's consider a home user making a consumer purchase over the Internet. In this situation, the user must submit customer information to the Web site (i.e., item, quantity, name of customer, address of customer, payment type, credit card number, etc.) to complete the purchase. Therefore, from the user's point of view (even if he or she is not explicitly aware of the risk or is seemingly unconcerned about it), risk exists. To identify the risk in this situation, we can state that the asset is the customer's information, the threat is anyone on the Internet with malicious intent, and the vulnerability is any technology weakness that allows the information to be observed and captured.

Threats and Vulnerabilities

Probably the most readily identifiable components of risk, to system and network administrators, are threat and vulnerability. The combination of threat and vulnerability yields a potential for undesirable outcomes, as when threats exploit vulnerabilities in an asset to compromise the asset's confidentiality, integrity, and/or availability. These undesirable outcomes are referred to as impacts.



3.1.1 Risk Impact

Understanding a risk's impact forms the basis for evaluating outcomes of risk: loss, destruction, modification, and interruption. Impact is the actualization of risk. To evaluate the outcome of a risk, we start by developing evaluation criteria for risk scenarios.

For example, consider a home user who sets up a personal Web server to display his or her resume. As a risk management process, our home user identifies his or her assets, considers possible negative outcomes, and characterizes the potential impact of an asset's compromise. Here, our home user recognizes that one asset is the Web server itself, while another is his or her resume. Possible negative outcomes to the resume asset include

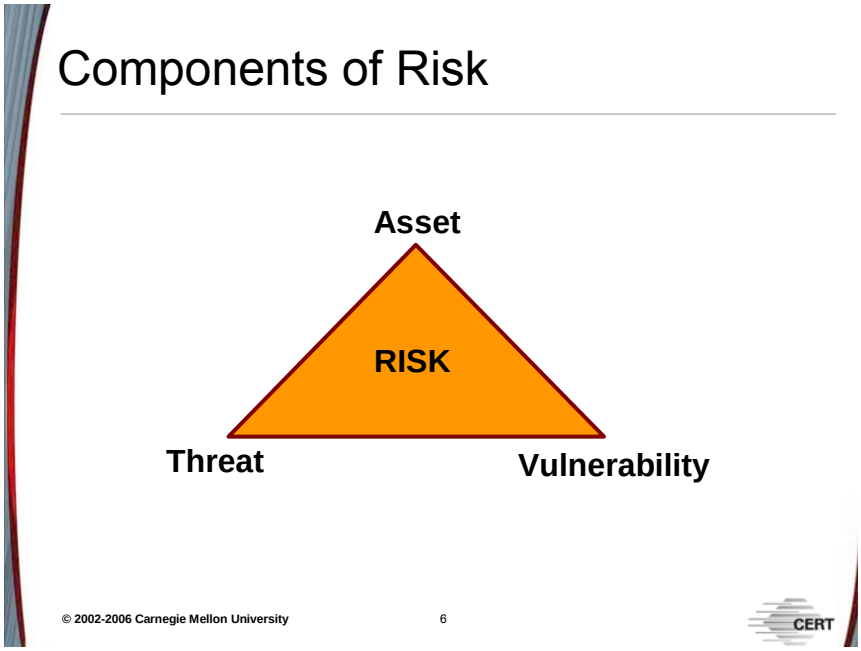
- Destruction of the resume file
- Modification of the resume content
- Pirating of the resume data
- Interruption of the resume's presentation to the Internet

After considering these potential negative impacts, our home user should be able to define potential failure conditions, such as

- Destruction, causing an expenditure of effort to restore or recover the information
- Modification, causing a prospective employer to consider the candidate either adequate or inadequate for the position, depending on what information was changed
- Theft, causing a loss of creative and competitive marketing of the individual's skills or background

- Interruption, causing an inability for potential employers to view the candidate's information

Finally, the home user must decide whether he or she really cares about the potential impacts to his or her assets. If yes, then mitigation strategies should be supplied; if no, then the potential risk impacts are accepted in addition to the consequences of loss or harm that could be suffered through exposure to the risk.



3.2 Components of Risk

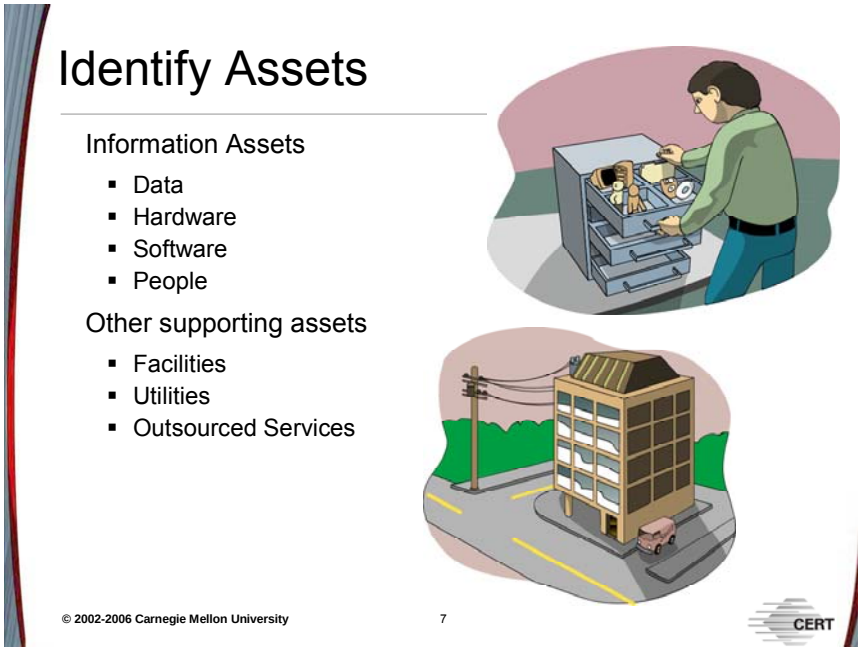
For a risk to exist, the following must be present:

- Assets of value to an organization or individual that must be protected (critical assets)
- Threats to these critical assets (possibility of disclosure, modification, destruction, or interruption)
- Vulnerabilities of these critical assets that may provide an opportunity for threats to act on the assets in a manner that discloses, modifies, destroys, or interrupts the assets

Identifying assets involves a discussion within the organization to determine what categories of assets exist, who owns each asset, and what level of protection is necessary for each asset. This exchange of information should take place between managers, staff, and information technology personnel on a periodic basis and as part of the organization's review of its information security policy. These events are important as a means of identifying assets and risk mitigation plans because they enable the organization to identify its current protection strategy for each asset and any changes to the asset's priority. This priority discussion allows a ranking of some assets over others, and it should be documented and reflected in organizational policy, recognizing that assets may be mission critical, non-critical but sensitive, or general in nature.

Relating threats and vulnerabilities to an asset is part of risk assessment and requires that those responsible for protecting information assets have an appreciation of the range of threats and vulnerabilities. Once the range is known, the likelihood of any one threat acting adversely on an asset must be understood.

The results of risk analysis identify the strategies (plans, policies, technological mechanisms) that can help mitigate the risk. Analysis includes evaluating the risk to an organization and measuring that risk against the impact to the organization if the risk is realized. For example, a determined risk for a medical organization may be that "modification of paper medical records by unauthorized individuals can lead to loss of life, financial or punitive penalties, or loss of customer confidence." In this case, this risk is actually stated as a risk scenario that includes assets (paper medical records), threat actors (personnel exceeding their privileges or unauthorized outsiders), outcome (modification of the records), and impact (public safety, financial, customer confidence, legal). Risk analysis determines which risks are viable (that is, non-negligible) and what degree of impact (high, medium, or low) the risk has on the organization when evaluated.



3.3 Identify Assets

An asset is anything of value to an organization. Typically, assets are classified as information assets (people, hardware, software, systems), other supporting assets (facilities, utilities, services), or critical assets. Critical assets may include information or other supporting assets. Later in this section, we will describe examples within each category.

It is important to note that your organization may choose to classify assets within different categories according to sensitivity or function. Asset definitions may be highly subjective, and asset value even more so; therefore, an easier way to approach assets and asset value can be to consider the worth of the asset (in both tangible and intangible terms) to the organization. By examining the costs associated with the value and intrinsic value of an asset (qualities of the asset's existence), you may discover a more meaningful definition and value of the asset.

Information includes

- data being processed on, stored on, or transmitted between systems
- backup and archive data (on-site and off-site storage volumes and locations)
- paper documents
- escrowed encryption keys
- software distribution media

Hardware includes

- desktop computers
- servers
- mainframes
- network equipment (routers, switches, firewalls)
- wiring infrastructure
- wireless support infrastructure

Software includes

- commercial off-the-shelf (cots) software
  - operating systems
  - desktop software
  - mainframe applications
- custom software
  - in-house effort
  - outsourced effort
  - ad hoc scripts
  - undocumented tools used by employees

People include

- senior and middle management
- technical and non-technical staff
- public relations
- help desk, facilities, security
- contractors, third parties (Computer Security Incidence Response Teams [CSIRTs])
- government, police, fire

Facilities include

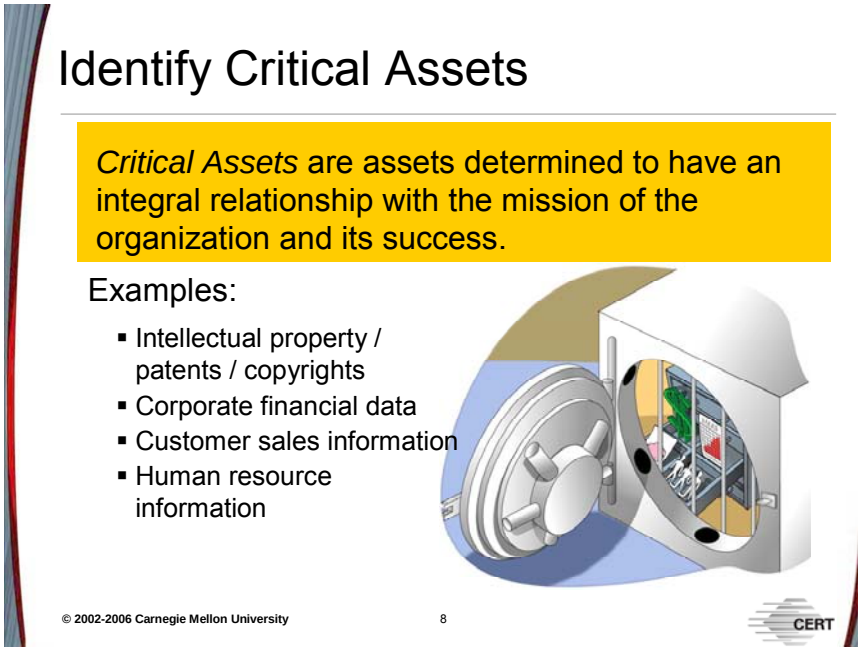
- heating/ventilation/air conditioning (hvac) support
- power
- water
- telephone
- security

Utilities include

- power
- water
- telephone
  - leased lines (t1, t3, isdn)
  - voice lines
  - cell phones
- pager services
- service-level agreements
  - hardware maintenance
  - HVAC support

Outsourced services include

- off-site services
  - information storage
  - Web services
- consultants
- utilities
- legal services
- public relations
- managed or monitored security
  - physical
  - network



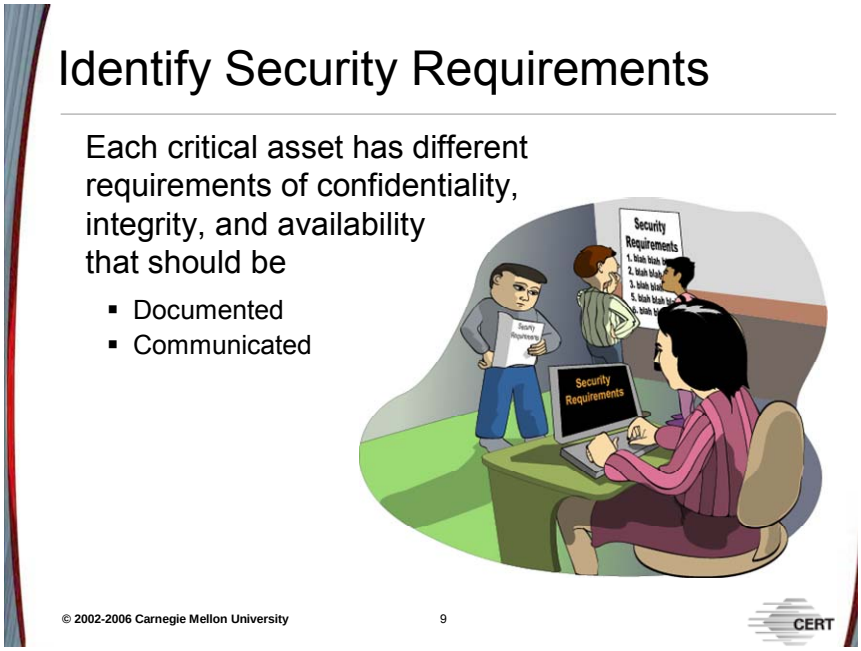
3.3.1 Identify Critical Assets

Critical assets are assets that have an integral relationship with the mission of the organization. This means that loss or damage to a critical asset would cause disruption to the operational or functional mission of the organization to a point where the mission fails. This concept recognizes that each individual organization will define a different and unique set of critical assets that align with mission success or failure.

Examples of critical assets include

- intellectual property
  - patents, copyrights
  - software code under development
  - systems acquisition or development projects
- corporate financial data
  - payroll information by employee, department, organization
  - financial earnings, revenue, and loss statements
  - stock dividend information
- customer sales information
  - names, addresses, credit card / account numbers, purchase histories, demographic information
- human resource information
  - names of employees, departments, salaries
  - hiring, administrative punishment, and disability information
- network architecture information
  - network topology diagrams
  - desktop or systems replacement plans
  - strategic infrastructure plans
  - vulnerability assessment reports
  - types and locations of infrastructure (general purpose, storage, server, networking, and security devices)

- U.S. Government or military classified information
  - compartmentalized projects
  - deployment and strategic plans
  - intelligence information, logistic movements/support
  - technical specifications on equipment, weapons, projects

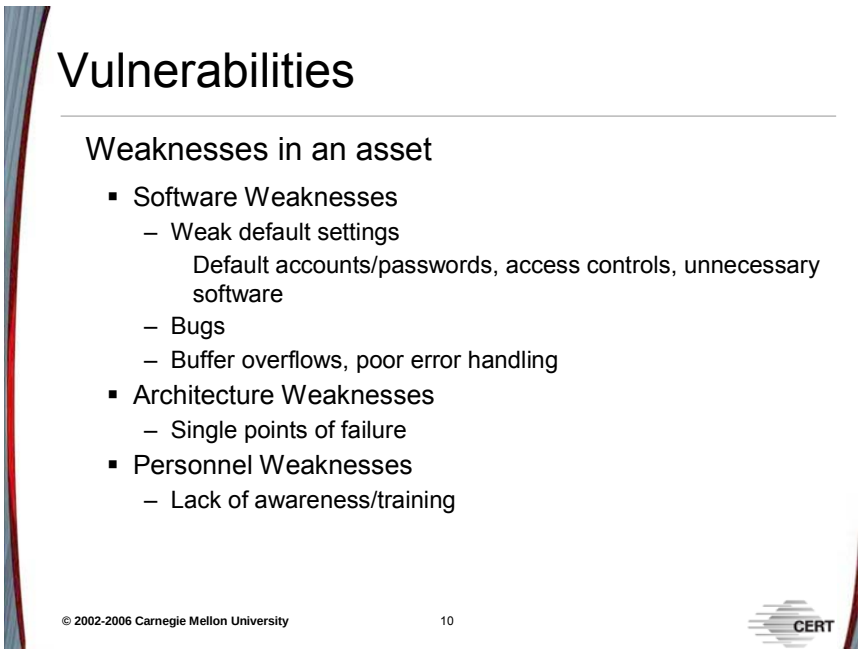


3.3.2 Identify Security Requirements

Each critical asset has different requirements for *confidentiality*, *integrity*, and *availability* that should be

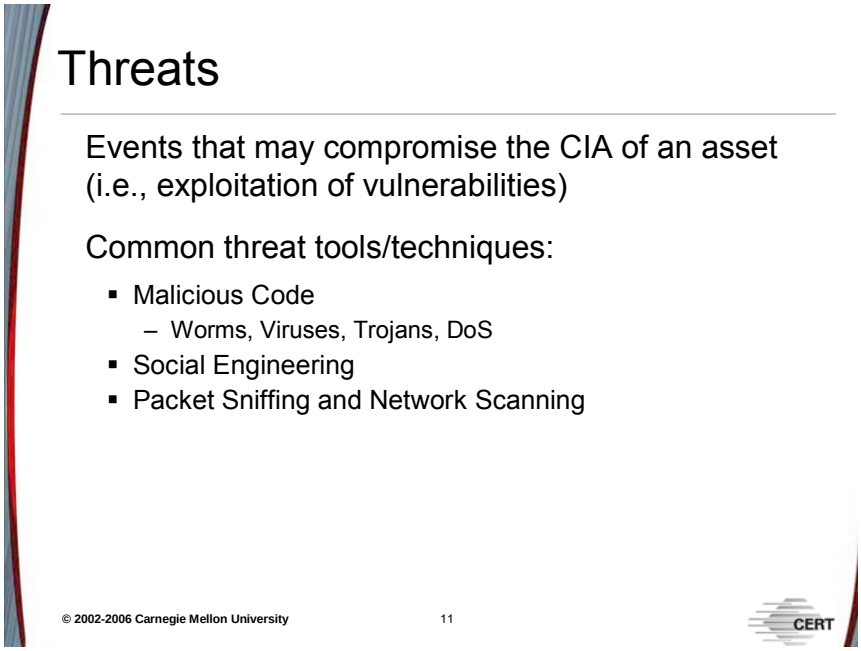
- documented, describing the requirements, the responsible information/asset owner(s), and the party charged with the asset's protection, as well as under what conditions and to what degree the requirements must be enforced
- communicated throughout the organization, especially from the owner of the asset to the person(s) responsible for its safety and security (the information/asset custodian)

Security requirements should be understood at all levels of the organization involved in the asset's protection. They should be described with enough detail for a specific requirement to be placed on the responsible owner (manager, user, system/security administrator, etc.) or the technology protecting the asset. They should be documented in security policies and plans.



3.3.3 Vulnerabilities

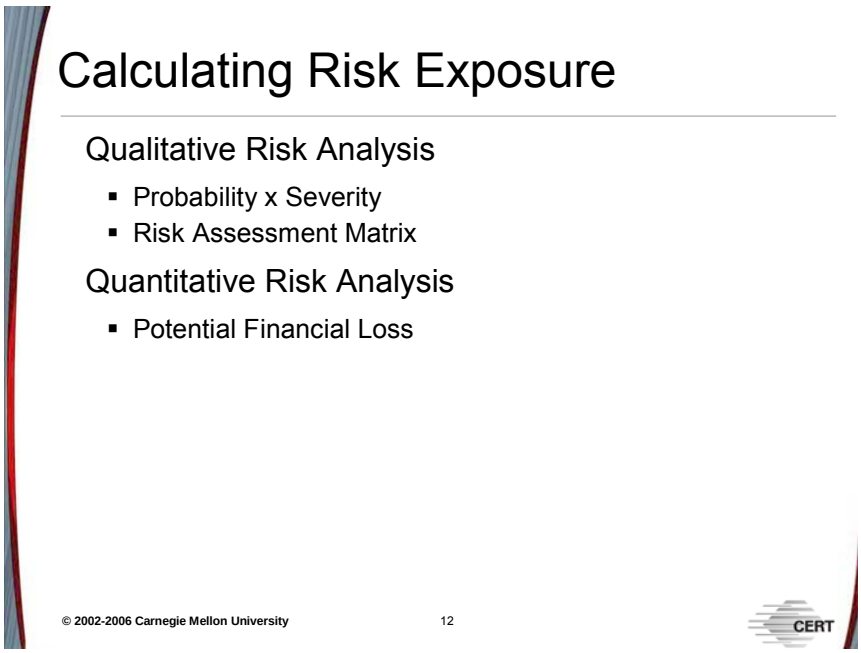
Vulnerability is the absence or weakness of a safeguard. It can also be described as a weakness in an asset or in the methods of ensuring that the asset is survivable. The examples of vulnerabilities listed on this slide provide a small sampling of the numerous classes of vulnerabilities that commonly exist.



3.3.4 Threats

A threat is any event that will cause an undesirable impact or loss to an organization if it occurs. Examples of threats include the following:

- intrusions into and disruptions of information systems
  - viruses, worms, and Trojan horses
  - denials of service
  - sniffing network traffic
  - stealing data assets
- loss of assets that are single points of failure
  - critical data that is not backed up
  - a single, critical piece of network infrastructure (i.e., a core router)
- keys that are used to encrypt critical data



3.4 Calculating Risk Exposure

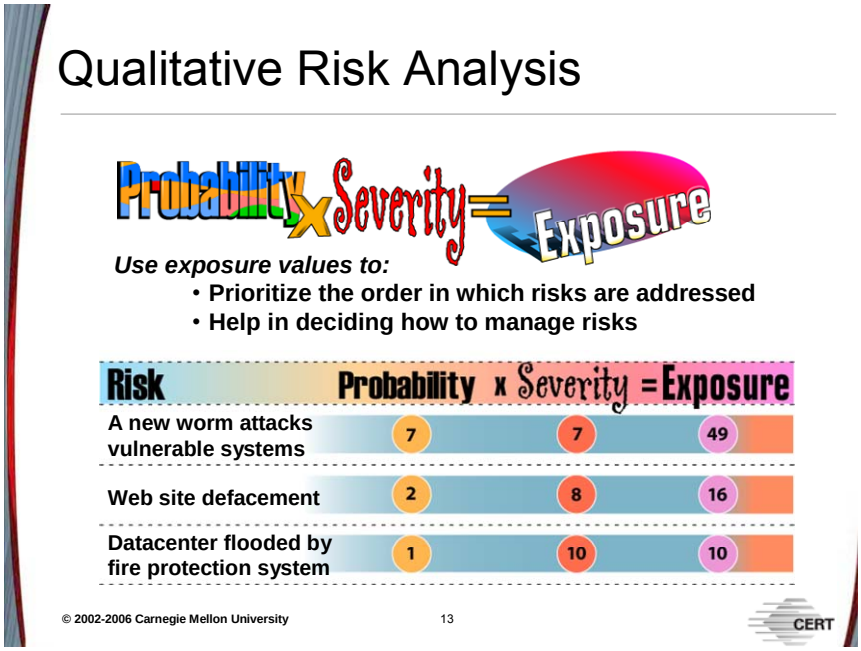
Risk analysis is the process of identifying security risks, determining their magnitude, and identifying areas in need of safeguards. Risks are traditionally captured as a description that can then be measured both qualitatively and quantitatively. Qualifying a risk means understanding the potential negative impact with respect to the asset as well as the likelihood of the threat. This impact occurs when the asset is destroyed, modified, interrupted, or disclosed. To home users, qualifying risk often means evaluating the impact of having their personal information disclosed. In this case, the users will probably be most concerned with their financial liability, chance of identity loss, and laws and regulations to which they may be subjected, as established by a qualitative scale (or criteria) for evaluating the risk (such as high, medium, or low).

Quantifying risk means understanding the possibility of the risk existing or coming to fruition. Here the home user attempts to measure the probability or likelihood of someone performing several different attacks whose goals are to retrieve his or her personal information. This measurement takes into account how likely it is that

- someone may observe the information in transit between the home user and the Web site (and possibly decode the encrypted network traffic)
- the software making the exchange of personal information is vulnerable to attack
- the user will be singled out for exploitation over all of the other Web commerce transactions happening at the Web site of purchase
- an attacker might gain access to the information once it has arrived at the Web site

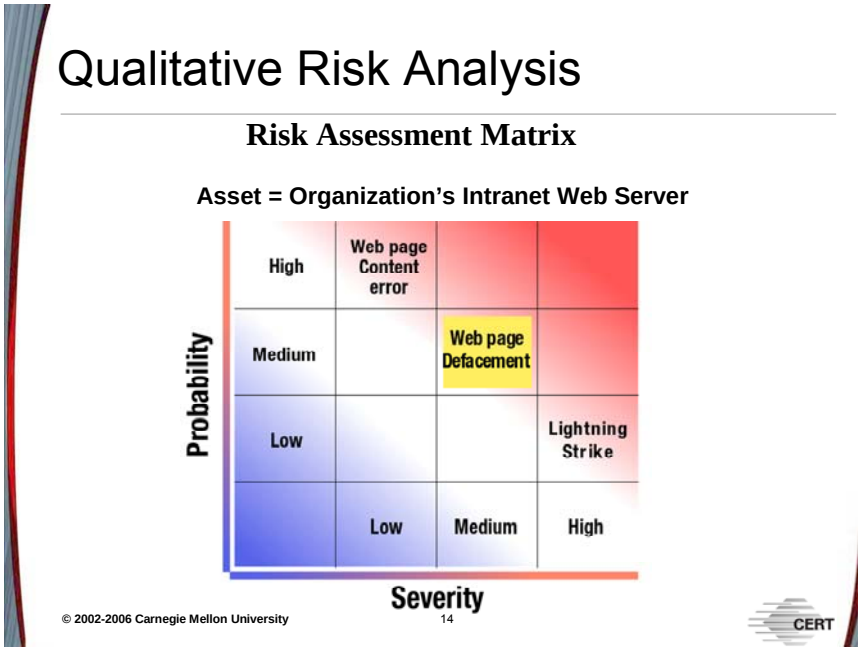
These are just a small sample of the risks involved in this simple transaction. For the individual to really understand these risks, he or she must appreciate the potential impact of these risks. This demands an understanding of the potential that threat sources (humans, system problems, viruses, etc.) have in exploiting and abusing vulnerabilities that result in risk. This potential falls into a continuum ranging from negligible to actual, over the life of the information being transmitted, stored, and processed.

Quantitative risk analysis can be a major project and can consume considerable organizational resources. It attempts to assign independently objective numeric values (hard dollars, for example) to the components of risk assessment and to the assessment of potential losses. Qualitative risk analysis addresses more intangible values of loss and typically attempts to produce scenarios so risk can be anticipated and managed. However, threat frequency and impact data is still required to conduct a qualitative risk analysis.



3.4.1 Qualitative Risk Analysis

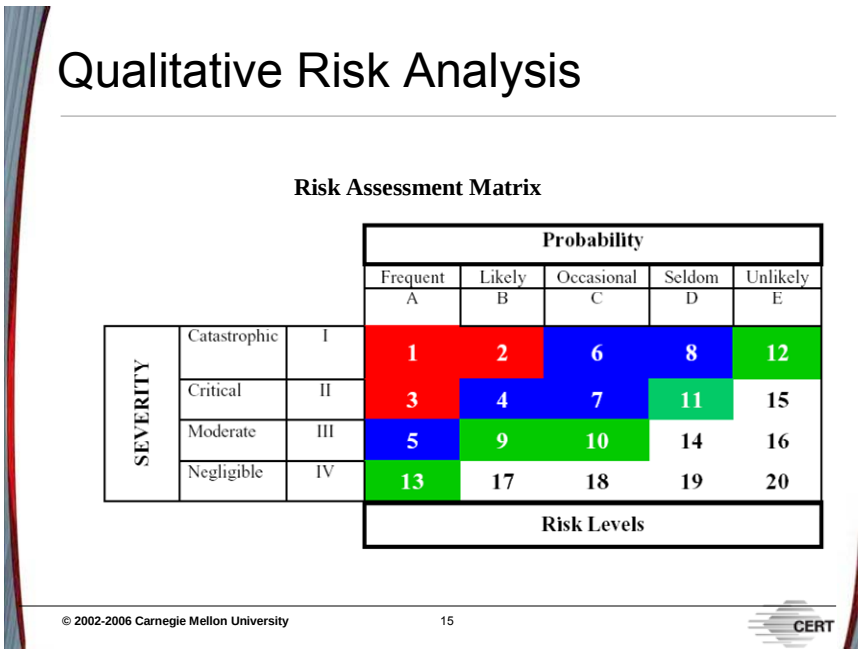
This slide shows a simple exposure table that supports a qualitative risk analysis. Threat scenarios are described for assets (typically critical assets), and data from the exposure table is used for making decisions regarding risk management. The table also provides a starting point for determining which risks are of greatest concern when it comes to mission survivability. If a decision is made to mitigate the risk, typically a cost/benefit analysis is conducted to select safeguards.



Simple Risk Assessment Matrix

Even simpler than the exposure table is the risk assessment matrix. It simply categorizes threats into levels of degree based upon probability vs. severity.

If a threat is in the High/High box in the matrix (high probability and high severity), an organization is likely to manage the risk associated with that threat first.

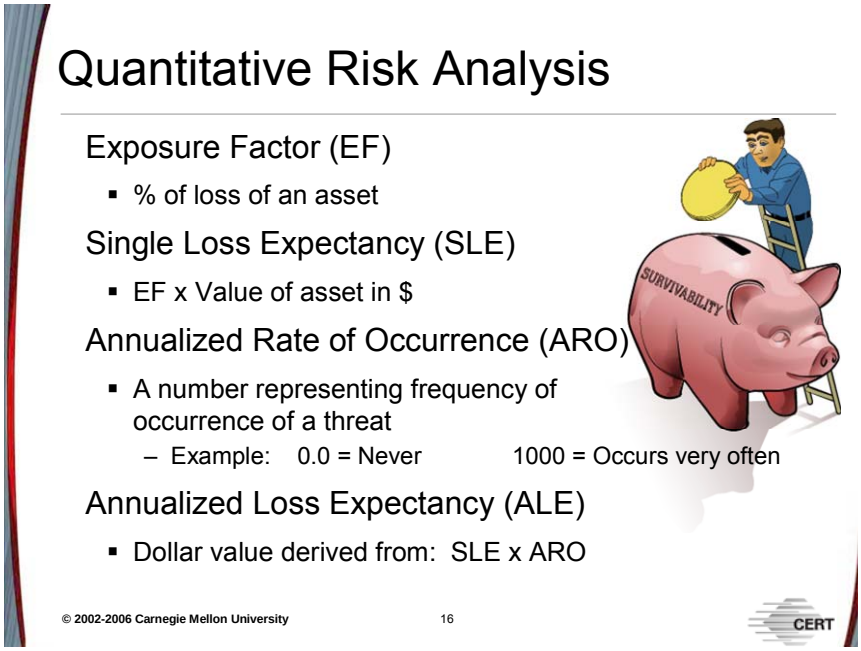


Detailed Risk Assessment Matrix

The above slide shows a more detailed risk assessment matrix (again based on the factors of probability and severity) that can be used when making risk management decisions. Here, threats of varying probability are categorized according to four levels of severity:

- Catastrophic – Complete mission failure
- Critical – Major mission degradation
- Moderate – Minor mission degradation
- Negligible – Less than minor mission degradation

The lower the risk level rating number, the more critical the risk is to the asset.



3.4.2 Quantitative Risk Analysis

Managers in IT are often faced with the dilemma of justifying their expenditures on survivability and security. Ideally, resources allocated toward survivability should be seen as an investment in the mission of the organization. But because the old paradigm (security seen as an overhead expense) is still an operational reality, IT managers often justify expenditures with forms of quantitative risk analysis. The terms shown on this slide are pseudo-standards that can help calculate risk in relation to actual dollar figures. Their usage helps to provide more reliable cost-benefit analysis.

- Exposure Factor (EF)
  The exposure factor describes the effects a realized threat would have on a particular asset as a percentage of loss of the total value of the asset. For example, loss of some hardware would have a small EF, whereas the catastrophic loss of all computing resources would have a large EF. The EF value is necessary to compute the Single Loss Expectancy (SLE), which in turn is necessary to compute the Annualized Loss Expectancy (ALE).
- Single Loss Expectancy (SLE)
  The single loss expectancy is the dollar figure assigned to an organization's loss from a single threat event. It is derived from the formula EF X asset value in dollars = SLE. For example, an asset valued at $10,000 that is subjected to an EF of 50 percent would yield an SLE of $5,000.
- Annualized Rate of Occurrence (ARO)
  The annualized rate of occurrence is a number that represents the estimated frequency with which a threat is expected to occur. This value can range from 0.0 (for threats that never occur) to a large number (for threats that occur frequently, such as misspellings of

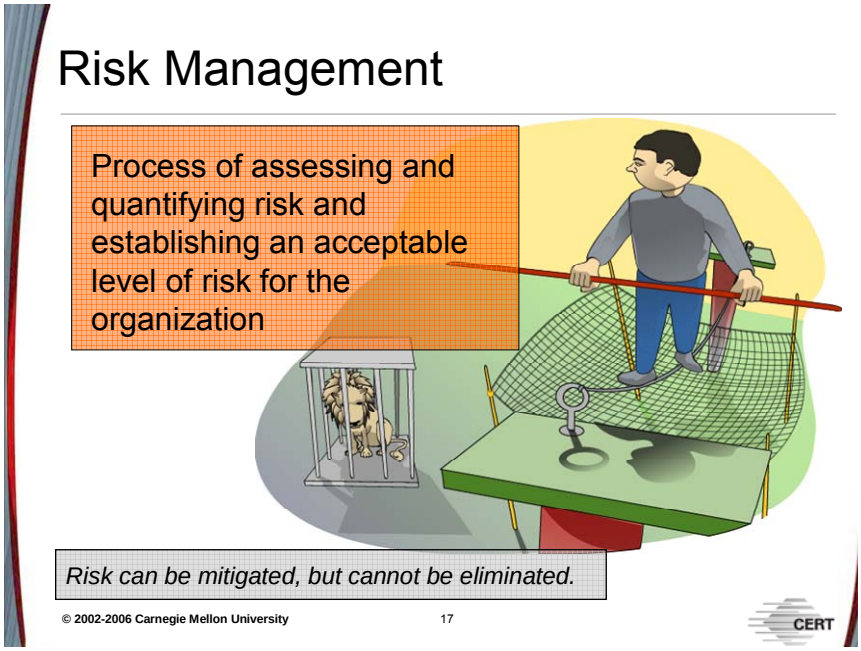
names in data entry). This number is usually created based on the likelihood that the threat will occur and the number of individuals that could cause it to occur. The loss incurred by this event is not a concern here, only how often it occurs. For example, flooding of an organization's data center by the fire control system could be estimated to occur once every 1000 years and will thus have an ARO of .001. However, the expected frequency of 100 help desk analysts making access control errors when administering accounts could be estimated at 15 times per year, resulting in an ARO of 1500.

- Annualized Loss Expectancy (ALE)
  Annualized loss expectancy (ALE) is the annual financial loss an organization expects from a threat. It is calculated by multiplying the single loss expectancy and annualized rate of occurrence (SLE X ARO = ALE). For example, a threat with a dollar value of $10,000 (SLE) that is expected to occur 5 times per year (ARO) will result in an ALE of $50,000 [Krutz 2001].

Generally speaking, if an organization's information survivability expenditures are less than the sum of the calculated ALEs, than some quantitative return on investment (ROI) figures can be discerned.

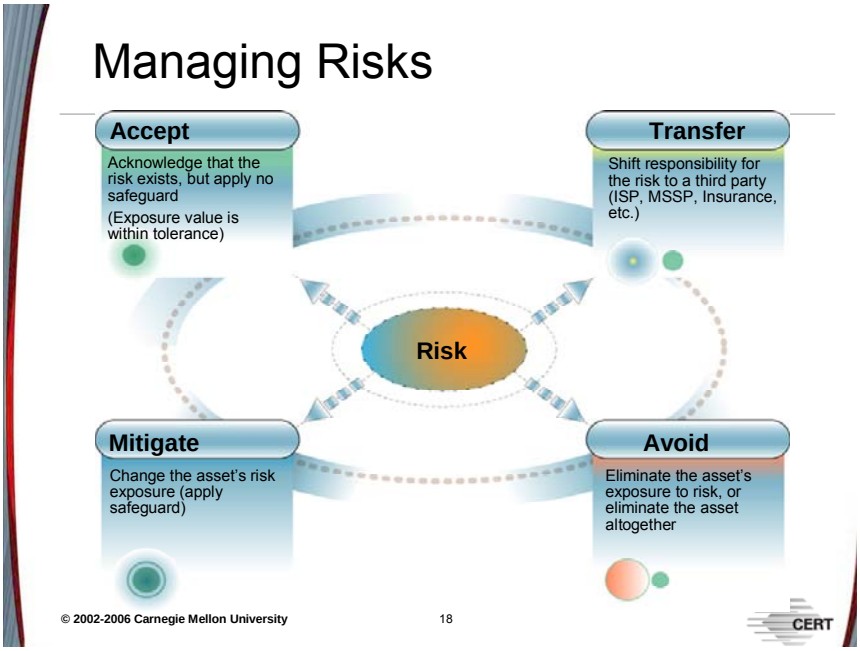
Summary of the Assessment Step

Risk is the probability and severity of loss from exposure to a threat. The assessment step involves the application of quantitative or qualitative measures to determine the level of risk associated with a specific threat. Specifically, this process evaluates the probability and severity of an undesirable event that could result from the threat.



3.5 Risk Management

Risk management should be a well-defined and established process. Effective risk management can save resources, reduce mishaps, and even save lives.



3.5.1 Risk Management Strategies

Organizations have four options when deciding how to manage risks:

1. Accept

If an organization chooses to accept risk, it does so with full knowledge of the potential threats and vulnerabilities to the asset. It may be that the asset's exposure is acceptable or within some level tolerance. For example, an organization recognizes the threat that usernames and passwords could be compromised by administering systems remotely with telnet, a protocol that transmits data, including passwords, in plaintext. However, the organization decides the risk is not great enough to warrant a safeguard (i.e., encrypted sessions with SSH or IPSec).

2. Mitigate

Mitigating risk is the process of actively applying safeguards to reduce an asset's level of exposure. In the above telnet example, the organization could mitigate the risk by (a) denying all management traffic to the remote systems that is not encrypted and authenticated, and (b) writing an organizational policy that acts as a deterrent (i.e., any attempt to compromise access controls on organizational systems will be met with stiff disciplinary action).

3. Transfer

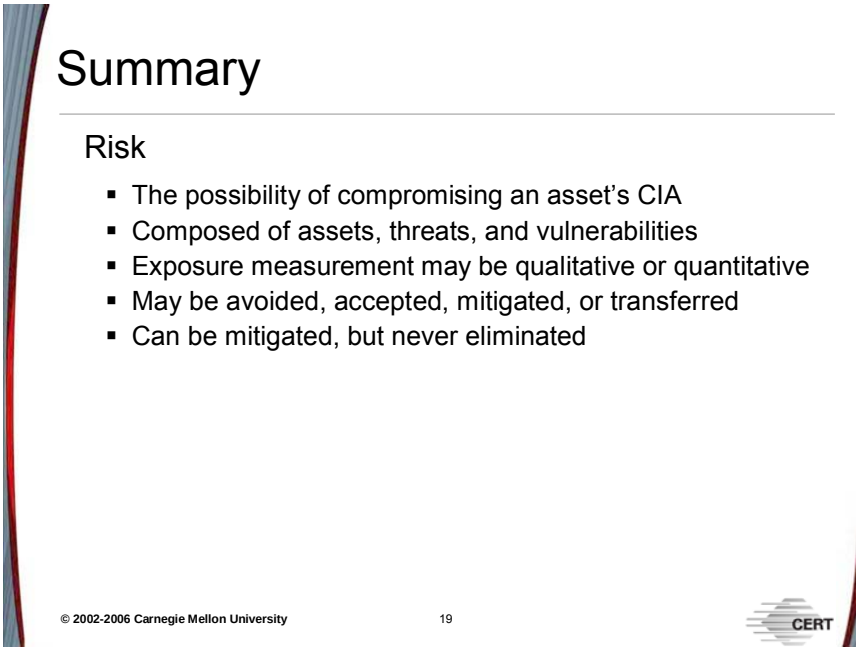
Transfer of risk occurs when an organization decides to contract with a third party to mitigate the risk. For example, an organization can transfer the risk of losing data (and support the goal of mission survivability) by contracting with a service provider that maintains an off-site data backup and recovery capability. Although risk transference does not change the probability or severity of a threat, it may decrease the probability or severity of the

organization's risk. At a minimum, the organization's risk is greatly decreased or eliminated because the possible losses or costs are shifted to another entity.

4. Avoid

Avoiding risk means that the organization eliminates the asset's exposure or even the asset itself. An example might be the replacement of historically vulnerable platforms (like Internet Information Server) with hardened platforms like a Bastille/Apache Web server solution [Bastille 06].[4]

[4] See http://www.bastille-linux.org/.



3.6 Summary

Sustaining and improving information security is a continuous risk management activity. Risk is comprised of assets (something of value to the organization), threats (concerns related to undesirable outcomes), safeguards, and vulnerabilities (weaknesses creating the possibility for threats to negatively impact the organization).

Risk analysis, a major component of risk assessment, helps to identify the possibility of certain risks and the impact when risks are realized. Because information cannot be realistically managed to have no risk, at some point risk must be accepted.

Review Questions

1. What are the components of risk?
2. Why do we prioritize one asset over another?
3. What two properties are analyzed and calculated as part of a qualitative risk assessment?
4. What are the four options available in managing risk?

Module 4: Identity Management

This module discusses methods of authentication and protection of identity and privacy.

Instructional Objectives

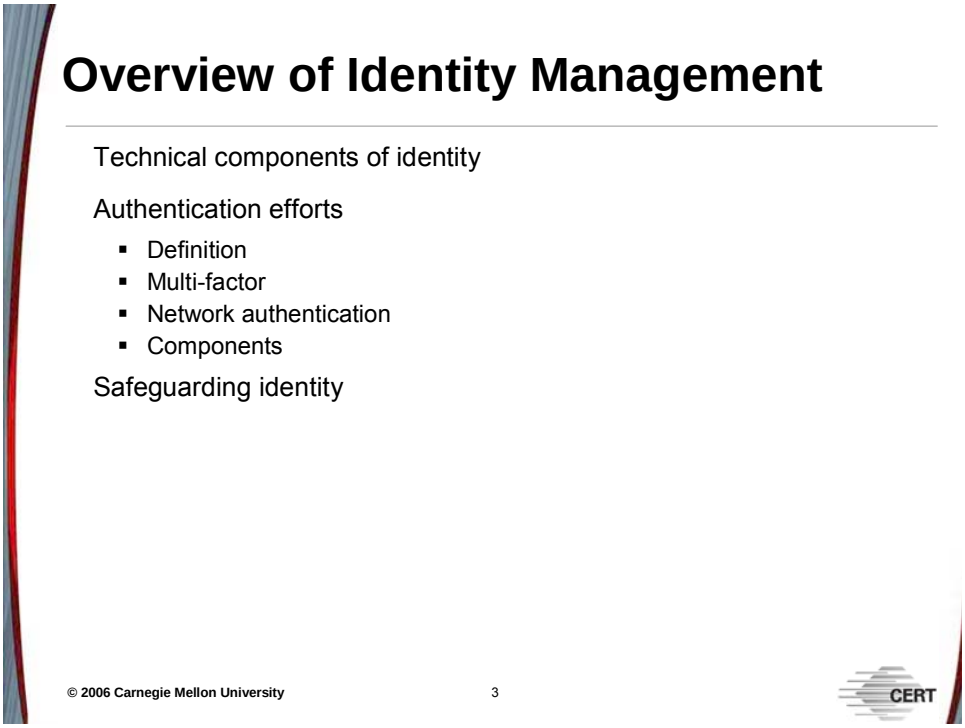
After completion of this module, students will be able to

- List the common identity credentials
- Define authentication
- Describe multi-factor authentication
- Compare symmetric vs. asymmetric keys
- Survey components of authentication systems
- Understand basic identity protection measures

CERT

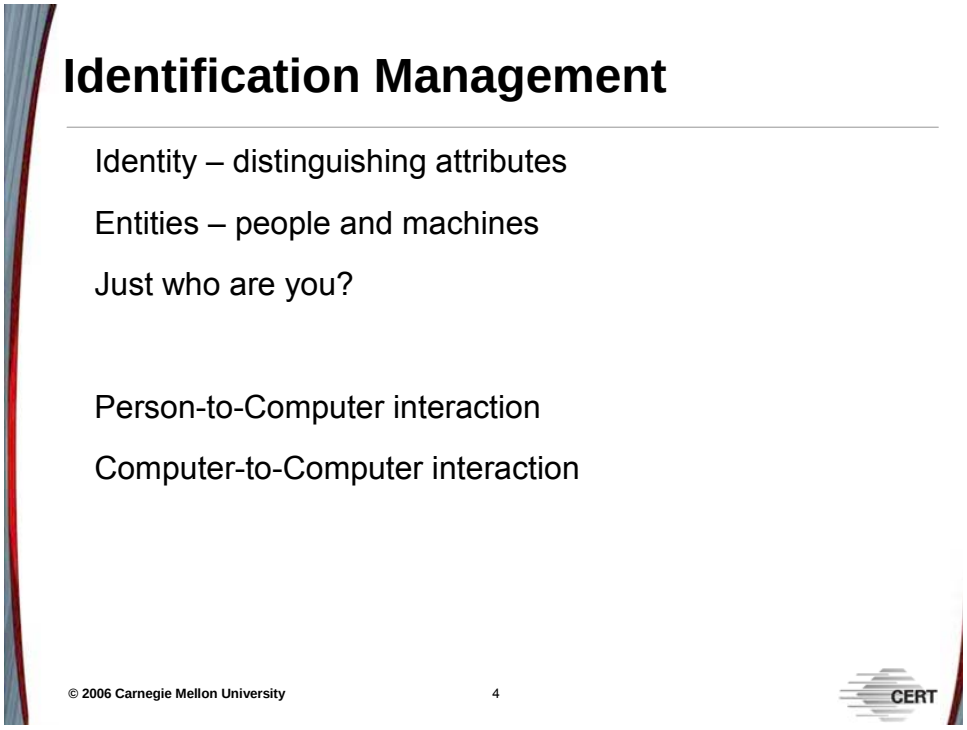
This instructional module will enable students to complete all of the above learning objectives.

4 Overview of Identity Management



As one of the eight major components of Defense-in-Depth, Identity Management is a critical aspect of any comprehensive information security effort. It primarily deals with *authentication*—determining whether a user of a computing system is in fact the user he or she claims to be. Because different users have different roles and responsibilities, systems must be able to distinguish one user from another with a high degree of accuracy.

The concept of *identity* is challenging to model in the digital world. An identity comprises the characteristics of a person that allow one to distinguish that person from any other person. In the early days of computing, simply entering a username was deemed sufficient proof of identity, and the user's claim of identity generally was accepted without question. It quickly became apparent, however, that people could take advantage of this situation to gain access to and control of resources that did not belong to them. Therefore, it became necessary to authenticate users, and more complex identity management processes were born.

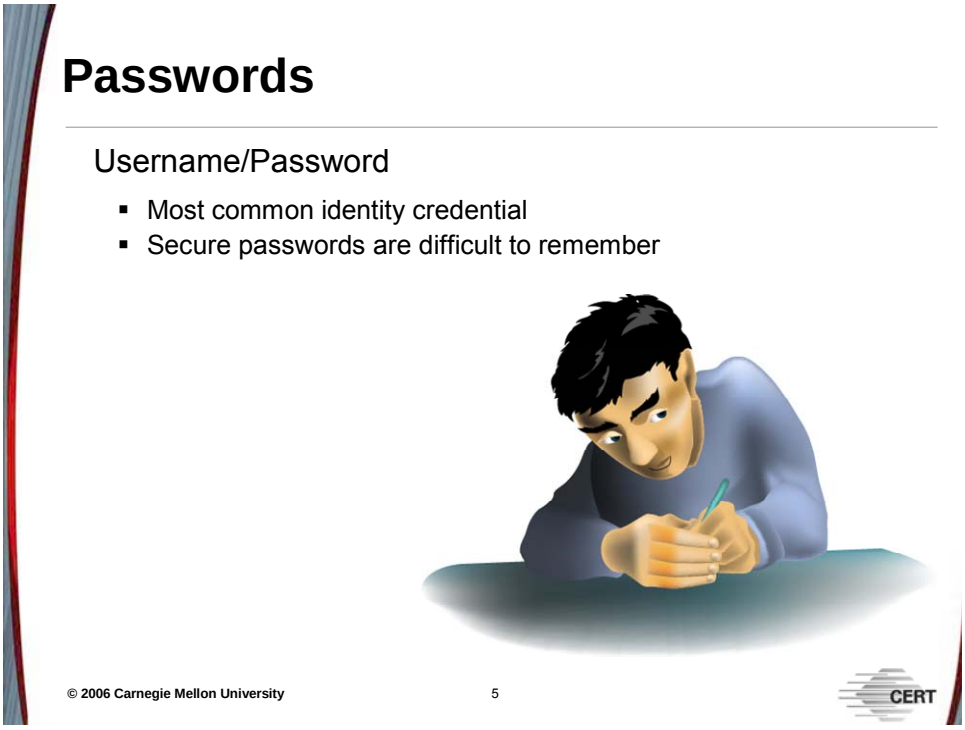


4.1 Identity of Technical Components

A word about the term *user*: The word *entity* is often a better description for the bearer of a digital identity because users are not the only participants in the authentication process. Just as often, computing services such as Web servers need to be identified, because users are keenly interested in authenticating the e-commerce Web servers with which they communicate and conduct financial transactions. Without authentication, they would have no assurance regarding to whom they were submitting their credit card information.

In this module, we will explore methods of digital identification and examine how they can be used on a computer, within a network, and across the Internet. We will look at person-to-computer and computer-to–computer identity management. Further, we will discuss aspects of privacy and identity protection.

Keep in mind as we discuss various methods of identity management that organizations requiring different levels of security will choose different methods. For businesses dealing with information that is rarely sensitive, usernames and passwords may be sufficient. Other organizations may need to employ a combination of sophisticated biometrics and cryptography. The vast majority of organizations will fall somewhere between these two extremes. Think about your organization's place along this continuum as you proceed with this module.



4.1.1 Passwords

The most common form of identification today is a username/password pair. A *password* is a sequence of keyboard characters known only to its owner. It is valuable in the authentication process because the computer system knows the user's password; so when the user claims to be *A*, the fact that the user knows *A*'s password suggests that the user is in fact *A*.

Username/password pairs are so widely used because users generally can pick passwords that are easy for them to remember but hard for others to guess. This method of identity management also requires the least amount of administrative work on the front end.

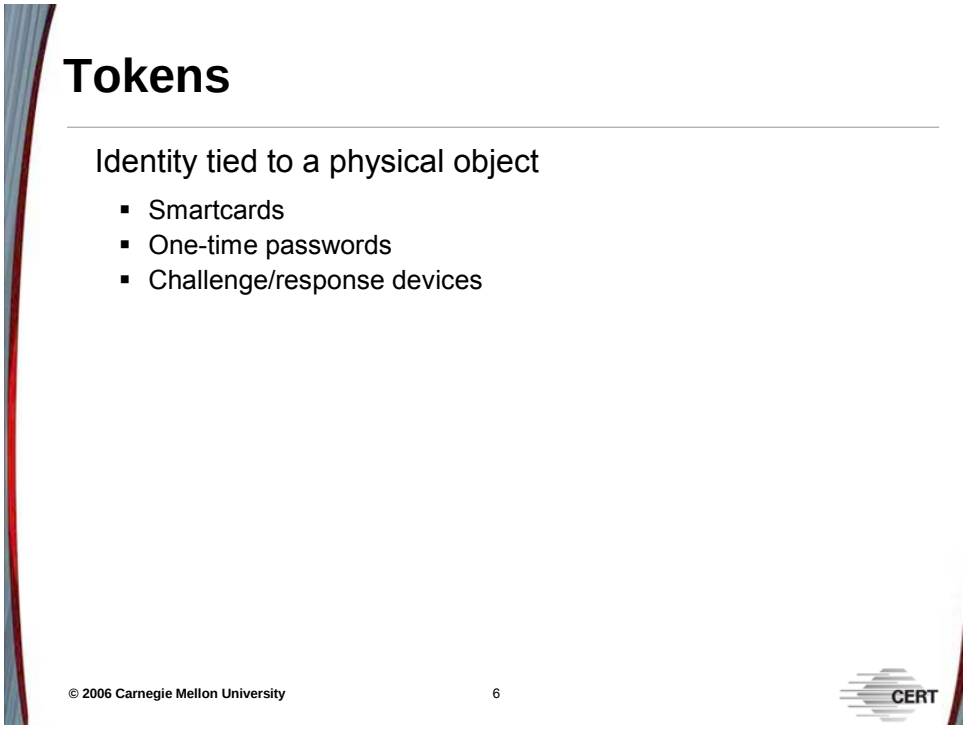
However, passwords also pose some problems. Any password that is simply a dictionary word, or a slight modification of one, provides little security against an attacker who tries to compromise it. Instead, it is necessary to build passwords that are long sequences of letters, numbers, and symbols. But truly robust passwords are often difficult to remember. They actually can have an adverse impact on security because users are likely to write the password on paper or store it somewhere else that is not secure.

There is also an administrative cost to passwords that is often not realized at the outset. Although establishing an initial password is not labor intensive, maintaining passwords can be. Industry studies indicate at least 25 percent of help-desk technical-support issues involve password problems.

Lastly, it is important to note that in most modern applications, a password is not stored or transmitted as plain text. Rather, it is transformed into a different value by a hash function. A

hash function is a procedure that turns an arbitrary-length string of characters into a unique fixed-length value. The SHA-1 and MD-5 hashing algorithms are the most common today. With these algorithms, it is highly improbable that two different passwords will hash to the same resulting value.

For more information about passwords, see the Wikipedia entry at http://en.wikipedia.org/wiki/Password.



4.1.2 Tokens

Beyond being tied to something one *knows*, identity also can be tied to something one *has*. Generally, tokens are physical objects given to users for identification purposes. The most popular tokens today are smartcards. These cards contain the user's name and photograph, along with an electronic chip that stores identity information digitally. The cards are convenient to carry and often eliminate the need to remember passwords. They can be used for physical identification as well as electronic identification.

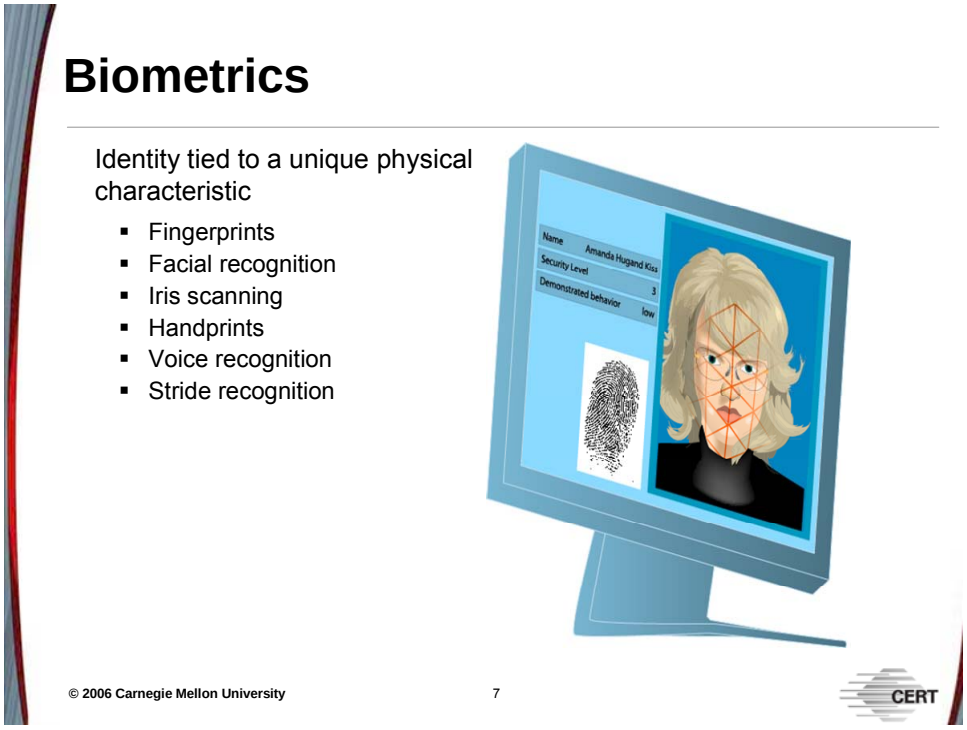
Smartcards are rapidly growing in use as more and more organizations implement card systems. Many U.S. federal agencies use such measures. The Department of Defense has successfully implemented a Common Card Access program, and other agencies are following suit. In 2004, the Department of Commerce was charged by a presidential directive (Homeland Security Presidential Directive 12[5]) with ensuring all federal agencies adopt a common credentialing system for employees and contractors who require physical or digital access to agency resources [NIST 06]. The Department of Homeland Security (DHS) plans to test the concept within the Transportation Security Administration by implementing the Transportation Worker Identity Credential program.

Much of the appeal of smartcards is the opportunity they provide to converge physical security with information security. As physical security controls become more reliant on electronic operation and administration, the two security fields are merging into a single realm. An industry consortium already has formed to focus on this issue of convergence.

[5] See http://csrc.nist.gov/publications/fips/fips201-1/FIPS-201-1-chng1.pdf.

The Open Security Exchange is a group of leading security organizations working to develop common standards and best practices for unified identity management across the physical and digital environments.

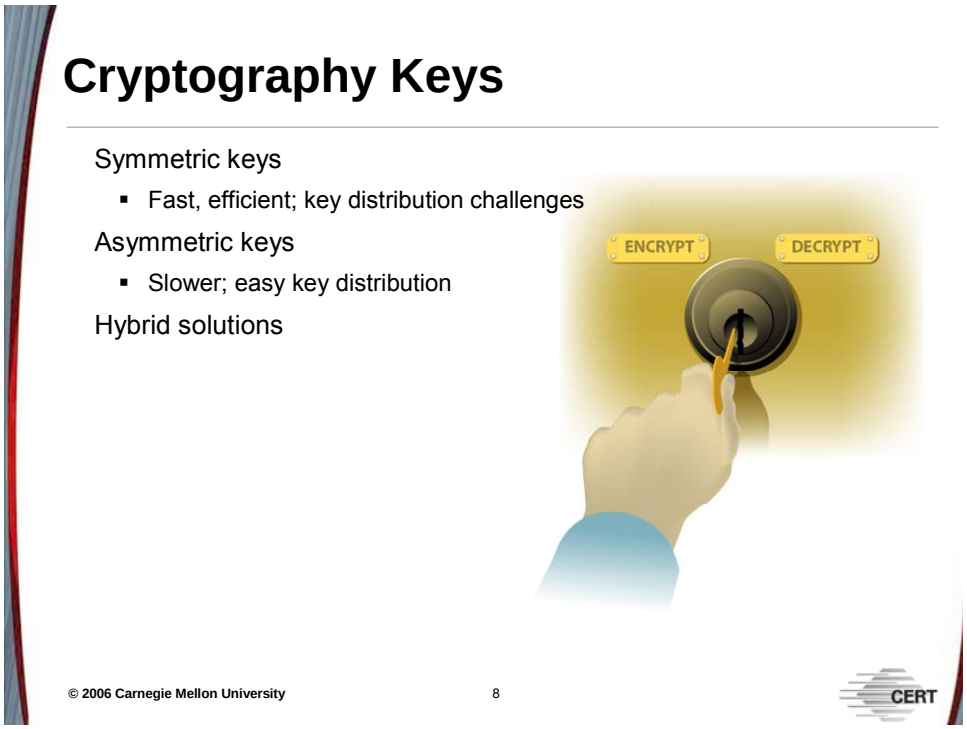
Other types of tokens also exist to provide digital identity management. These tokens usually are hardware devices. They may supply a stream of one-time passwords—passwords generated on the fly for users and then immediately deactivated after successful use—or they may provide responses to challenges posed by the system the user wants to access. One-time passwords are implemented in such a way that the authenticating system knows to expect certain passwords in a certain sequence from a certain user's device. Generally, these passwords are random enough to assure low probability of an attacker guessing the next password in the sequence. Challenge/response tokens are a means for a user to prove his or her identity by transforming a challenge code into a response that only the user with the token can provide.



4.1.3 Biometrics

Identity can be most effectively proved by matching certain physical characteristics of the user. Biometrics is the field of measuring uniquely distinguishing physical characteristics, such as fingerprints, facial features, and iris patterns, and storing digital representations of these characteristics to identify users. There are even proponents of using DNA to identify individuals. Unlike passwords, biometric characteristics are not easily lost or forgotten. A claim of identity supported by biometric evidence, therefore, can be more substantial than a claim supported by a token or password.

Biometrics are increasingly in use. Fingerprint readers especially are becoming more common as a means of verifying identity. There are low-cost devices available to provide this enhanced level of security. Some are embedded in notebook computers or can be added as peripheral devices, allowing users to log in to the computer with a single finger swipe. These login mechanisms can replace traditional username/password pairs and can be tied into central authentication servers in networked environments.



4.1.4 Cryptographic Keys

Cryptography provides a means for digital entities to prove their unique identity. It uses a *key*—a string of bits—in conjunction with published algorithms to encrypt and decrypt data. In effect, this key becomes like a password—something an entity knows. An entity can state his or her identity as the key holder and then back up that claim with use of the key.

There are two main types of keys: symmetric and asymmetric. Symmetric keys are keys that both encrypt and decrypt data. The same key is used on both ends of a data exchange by the sender and the receiver. A wide variety of algorithms use symmetric keys: Digital Encryption Standard (DES), Triple DES, Advanced Encryption Standard (AES), RC4, Blowfish, Twofish, and the International Data Encryption Algorithm (IDEA) are examples. The AES algorithm has superseded DES as the U.S. government's recommended algorithm. Symmetric-key algorithms have the property of being highly computationally efficient so that they run fast. This speed makes them suitable for encrypting large amounts of data. To do so, these algorithms use either a block cipher method, in which data is encrypted in "blocks" of a certain size, or a stream cipher, in which each plaintext character is encrypted one at a time. Picture a conveyor belt with several piles of letters poised to enter the encryption machine one pile at a time (a block cipher) versus a stream of letters laid out end-to-end on the belt and sent continuously through the encryption machine (a stream cipher).

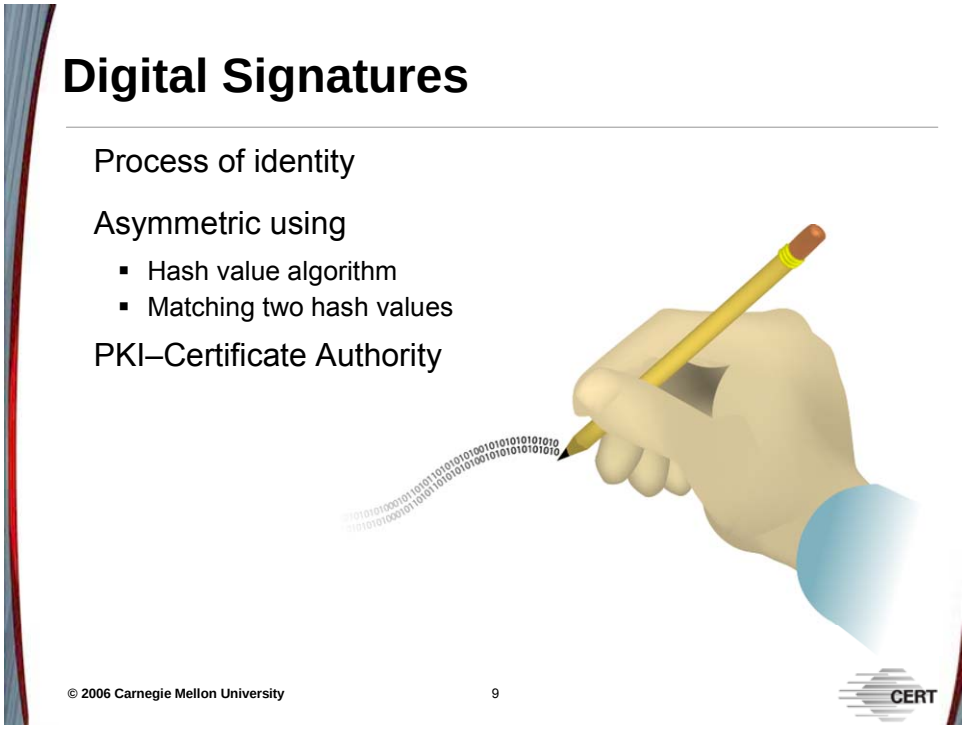
One downside of symmetric keys is the necessity of sharing the secret key. Not only are there challenges to securely exchanging a shared key, but there is also an issue of scale. A user needs to have a shared key with each of the entities with which he or she needs secure communication. This makes key management a burdensome task.

Asymmetric cryptography tackles this hurdle by making use of *key pairs*. A key pair is a set of two keys, one private and one public. These two keys are created using a mathematical procedure such that the two keys complement each other. In this method, the sender uses the destination entity's public key to encrypt the data. Once the data is encrypted, it can only be decrypted by the holder of the corresponding private key. Even the person who encrypted the data cannot decrypt it because that person only has the public key, not the private key. When the destination entity receives the cipher text (encrypted data), that person can use his or her own private key to decrypt the message. Keys can be produced so that it is computationally infeasible to determine the private key based on the public key. Asymmetric encryption algorithms are based on complex mathematical operations, such as prime number factorization in the case of the RSA algorithm or discrete logarithms in the case of the El Gamal algorithm.

The existence of a public key that can be openly distributed greatly reduces the key management challenges inherent in symmetric key implementations. However, that simplification comes at a cost. It is more computationally complex to use asymmetric algorithms, so they run relatively slowly. It is therefore not practical to encrypt large amounts of data using this method.

To combine the best properties of both systems, hybrid systems have been developed. Protocols such as SSL (Secure Sockets Layer, widely used in e-commerce) and some implementations of IPSec (a network-layer security protocol that addresses some of the security weaknesses of the original Internet Protocol) use both mechanisms to utilize their strengths. They use asymmetric keys to handle the exchange of a shared, symmetric key. Generally, at the beginning of a communication session, the protocol involves encrypting a random shared key with the other party's public key. Then the receiving party decrypts the message with his or her private key to extract the shared key. After some confirmation process, both parties encrypt and decrypt all messages using the negotiated, efficient, symmetric key algorithm.

So far, this cryptography discussion has presented some background on keys. But keys do not necessarily confirm identity more strongly than passwords do. Instead, the strength of cryptography provides for identity management through the use of digital signatures and certificates, which we'll discuss next.



4.1.5 Digital Signatures

We noted in the previous section that the private and public keys of a pair are complementary. For the digital signature process, which is designed to confirm a message sender's identity, they are used in reverse. The signer uses his or her own private key to encrypt a message, and anybody can use the signer's public key to decrypt it. If the decrypted value is correct, then the message really did come from the holder of the private key. Because it is inefficient to use asymmetric algorithms on large sets of data, digital signatures are typically used on a hash value of the data, generated by a fast, efficient hashing algorithm. Remember, a hash function such as MD5 or SHA-1 turns an arbitrary-length string of data into a unique fixed-length value. To verify a digital signature, one calculates a hash of the data, using the same algorithm that the signer used, and then compares it to the hash value discovered by decrypting the signed data with the signer's public key. If the two hash values match, then the verifier can have high confidence that the data did come from the private key holder.

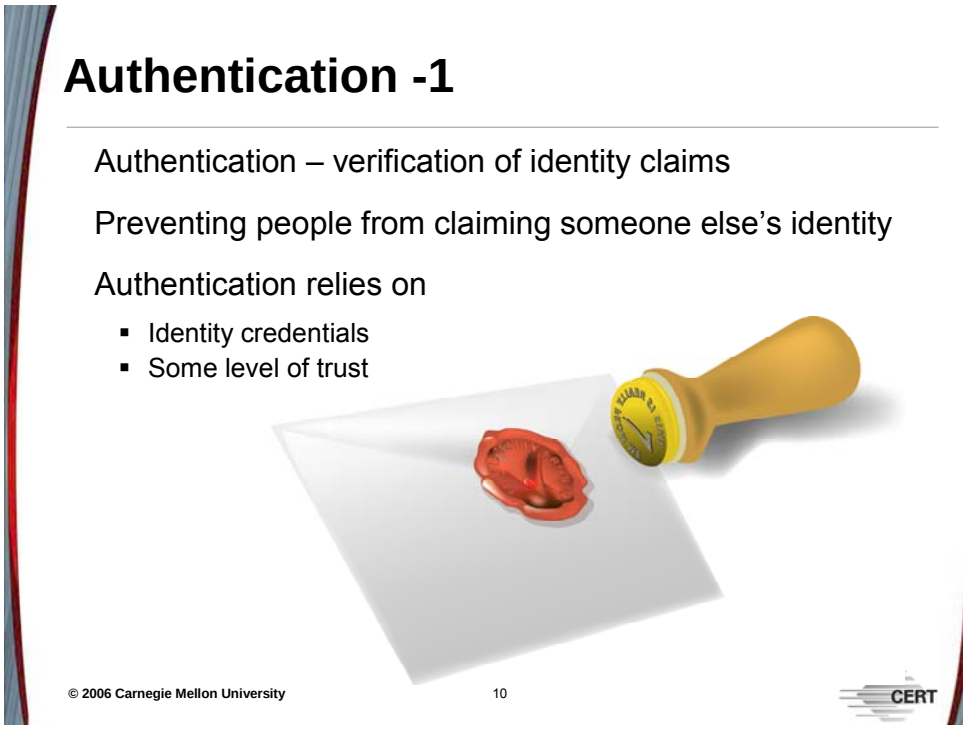
4.1.6 Certificate Authority

One problem with digital signatures is that anyone can obtain a public/private key pair and claim to be a certain person with a certain name. If you have never met the message sender, how do you know they are who they say they are? This is, at its root, a trust problem. Trust is essential within the realm of digital identity because of the dynamic nature of digital relationships. One solution to this problem of trust is the use of certificates for identity management. In the same way that a state issues a driver's license to bind an identity to certain driving privileges, a trusted third-party organization can bind the identity of a person

to a particular public/private key pair. That is the role of a certificate authority: to attest that the holder of a public key is in fact the entity it claims to be. The authority does this by issuing a certificate, which is a digitally signed set of data that associates identity information with a public key. With this certificate, there is no need to take an identity claim at face value; rather, one can confirm the identity claim with a disinterested third party: the certificate authority. This reduces the complexity of trust relationships; one need only trust the certificate authority, rather than trusting everybody who claims to be somebody.

In practice, most people use this form of identity establishment whenever they conduct an online e-commerce transaction. Using SSL/TLS, the user receives a certificate from the Web server, so that the user can be sure he or she is communicating with the correct Web server. The user's browser checks the Web server's claimed identity against a list from the appropriate certificate authority and displays any discrepancies between the claimed identity and the certified identity. Users can decide to trust or distrust various Certificate Authorities based on their understanding of each authority's competence. To facilitate e-commerce, most browsers default to trust most commercial certificate authorities.

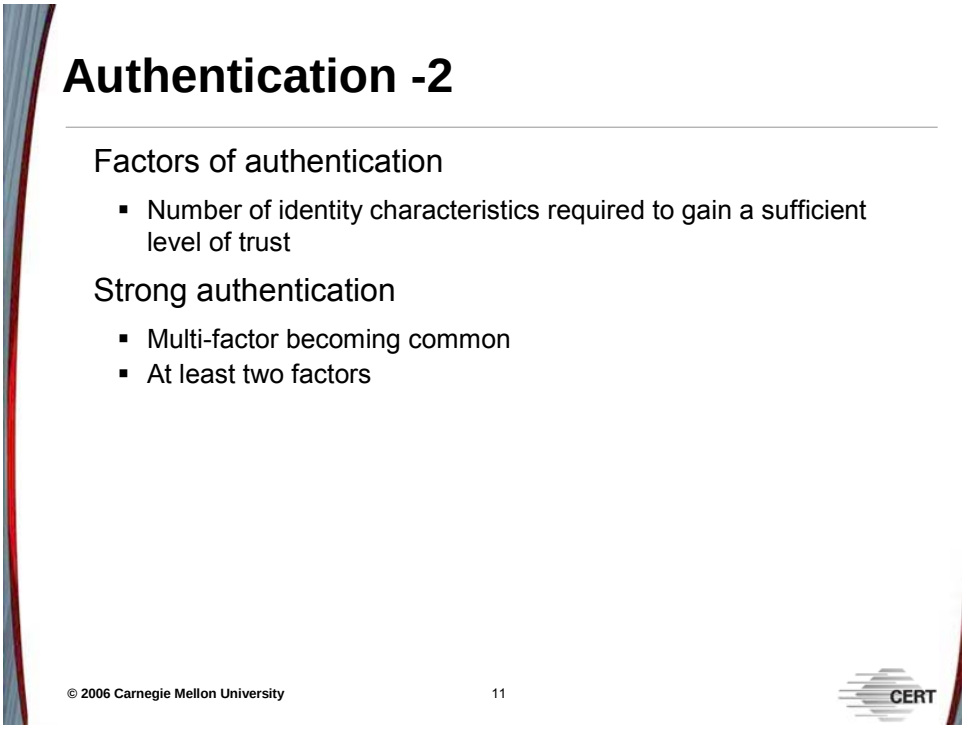
It is possible for users to obtain their own digital certificates. Most browsers allow the creation of digital certificates at no cost; however, these certificates will have a low trust level because they have not been authenticated by a certificate authority. Certificate authorities such as VeriSign will authenticate any person's certificate for a fee that varies depending on the degree of authentication required.



4.2 Authentication Efforts

4.2.1 Definition

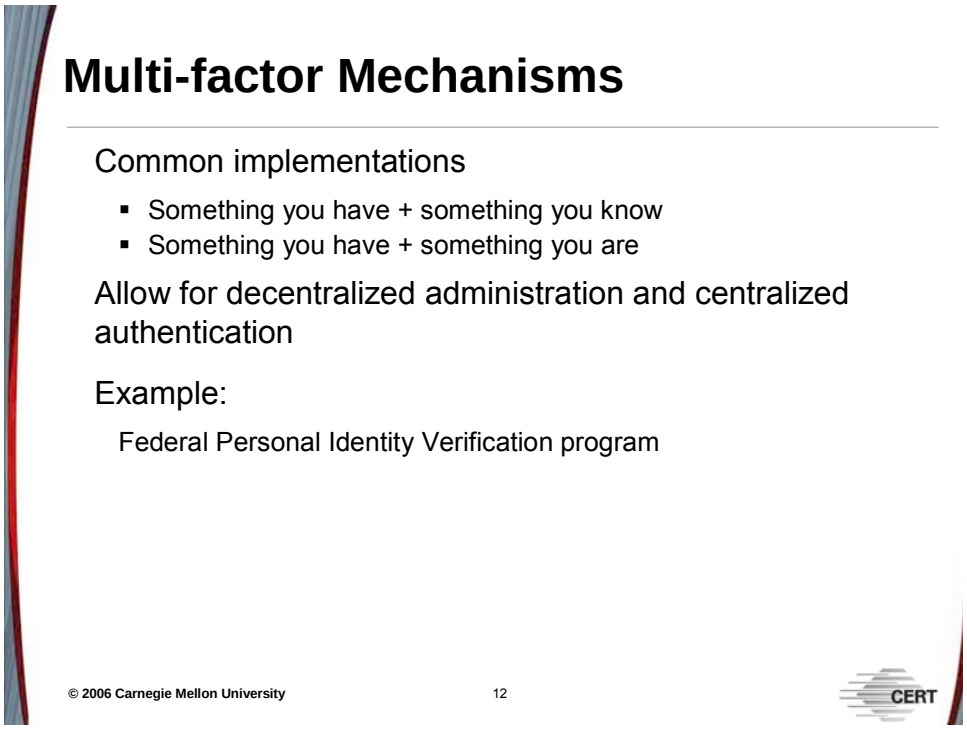
Authentication relies on identity and trust. The previous discussion of the most common forms of digital identification has shown that identity management is far easier when a third-party authority is involved. This is because identity characteristics alone cannot prove an identity claim. Even in the offline world, assurance of identity always relies on an authentication authority. When somebody offers a driver's license as a proof of identity, the verifier has to trust that the state that issued the license performed due diligence. The verifier for the state would have to have seen the person's birth certificate or some other official corroboration of identity before providing the state's own certification that a certain photograph really is valid for a certain person.



4.2.2 Multi-factor Authentication

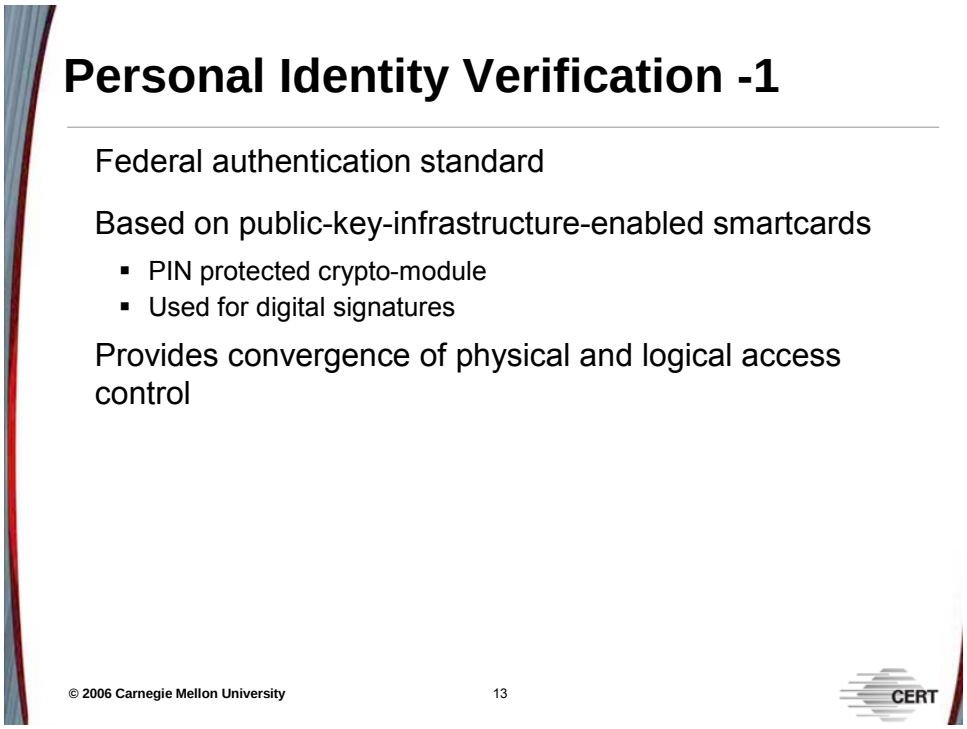
Identity management is not failsafe. Passwords can be forgotten, stolen, or guessed. Tokens can be lost or stolen. Cryptographic keys can be compromised. Thus, there is a move throughout industry and government to adopt the concept of *multi-factor authentication* or *strong authentication*. This approach involves using a combination of identification measures to verify an identity claim. This is consistent with the concept of a layered defense and a Defense-in-Depth mindset. Instead of using just a password, a system might require a password and a valid fingerprint. Most common is the implementation of a smartcard token that contains a PIN-protected public/private key pair. More stringent security needs may dictate use of both a signed private key and a biometric measure.

As an everyday example, a user of a shared workstation in a high-traffic area might use multi-factor authentication to reduce the risk of data theft if his or her password is compromised via "shoulder surfing." Assuming he or she still possesses the other half of the key, such as a smart card or biometric identifier, his or her data will remain uncompromised.



The purpose of engaging multi-factor mechanisms is to add an element of complexity that can stymie an attacker trying to forge an identity. By combining something you know with something you are or something you have, you can drastically reduce the risk of somebody simply learning that which you know or stealing that which you have.

One example of a system employing multi-factor authentication is the Federal Personal Identify Verification program, which we'll discuss in the next few slides.



4.2.2.1 Personal Identify Verification

Large organizations are moving toward multi-factor authentication. For example, U.S. federal agencies have been working toward an implementation of such a system. The National Institute of Standards and Technology (NIST), a Department of Commerce agency, has published a Federal Information Processing Standard (FIPS 201)[6] to detail the policy of a Personal Identity Verification (PIV) program [NIST 06]. Designed to provide secure and reliable forms of identification for all government employees and contractors, this standard provides a model with three primary components:

1. PIV card issuance and management
2. PIV front-end system
3. access control that provides identification and authentication.

[6] See http://csrc.nist.gov/publications/fips/fips201-1/FIPS-201-1-chng1.pdf.

Personal Identity Verification -2

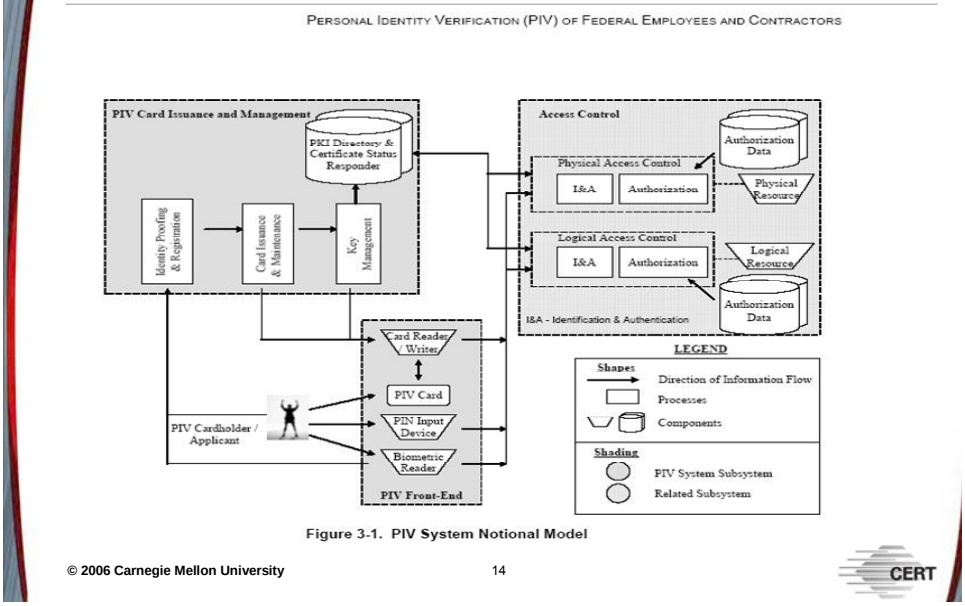


Figure 3-1. PIV System Notional Model

CERT

The issuance and management component of PIV provides for the administration of the credentialing program. It is composed of three critical tasks.

First, it is responsible for the identity proofing of card applicants. This process verifies an applicant's identity using other identity documents to ensure the PIV card is issued to a legitimate government employee or contractor. This proofing is the foundation of trust that upholds the rest of the system. The identity and authentication systems must rely on it. In fact, a straightforward means of attacking the system is for an attacker to get official credentials for a fake identity.

Second, after validating the applicant's identity, the issuance and management component must build a card with the appropriate identification characteristics. These can include a photograph, name, biometric data, and a PIN for unlocking the embedded identity credentials.

Finally, the issuance and management component is responsible for key management of issued cards, which includes a local Public Key Infrastructure (PKI). Each card is designed to generate public/private key pairs for asymmetric cryptographic functions. (In fact, the PIV cards are required to have at least one key, but other keys also may be added for other functions like card management, digital signatures, and key management.) The public key is then signed by the local certificate authority, and that certificate is embedded into the card as well. To add flexibility for large organizations, certificates can contain certification authority chains. This allows some measure of distributed key management among various departments. So long as all chains have a common root or connection, credentials can still be authenticated throughout the whole organization. This public key infrastructure is essential

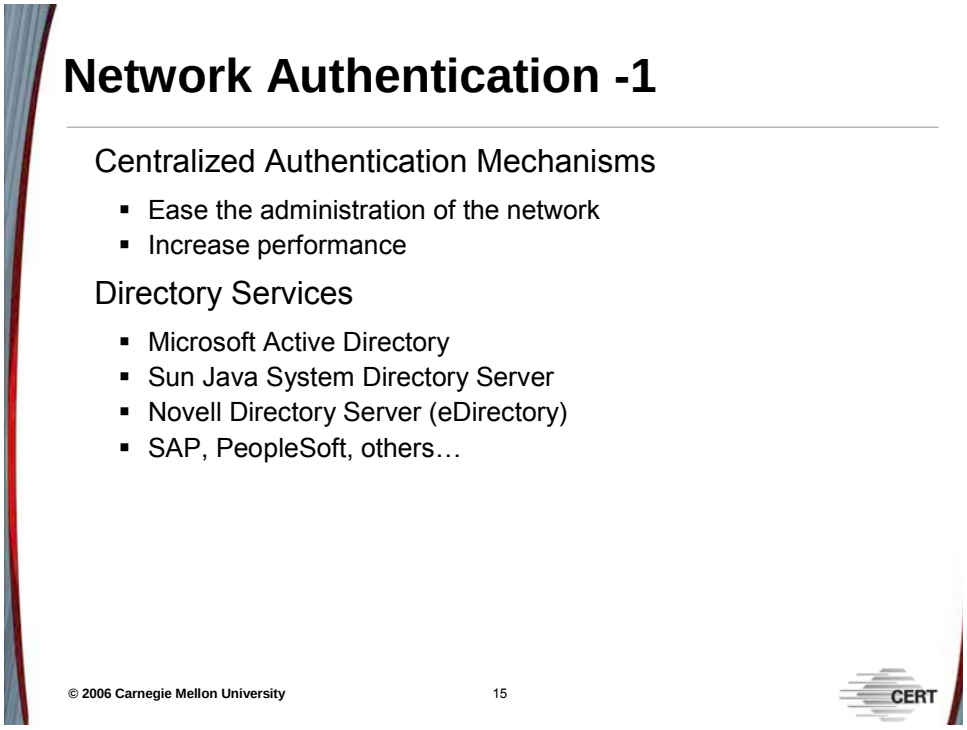
to efficient management of the PIV program and allowing common credentialing throughout the government; however, it broadens the scope of trust that the authentication process must assume. There is a separate Common Policy Framework for the U.S. Federal PKI, under which PIV Certification Authorities must operate.

The PIV front-end component serves as the user interface for cardholders. It consists of the collection of hardware devices necessary to electronically authenticate the cardholder. At its simplest, the front-end contains a card, a reader, and a PIN input keypad. This PIN input allows the user to activate the card, releasing the card's protections on the user's credential data, including the necessary private keys. The PIN provides a form of two-factor authentication that mitigates potential loss or theft of the card, because the card would not be useful without knowledge of the PIN to activate it. Another means of higher security is the addition of a biometric reader. The biometrics stored on the card can be checked against actual data taken from the reader. These types of front-end components are placed at both physical and logical access points. Again, a substantial advantage of the PIV program is its convergence of physical and digital identification needs. A physical access point might be at a door allowing people into a room, and a logical access point might be at a computer, allowing a user to log in to the network. The PIV card would work for both purposes.

The third part of the PIV system is the Access Control component. This component interacts with the front-end system to accomplish both authentication and authorization. In a non-networked, physical access-based environment, a guard could use the photo and data printed on the card to authenticate the individual. However, the strength of the smartcard system is in leveraging technology to make such determinations. The identification and authentication process involves querying the card in the front-end reader for identity information. Depending on the level of security necessary, different authentication routines may be used. It may be enough to trust a valid digital signature of the card's data. It may be necessary to compare biometric reader results to those stored on the card. Furthermore, the identity and authentication process can authenticate the user by testing the user's private key against a provided challenge. In this situation, the card is required to sign the challenge with its private key. If it is successfully verified with the card's public key, and the key was signed by a trusted certificate authority, the user is considered authenticated. At this point, other card data is passed to the authorization process, which compares the access request against the privileges and limitations that govern the authenticated user.

This variety of authentication techniques allows for a range of identity assurance. Based on the rigor of the identity-proofing process, the security of the card issuance and maintenance process, and the strength of the technical means of authentication, the access authority can have SOME, HIGH, or VERY HIGH confidence in the cardholder's claimed identity.

This example of the U.S. Federal Personal Identity Verification standard demonstrates the scope of the identification and authentication efforts necessary to successfully manage information security in large organizations.



4.2.3 Network Authentication

Modern information systems are built to facilitate data exchange throughout a network of associated computers. To provide a secure network environment, authentication mechanisms have evolved into complex, efficient protocols. Because authorization depends on authentication, these measures must be supported at all layers of the computing environment. Before a user can run a program, for example, the operating system must determine if it is allowable. Before a program can access a remote resource, the remote service must know who is making the request.

Nearly all operating systems have the capability to authenticate users against a local account database. This provides a solid basis for authenticating the user of resources on that particular machine. However, the considerable administrative effort of managing separate accounts for the same user on different machines has led to the development of centralized authentication services. This allows administrators to maintain account identity characteristics at a central place on the network so that network resources can pass credentials to the central service for authentication. This approach not only eases the administrative burden, but also enhances the user experience by removing the need to log on at every different resource.

Central authentication service is typically bundled into an enterprise's directory services resource. Nearly all major operating system vendors provide a directory service such as Microsoft's Active Directory, Sun's Java System Directory Server, or Novell's Directory Server. Other software providers such as SAP and PeopleSoft also have developed centralized authentication mechanisms.

Network Authentication -2

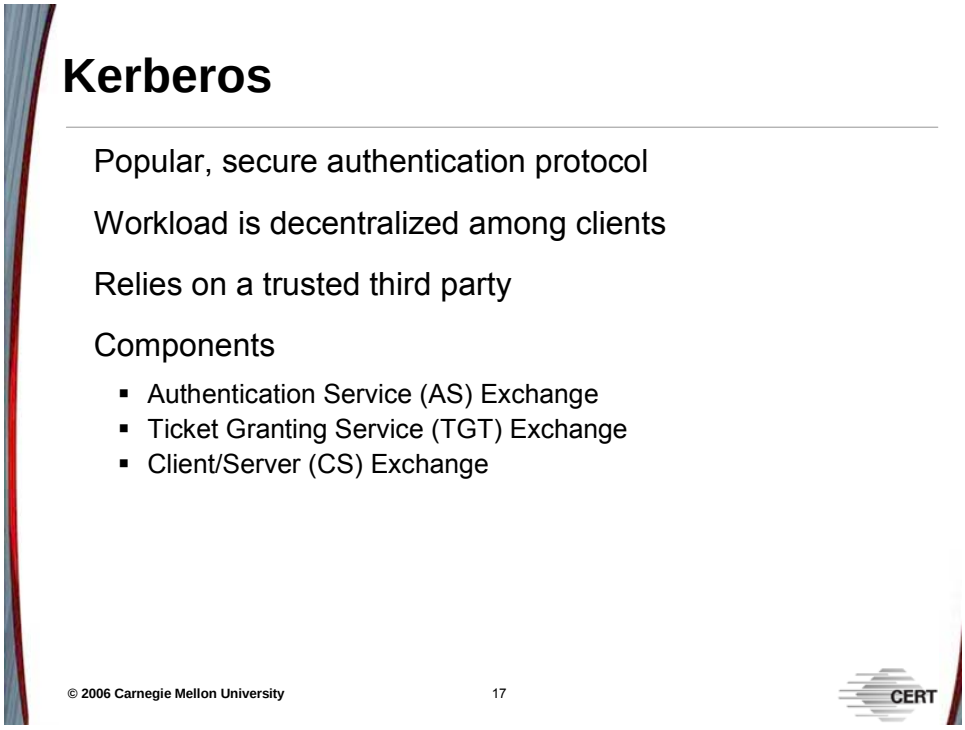
Common Protocols

- LM, NTLM
- Radius, TACAC
- PAP, CHAP
- Kerberos

Issues

- Excessive communication?
- Mutual authentication?

With the increase in the number of authentication methods, clients, and servers, many directory services focus on interoperability. To that end, they are generally capable of negotiating a compatible protocol between clients and servers. One such protocol that has become popular as it matures is the Kerberos authentication protocol. Kerberos was developed at the Massachusetts Institute of Technology and has been implemented by Microsoft as its default protocol since the Windows 2000 operating system. Kerberos offers several advantages over earlier protocols like LAN Manager (LM) and NT LAN Manager (NTLM) because it allows for mutual authentication of two parties and decentralizes the connection management aspects of the system, which we will discuss on the next slide.



4.2.4 Components

4.2.4.1 Kerberos

Kerberos provides highly efficient, secure exchange of authentication data [Kohl 93]. The protocol relies on all parties trusting a central key manager. All the work of managing the authentication process is handled by the client, which initiates all connections and keeps track of its established keys. There are three components of the Kerberos protocol:[7]

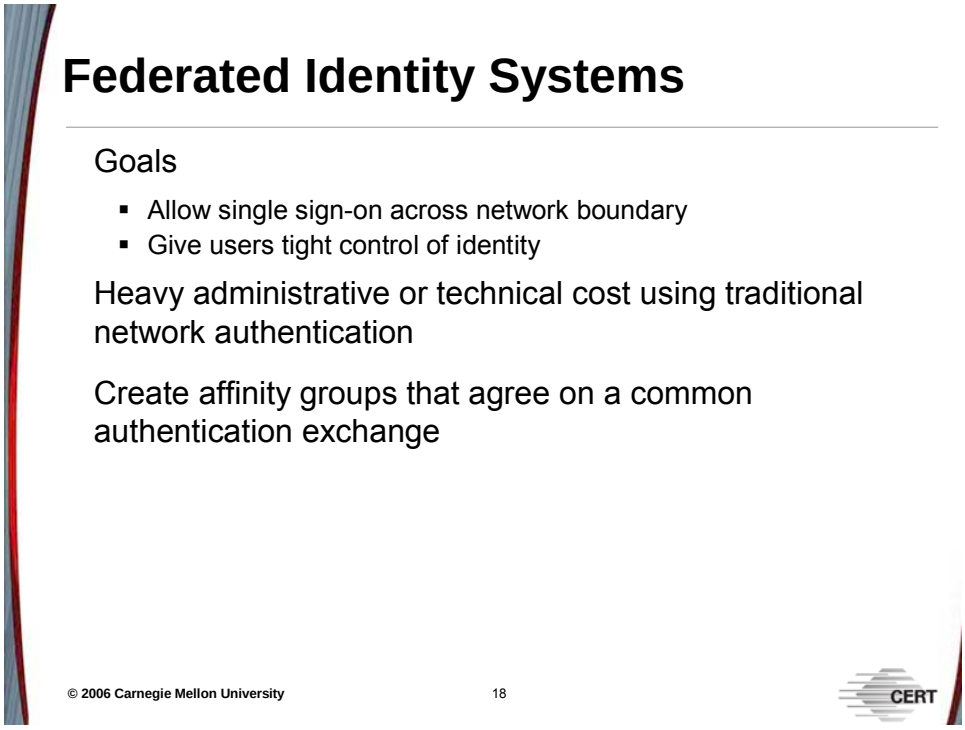
1. Authentication Service (AS) Exchange. The first step toward using Kerberos for network-wide authentication is to establish a shared key with the key distribution center. The client and the key manager already share a secret—the user's password. However, to reduce the exposure of this long-term secret, the AS exchange only uses the password to establish a new short-term shared key. This logon session key, which is shared between the user and the key manager, serves as proof of identity until the key expires. Along with this key comes a ticket for the Ticket Granting Service.
2. Ticket Granting Service (TGS) Exchange. When a user needs to connect to a server, the user asks the central key manager for a key specific to that connection. Using its logon session key from the AS exchange, the client encrypts a request for a ticket to connect to the server. Assuming the request decrypts properly with that user's logon session key, the key manager creates a new key for use between the user and the server. The manager returns two copies of the new key: one encrypted with the user's session key and the

[7] See http://www.ietf.org/rfc/rfc1510.txt.

other encrypted with the server's key. The client can store and reuse this ticket until it expires.

3. Client/Server (CS) Exchange. When a client wishes to make a connection to the server, that client forwards the server's copy of the Kerberos ticket to the server. Because the ticket is encrypted with a key that is known only to the server and the key manager, the server can assume the requesting party has been authenticated properly by the AS exchange.

The Kerberos standard has room for extensions that modify its mechanics to a small degree. The protocol is easily adaptable for use with a PKI-enabled smartcard instead of a password. In the AS exchange, instead of sharing a password to create a logon session key, the local security authority can forward the user's certificate to the authentication service. If the AS verifies an authoritative signature in the certificate, it returns the logon session key encrypted with the user's public key.



4.2.4.2 Federated Identity Systems

Federated identity systems aim to overcome the obstacles of authenticating across disparate networks. The demands of electronic commerce have grown so much that business partners often make network resources available to each other. Especially in the realm of Web services, authentication across organizational boundaries quickly becomes tenuous. If a supplier publishes an ordering service, it must configure the service to be accessible only to authenticated users. This generally involves a substantial administrative workload to keep up with the changing user base. Distributing the authentication work to a user's home network introduces the problem of differing infrastructures and protocols. Technical challenges must be overcome to translate a Kerberos authentication from one network into a Lightweight Directory Access Protocol (LDAP) authentication on another.

Federated identity systems create an efficient means of achieving common acceptance of decentralized authentication. All the current approaches consist of developing a common language and protocol through which different entities can transfer authentication data. This common glue between disparate systems of partners makes the system extensible. It becomes easy to add or remove new partners.

Federations are created by a group of organizations that have some relationship that joins them together for a purpose. A common example is that of a travel agency that works routinely with hotels, car rental firms, and airlines. Any organization that deals with travel and sees a benefit to shared authentication across industry partners can join the federation.

A federation can be formal or informal. In the context of identity management, it essentially refers to an organization that will provide authentication information according to a common standard.

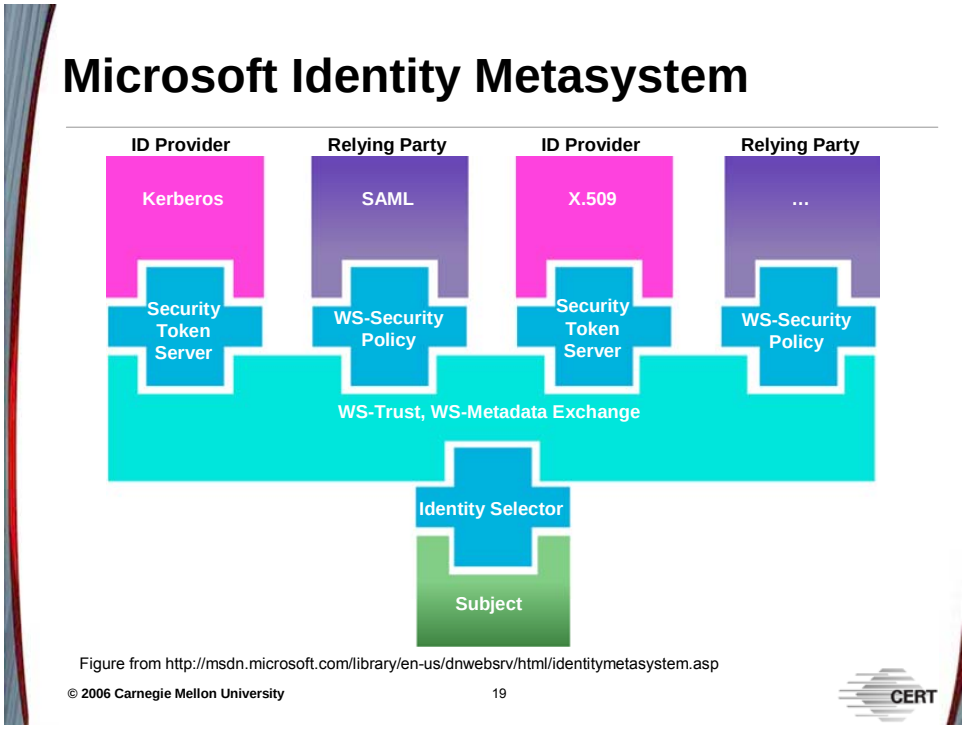
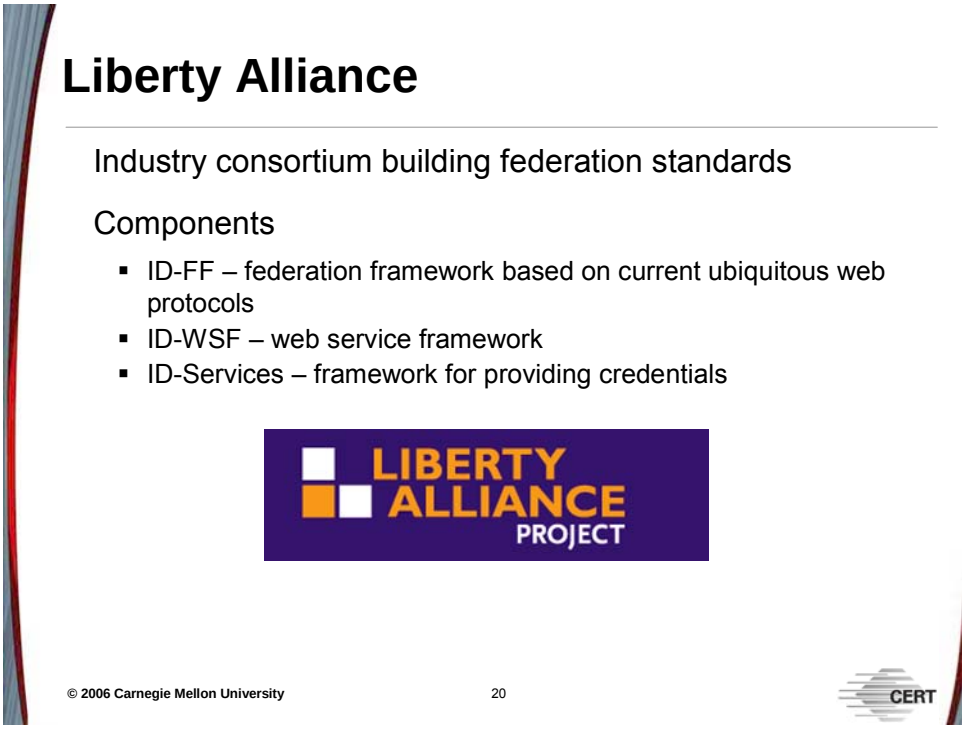


Figure 1: Identity Metasystem Architectural Diagram [Microsoft 05]

4.2.4.3 Microsoft Identity Metasystem

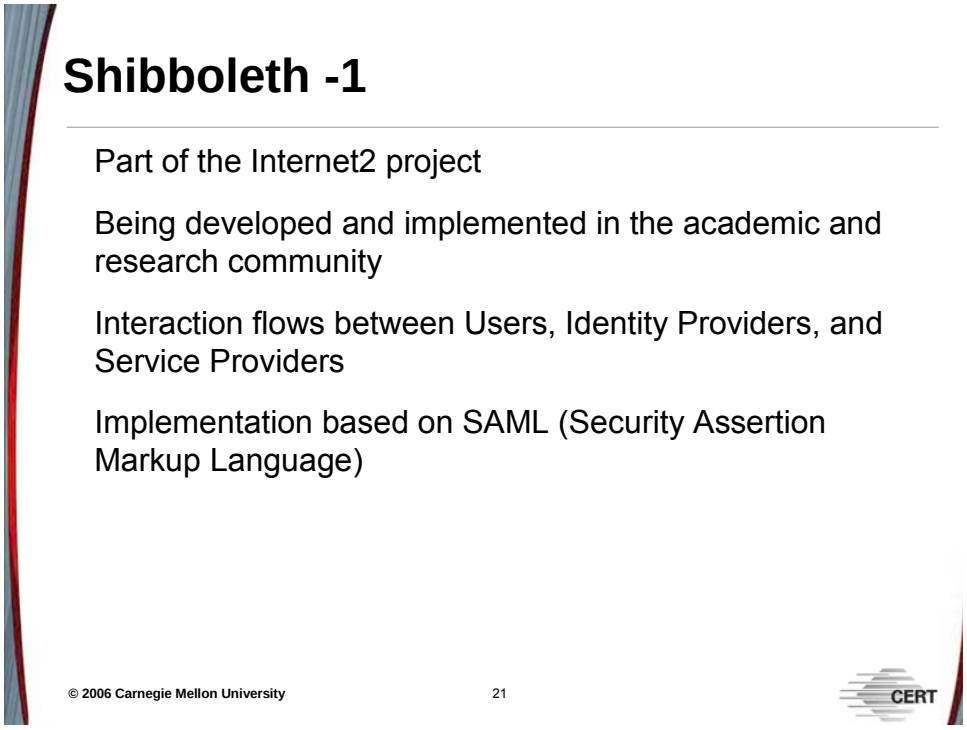
There are at least three large federated identity efforts under development by Microsoft, the Liberty Alliance, and Shibboleth. Microsoft is building a system called MS-Infocard. Having learned lessons from its MS Passport program, it is giving users control over their identities so they can determine what relationships they consider justifiable. MS-Infocard is built on a foundation of Web Services Trust Language (WS-Trust) Web service protocols. These Web services transform identity claims and authentication tokens received from an "ID Provider" into a format understandable by a "Relying Party."

For more information, see the Wikipedia entry at http://en.wikipedia.org/wiki/Infocard.



4.2.4.4 Liberty Alliance

Working along the same lines, the Liberty Alliance also is developing specifications for federated identity management. This consortium is composed of approximately 150 companies worldwide that understand that strong identity and authentication is in the best interests of both consumers and industry. They have developed standards for three components of the Liberty identity system: a federation framework, a Web service framework, and ID Services specifications. The federation framework provides technical standards for achieving federation and single sign-on using only the most widespread current technology. The Web service framework (ID-WSF) incorporates standards that can be implemented as Web services become more widely used. For more information, see http://www.projectliberty.org/resources/specifications.php.

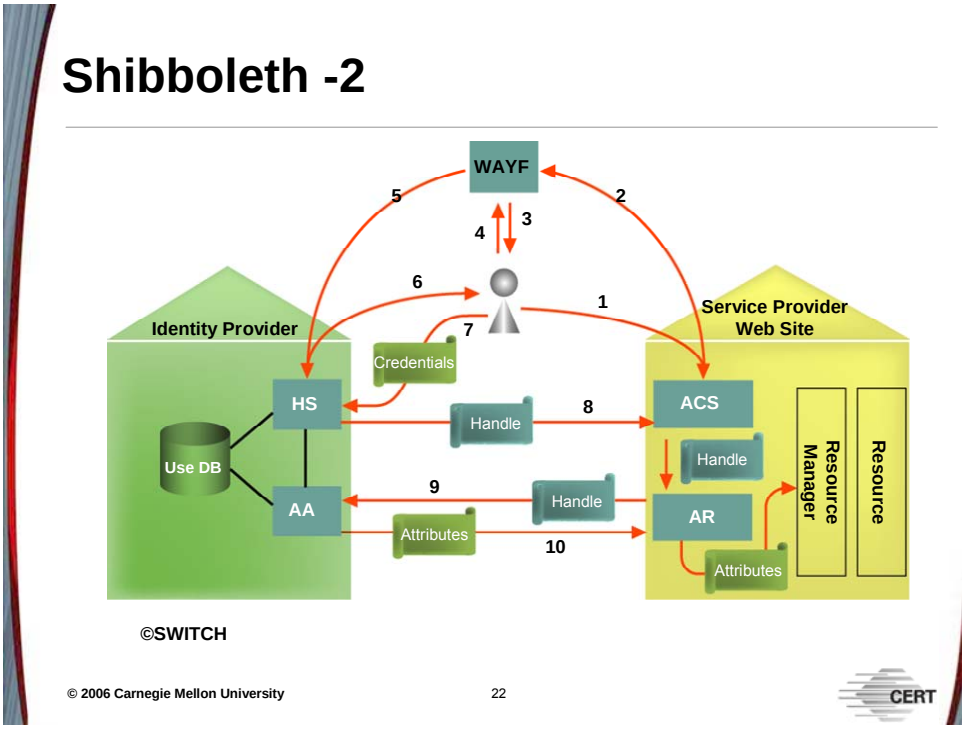


4.2.4.5 Shibboleth

A third effort, propelled largely by the academic community, is called Shibboleth.[8] Associated with the Internet2 project that connects many universities and research institutions, Shibboleth's approach to federated identity management considers the same types of standards and entities as the Microsoft and Liberty Alliance models.

In Shibboleth's case, the user interacts with an Identity Provider and a service provider. One of the goals of this system is to protect user identity by disclosing to the service provider only those characteristics the user has approved in advance. When very limited characteristics are approved, the service provider might only receive assurance that the user is authenticated by the Identity Provider. In other scenarios, a user may agree to provide a name or some demographic information that might help the service provider present additional applicable resources, so the Identity Provider also will disclose this information upon request.

[8] See http://shibboleth.internet2.edu/about.html.



The message flow of the Shibboleth protocol circulates through all three primary entities: the user, the Identity Provider, and the service provider. First, the user initiates a resource request to a service provider. The service provider redirects the user to his or her chosen Identity Provider, where the user is authenticated with local credentials. The Identity Provider generates a unique handle (identifier) for the user's authenticated session and sends it back to the user, who presents the handle to the service provider. The service provider then contacts the Identity Provider directly, requesting any necessary or allowable attributes pertaining to the generic authenticated handle that it holds. The Identity Provider returns attributes to the service provider based on the user-controlled disclosure policy.

All messages involved in this exchange are formatted using the Security Assertion Markup Language (SAML) standards produced by the Organization for the Advancement of Structured Information Standards (OASIS) Security Services Technical Committee. This common, extensible language promotes growth of the technology by making adoption easy and stable. SAML is an XML syntax that describes security information, including identity claims and authentication assertions.

Protecting Identity and Privacy

Identity theft is not exclusive to the digital realm.

The Internet makes it easier to use stolen identity.

Policy efforts must match technology efforts.

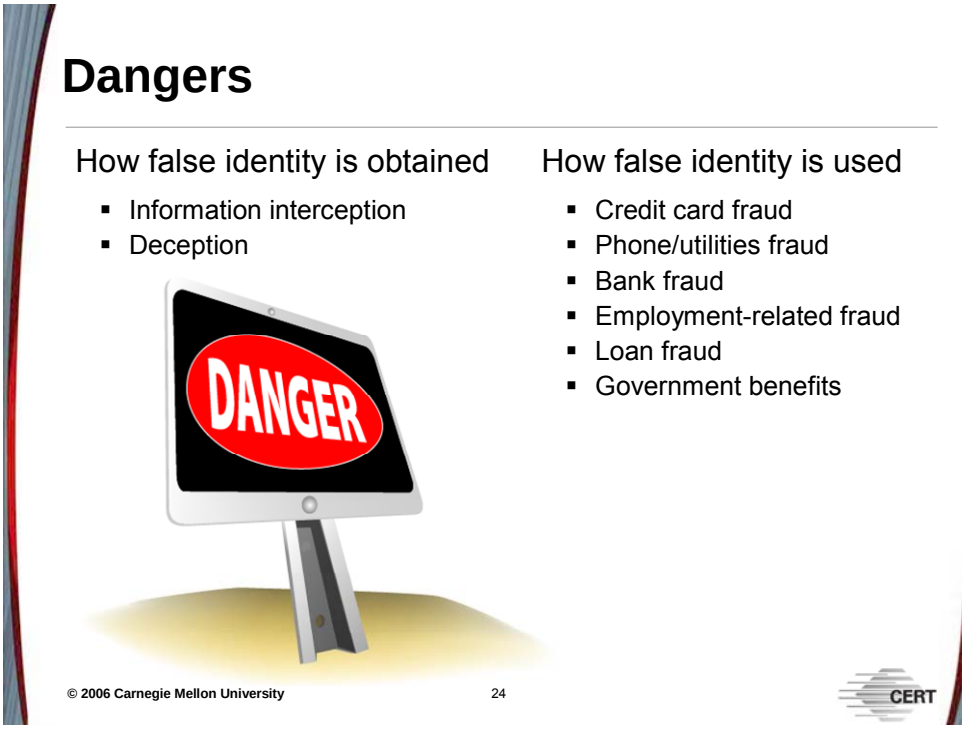
Principle players

- Federal and State Government
- Privacy Advocacy Groups
- Technology developers and managers

CERT

4.3 Safeguarding Identity

Protecting user identity is a growing concern. Many have likened the Internet to a lawless 'wild west' that will crumble if most people cannot operate safely. Fortunately, the significant technical progress made in clarifying and streamlining digital identity management has been matched by progress in policy dealing with the proper treatment of identity. Ongoing policy efforts have aimed to raise awareness of identity theft and of measures that can protect against it. The U.S. Federal Trade Commission (FTC) is charged with protecting consumers and is a major voice on identity theft issues. It has created an online information resource at http://www.consumer.gov/idtheft.



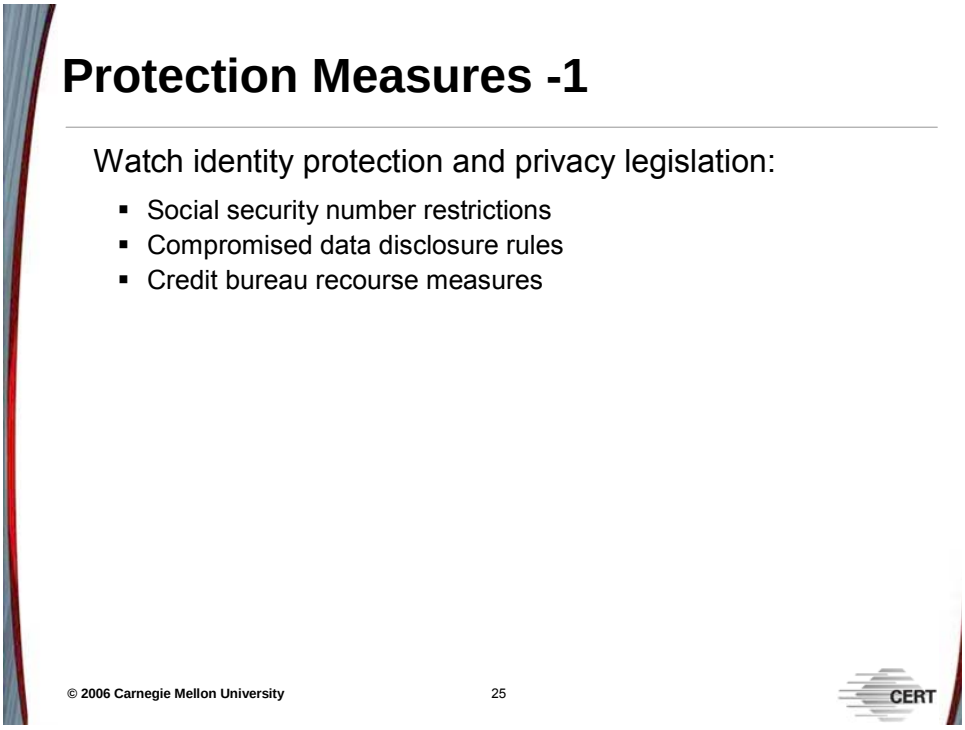
4.3.1 Dangers

Identity theft is not exclusive to the digital realm. However, widespread use of the Internet has exacerbated this longstanding problem. After all, the physical act of stealing credit card information out of a mailbox is inefficient for an attacker in terms of time invested to gain a single credit card number. On the other hand, an attacker can run programs to glean credit card numbers from Internet traffic with little effort. The potential scale and scope that thieves have on the Internet makes it possible for them to achieve high returns on their illegal investment. Identity information can be intercepted at data entry, in transit, or in storage. If interception poses a difficulty, thieves may simply attempt to deceive a user into thinking they are a legitimate entity.

Most thieves conduct their activities for financial gain, so it is not surprising that most identity thefts involve money. The FTC maintains an identity theft database called the Consumer Sentinel, which logged roughly 254,000 reports in 2004. It reports the following categories as the major areas of concern:

- credit card fraud (28%)
- phone or utilities fraud (19%)
- bank fraud (18%)
- employment-related fraud (13%)
- loan fraud (5%)
- attempted identity theft (6%)

- government documents or benefits fraud (8%)
- other fraud (22%)



4.3.2 Protection Measures

One way to protect identity is to stay alert to federal and state legislation that affects privacy. Many state governments have been working on privacy protection, specifically online identity protection, for at least the last five years. Some of the laws enacted

- restrict the use of certain identifying information, such as the Social Security number
- require corporate disclosure of personal information data loss
- provide victims of stolen identity with recourse to fix damaged credentials

The national trend in recourse has been to give consumers some control over their credit reports. In many states, there are requirements for credit bureaus to provide free credit reports for consumer review. Additionally, if consumers suspect identity theft, they can place a freeze on any new credit without the credit bureaus' explicit approval.

Protection Measures -2

Carefully manage digital identities.

Place security controls on corporate and home networks and computers.

Build security plans that consider insider threats.

Managing digital identities carefully can reduce exposure to online theft. But one reason this problem is so daunting in the first place is the proliferation of identifying information required to conduct online business. As federated identity systems mature, it should become easier to limit the information that different entities can collect about other entities.

In the meantime, a basic precaution is to encourage users to avoid using the same password for a banking Web site that is used for a casual Web site, such as eBay or Amazon. Identities should be kept separate and unlinkable.

Protection Measures -3

Be aware of common deception measures:

- Phishing
- Pharming
- Malformed URLs

Much identity theft occurs simply through tricks and deception. Awareness of common attack methods such as spam, “phishing,” and “pharming” can help prevent consumers from providing credentials to imposters. Further, proper use of common security controls can filter out a great part of the deception threat.

“Phishing” is the solicitation of access credentials under false pretenses. Generally, a user will get an email that looks like it is from a familiar organization asking the user to provide further information by clicking on a link. However, the link actually points to the attacker’s Web site, which has been crafted to look like the legitimate Web site. When the user logs in, his or her logon credentials are exposed. It should be noted that most online services never solicit information via email, so any emailed requests to log in somewhere should be treated as suspicious.

“Pharming” is based on the same idea as phishing; however, it uses a more sophisticated model for attracting victims. Rather than convincing users to take some action, pharming attempts to redirect users after they request a legitimate Web site. This is accomplished through DNS poisoning—changing the address of a Web site from the real address to an address containing a spoofed version of the site. This redirection can also be achieved through use of viruses that alter a computer’s host file, statically assigning a false address to a Web site and preventing a dynamic lookup of the real address.

Protection Measures -4

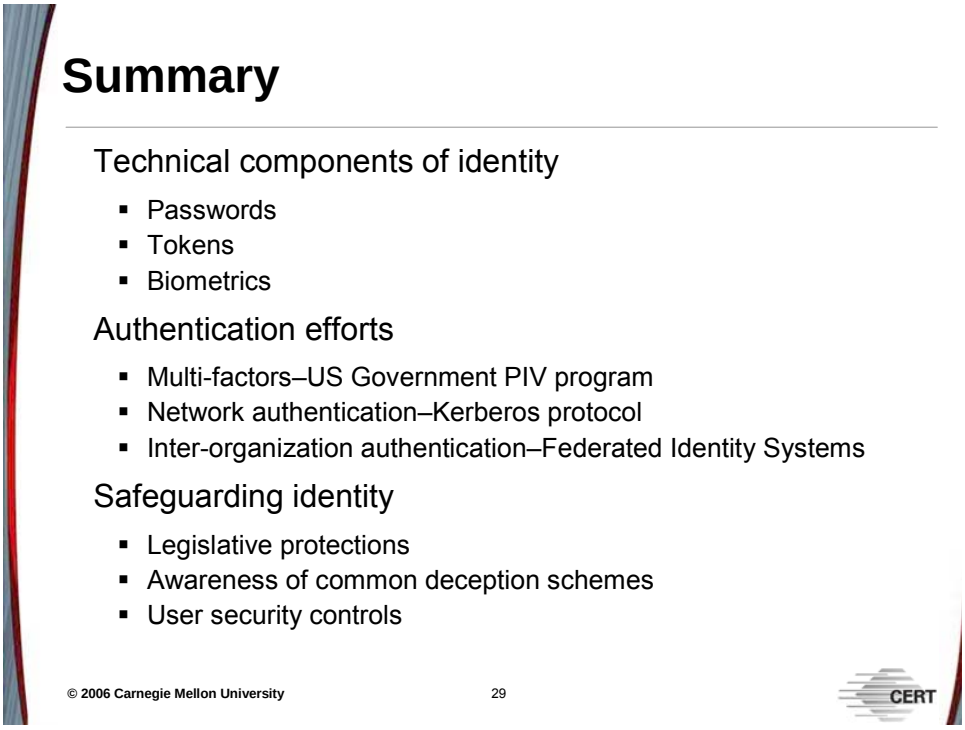
Use encrypted or authenticated protocols for Internet communication–SSL, HTTPS.

Pay attention to browser security controls.

Many authentication failures are reported through browser alert.

Internet browsers provide security controls that are generally sufficient to expose online deception. A common way to avoid exposure is to connect to sensitive Web sites using Secure Sockets Layer (SSL). The SSL protocol triggers visual security references in the browser, such as the yellow padlock icon and the "https://" designator. When using SSL, the Web server must provide a digital certificate proving its identity. If there is any discrepancy among the name of the Web site, the name on the certificate, or the certificate authority's signature, the user will be alerted. It is important to look at such alerts because some imposters provide similar-looking certificates that don't quite match. Users in the habit of clicking "Continue" in response to alert messages without inspecting the alerts are likely to become victims.

Unfortunately, information is not necessarily safe even when it is provided by a user to a real organization. A significant insider threat exists that must be met with operating controls and policies within that organization. A comprehensive information security management plan should be designed, such that the ability for insiders to misuse data that they are authorized to handle is restricted and auditable.



By this point, it should be clear that there are many ways to approach Defense-in-Depth. There is no simple checklist that, once completed, results in resiliency. Defense-in-Depth is a holistic concept that requires the integration of many different components to create an overarching culture of security, supported by technology that serves the business mission. It is not just about technology for technology's sake.

Keep that in mind in the next sections as we discuss Authorization Management, Accountability Management, and so on, with the end goal of building toward a more secure enterprise.

Review Questions

1. What identity mechanism relies on “something you have?”
2. Define authentication.
3. What is the difference between symmetric and asymmetric cryptography?
4. List three Federated Identity Systems.
5. Is it safe practice to enter personal information over an unauthenticated channel?

Module 5: Authorization Management

This module discusses best practices and implementation methods for managing authorization and the protection of information integrity and confidentiality.

Instructional Objectives

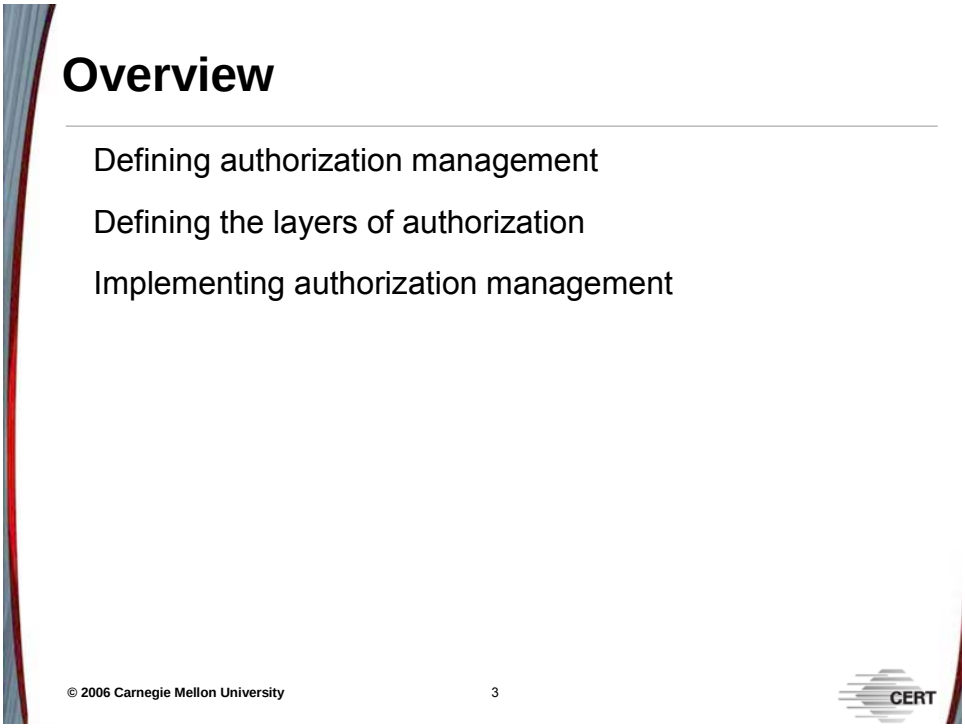
Upon completion of this module, students will be able to

- Define authorization management
- Explain its importance
- Define levels of authorization management
- Explain methods of authorization for each layer
- State best practices for authorization management

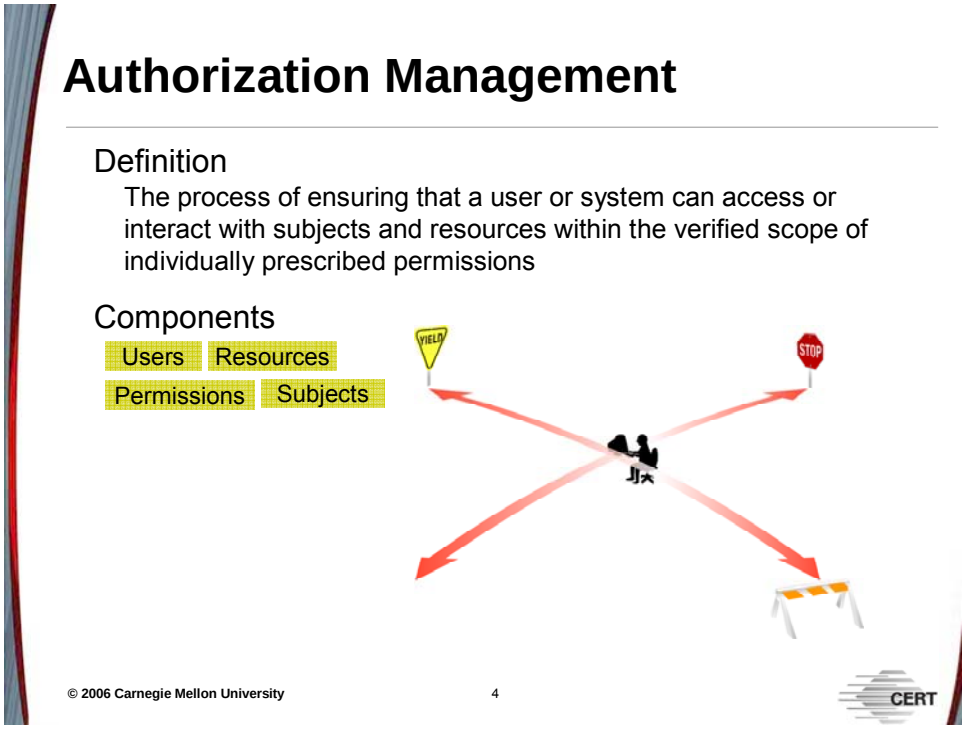
CERT

This instructional module will enable students to complete all of the above learning objectives.

5 Overview of Authorization Management



This module lays the foundation for understanding the various layers of authorization and access control. These can be implemented as a component of Defense-in-Depth information assurance.



5.1 Defining Authorization Management

In the context of information assurance, authorization management is the task of assigning and then keeping track of who has access to what in a networked environment. Authorization management should be viewed as a process within the larger framework of Defense-in-Depth rather than as a single task. Indeed, it is dependent on authentication of identity, which was discussed in the previous module in this course.

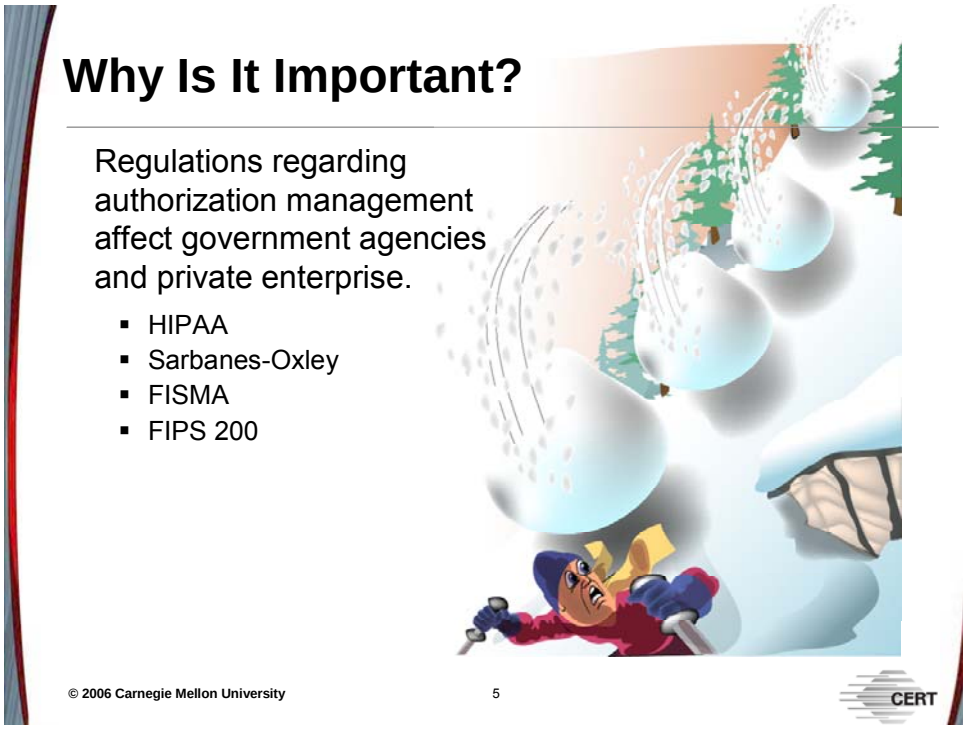
The first step in the authorization management process involves determining which organizational information and resources need protection. These key assets or types of assets should be identified through a formal process that becomes part of organizational policy. Key assets are the data or resources critical to the organization's mission. Identifying them on a highly specific level may require extended effort, as discussed in the Risk Management module of this course.

For government agencies, this process is defined in FIPS Publication 199, *Standards for Security Categorization of Federal Information and Information Systems.*

Once important assets have been defined, authorization management can be implemented. This begins with authentication—establishing the identity of a person or host that is attempting to gain access to protected resources. The concept of authentication, as stated above, builds on the material presented in the Identity Management section of this course.

The primary goal of authorization management is to achieve two of the three primary security goals: confidentiality and integrity. To review, these two principles are defined as follows:

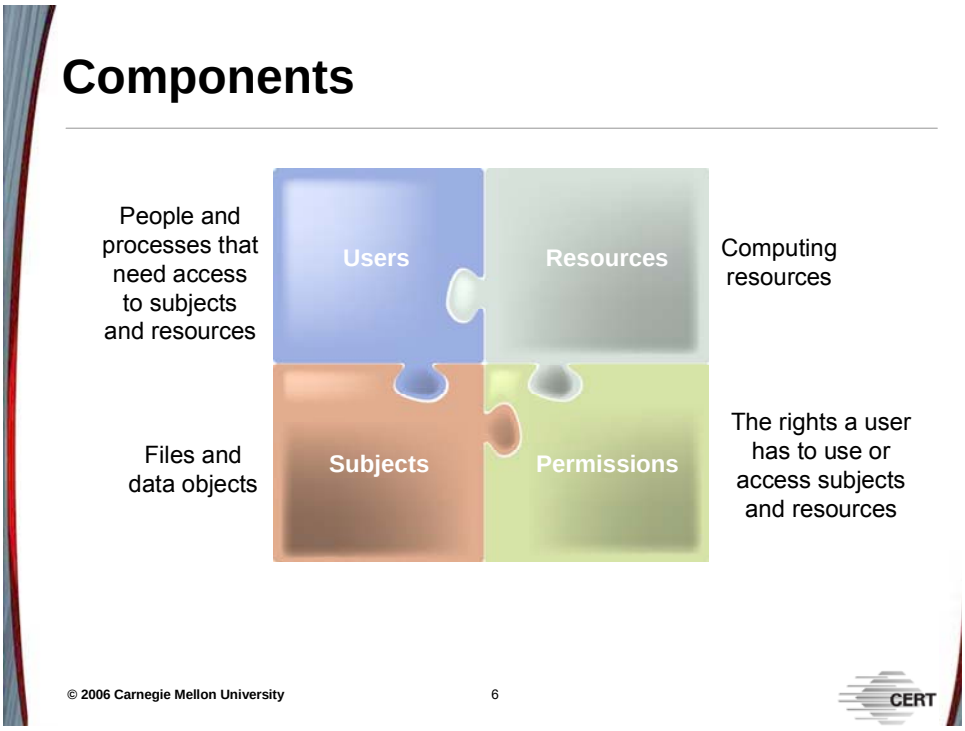
1. **Confidentiality –**the assurance that information can be read and interpreted only by persons and processes explicitly authorized to do so
2. **Integrity –** the assurance that information remains intact, correct, and authentic



5.1.1 Why Is It Important?

Authorization management, when using consistent and reliable authentication mechanisms, can be the most valuable tool for protecting access to elements of your infrastructure. Not only is this a fundamental necessity for information assurance, but it is also increasingly a means of ensuring compliance with government regulations.

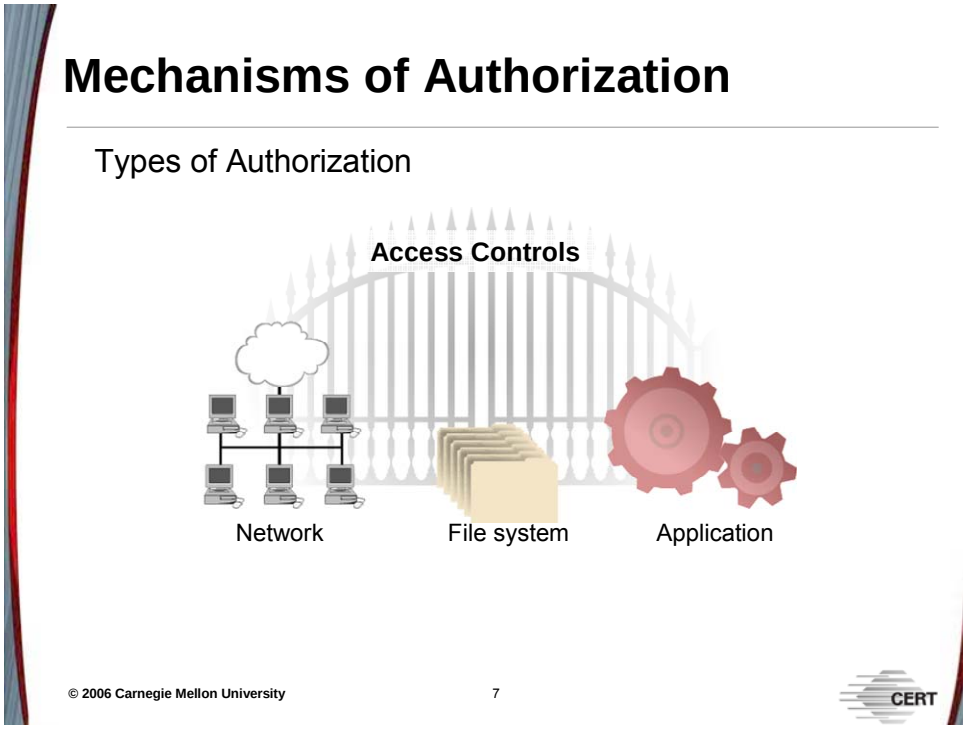
These regulations affect both government entities and private-sector organizations to varying degrees. As discussed earlier in the Compliance Management section of this course, the U.S. private sector must comply with laws such as Sarbanes-Oxley, Gramm-Leach-Bliley, and HIPAA, while government organizations face requirements laid out in FISMA, FIPS 200, and other federal standards.



5.1.2 Components of Authorization Management

The following are components of authorization management:

- **User –** an individual or host that requires access to protected resources. The identity of the user is established though secure authentication mechanisms (discussed in the Identity Management module).
- **Subject –** a file or object that a user can access. This definition includes files, folders, and other types of data records such as database records.
- **Resources –** any of a wide variety of network resources that a user may access. This category can include the network itself as well as network subcomponents such as printers and services.
- **Permissions –** rights granted to a user to access a given resource or subject.

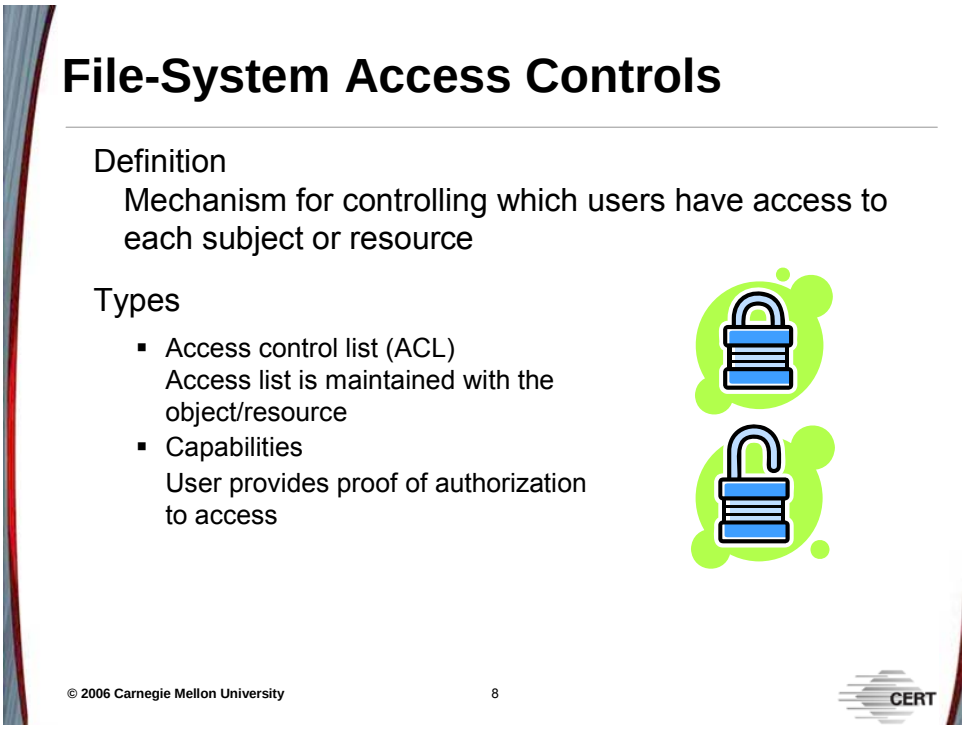


5.1.3 Types of Access Controls

There are several levels of authorization and access control. Access control involves restricting access to a particular asset so that only authorized users or groups can view, use, change, or delete the asset. In a networked environment, access can be controlled at each level of interaction: the file-system level, application level, or network level. Each level of access control presents different challenges and advantages, which we will discuss later in this module.

- **File-system access controls** are integrated into the operating system.
- **Network-level access controls** involve the utilization of network devices to regulate traffic across a network.
- **Application-level access controls** are permissions defined and applied at the application level to monitor and enforce rules for accessing applications and application data.

We'll next discuss these types of access controls in more detail.



5.1.4 File-System Access Controls

File-system access controls are an integral part of the operating system. In general, each operating system uses a different type of file system. The file system, in turn, may use a specific type of access control. Until recently, the primary means of file-system access control was access control lists (ACLs), but another model, the capability model, now is being integrated into newer versions of operating systems. Both models are discussed below.

5.1.4.1 Access Control List

ACLs are the most widely used means of file-system access control. They are used in most commercial or commonly used operating systems today. An ACL is a matrix that lists each user along with his or her privileges for accessing each specific object on the system. The ACL model creates a separate list that is stored for each object. A sample ACL for a file called “File 1” would look like the following:

User	File 1
John	rw
Alice	r
Howard	r

In this example, John would have permission to read (r) and write (w) to File 1. Alice and Howard would have only read (r) permission. In practical terms, File 1 might be a HR file

that contains important salary data. John might be the only person in the organization who should have rights to change this information, while Alice and Howard might work in payroll and need to access the records to make sure employees are paid appropriately.

Advantages

- The primary benefit of the ACL model is the ability to easily revoke a user's privileges to access a given object by editing the ACL of that object or resource.
- ACLs do not generally allow delegation of privileges. This can be either a benefit or a disadvantage, depending on the circumstance.
- The ACL model allows easy enumeration of all individuals who have access to a specific object.
- The ACL model can support file owners' ability to modify access controls.

Disadvantages

The ACL model limits access to resources based on role or job requirements. However, the model presents certain other disadvantages:

- Loopholes in permissions can occur when an object or resource is overlooked by a system administrator.
- The ACL model makes it very difficult to enumerate all of the objects or resources that a particular user is authorized to access.
- ACLs will not protect objects if an attacker has access to the physical drive. After all, if an attacker is able to bypass the operating system, ACLs will be useless.
- Allowing users to modify ACLs for objects they own can make it very difficult to monitor access controls and ensure that users are following organizational policy.
- To determine if all files have been configured with appropriate access controls, it may be necessary to manually check the ACL for each and every file or folder.

5.1.4.2 Capabilities

By contrast, the capability model stores a list of permissions for each user instead of for each file. This list is presented by the user, much like a token, to prove authorization when the user wishes to access a protected object. There are also implementations of the capability model wherein the system stores each user's access list in a centralized location. This practice can aid in the administration of access controls and can help minimize some of the potential drawbacks of the capability model.

The main difference between the two models is that in the capability model, access controls for a single user are kept in a single table, whereas in the ACL model, each file has its own list of authorized users. Think of the capability model as providing a full listing of all that a particular user is *capable* of accessing.

An example of a capabilities table is below:

User	File 1	File 2	File 3	File 4
John	rw	rwx	R	rw

This table shows that John has

- read and write access for File 1 and File 4
- read, write, and execute privileges for File 2
- read access for File 3

Systems that have capability-based access controls are used primarily in research or academia at present.

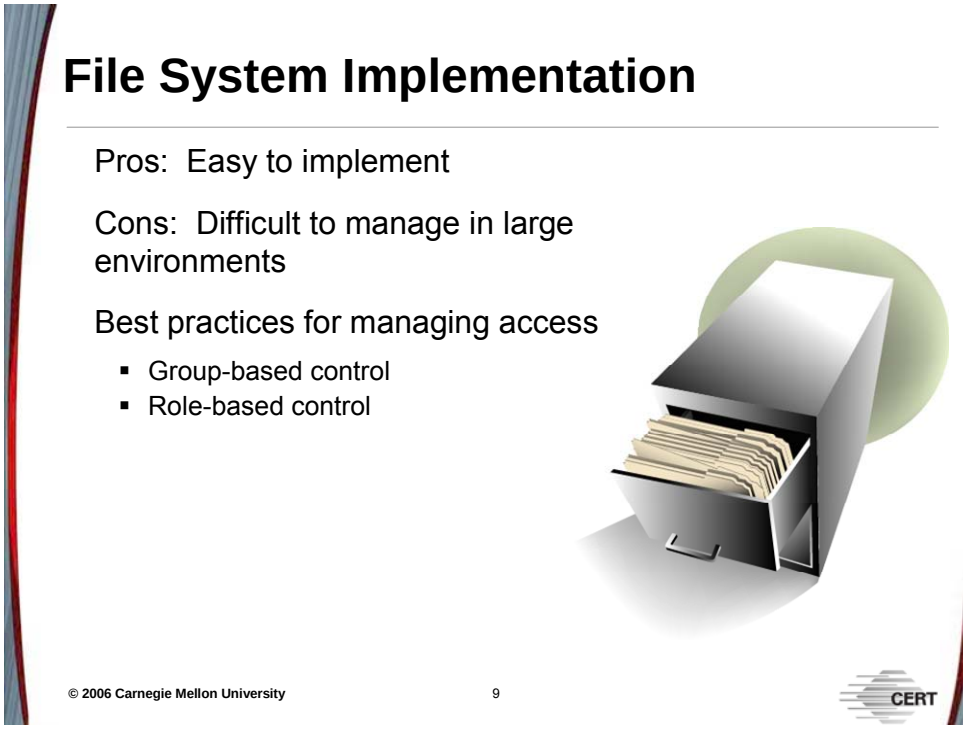
Advantages

- Using delegation to grant access can minimize the role of the system administrator; there is no need to edit the privileges associated with objects when a user with prior authorization directly delegates (grants) rights to another user.
- By default, the capability model grants a user or system only the ability to access (read) objects or resources when authorization has been explicitly given. In the capability model, this is referred to as *least privilege* access.
- Enumerating all of the objects for which a given user has authorization is a much simpler task than in the ACL approach.

Disadvantages

- Accountability and auditing can become difficult when using the capability model, since there may be no centralized list of users who have access to files and resources. This is especially true when delegation is used to grant access.
- It can be time-consuming to update the user list with new access permissions.

One of the few commercial operating systems to support the capability model is the widely used IBM AS/400, now called the iSeries.



5.1.5 File-System Implementation

5.1.5.1 Best Practices

There are many ways to implement and manage file-system access controls in an organization. Some strategies to make management easier and less likely to create loopholes are described below.

5.1.5.2 Group/Role Based Access Controls

By assigning classes of users to groups, you can reduce the amount of administration required for each individual user. Using a wider classification requires changing fewer access control lists when assigning or changing permissions on objects or resources.

Role-based management is another form of group management. Beyond general role groups such as human resources and marketing, you can also create groups based on more specific organizational roles. For example, groups might be created for administrators, account managers, or team members. Any job role requiring specific levels of authorization that differ from those of other roles would be a good candidate for role-based access controls.

Even when only a single user in an organization holds a specific role, it may be easier to manage access controls in a role-based manner. This approach will prove much more effective and efficient should the employee leave that particular role. The same privileges of the former employee could be applied to the new person who takes on the role, with no need

to change multiple objects or permissions. The new employee would only have to be added to the group for the privileges to be applied.

When implementing role-based management in an organization, it is important to use groups rather than assigning a single username to be used by all personnel performing a certain task. To see why, let's consider how this practice would apply to a system administrator role:

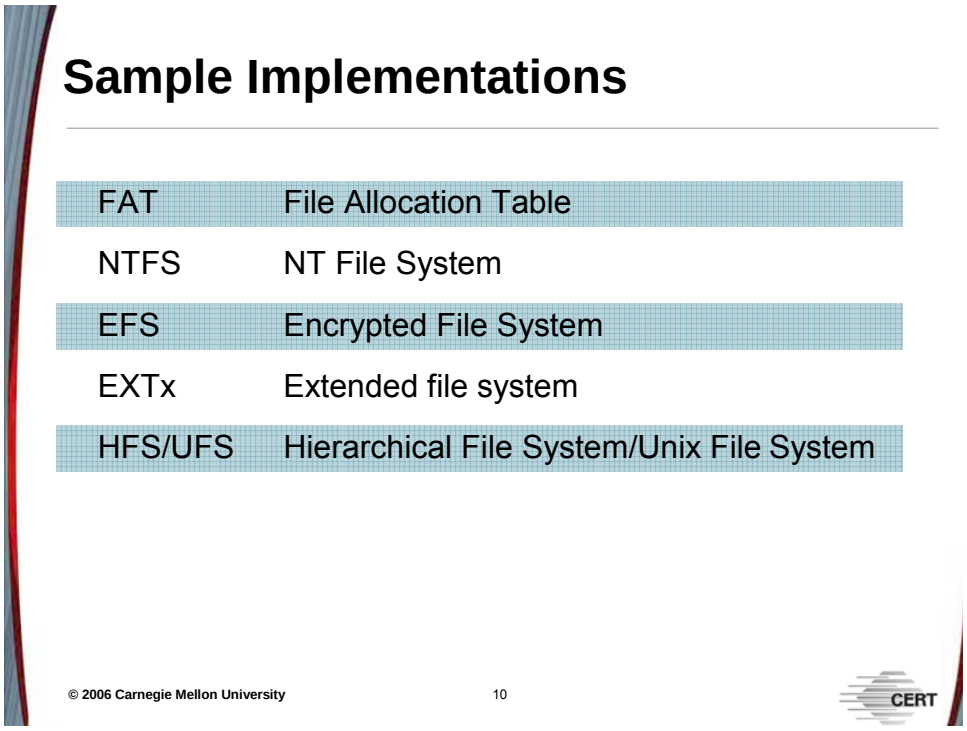
1. Create an account with administrative privileges and allow all administrative personnel to use this account when performing administrative duties.
   Several problems are caused by this scenario. Auditing is a problem since any number of employees could be using the account. Should any problems arise, it will be very difficult to determine which employee was logged into the account at the time of the activity in question. Additionally, if an employee leaves the company, security concerns may arise.
2. Create an employee-specific administrative username for every employee in the role of system administrator and then assign each username to the "Administrators" group.
   This scenario allows full auditing of each individual's activity.

5.1.5.3 Policy

Every organization should develop and follow an information security policy that establishes procedures and protocols for authorization management. Every tactical-level security policy should define the specific access controls that will be used to protect essential information and resources. The process for changing user and group permissions also should be clearly defined in policy.

Additional Practices

- Remove generic or default user groups such as the "Everyone" group in Microsoft Windows.
- Apply permissions to folders rather than to individual files when appropriate. This can reduce the amount of administration required to maintain security.
- Restrict access to system or configuration files to administrators only.
- Be careful with inheritance of permissions. Make sure new files or sub-folders created within a directory inherit the same permissions as the primary folder in which they reside.
- Create a separation of duties. This practice is key to maintaining control of access protection and keeps checks and balances in place. For example, the individual responsible for maintaining access rights should not be able to delete or edit access logs.



5.1.5.4 Implementations

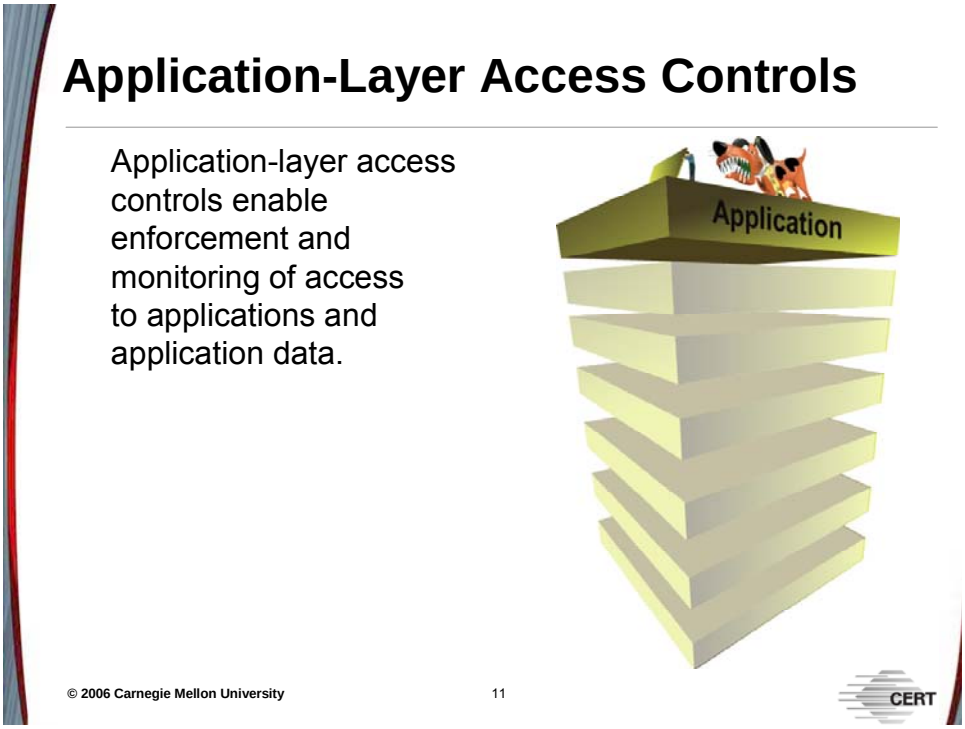
Each operating system generally has a specific file-system type. There is a large variety of possible types, but only a few are widely used. The most common file systems you will see are described below.

FAT – File Allocation Table – This was the primary file system for the Windows operating system up through the release of Windows ME. The FAT file system is relatively basic and is supported on most operating systems, making it a popular choice for disks that must be read by both Windows and Linux/UNIX hosts.

NTFS – NT File System – This is the replacement for the FAT system in the Windows operating system from Windows NT through XP and Windows Server 2003. The NTFS file system is not compatible with other operating systems. Its primary benefit is its support for Encrypting File System (EFS).

Extx – Extended file system – This was the original Linux file system. There are three different versions: Ext, Ext2, and Ext3. Ext3 is now the default file system in Debian, Fedora, and RedHat Linux distributions.

EFS – Encrypting File System – This is a companion to the NTFS file system used in Microsoft Windows. It allows files stored on an NTFS partition to be encrypted using a combination of symmetric and public key cryptography. This adds another layer of security to aid in maintaining confidentiality should an attacker gain physical access to the hard drive.



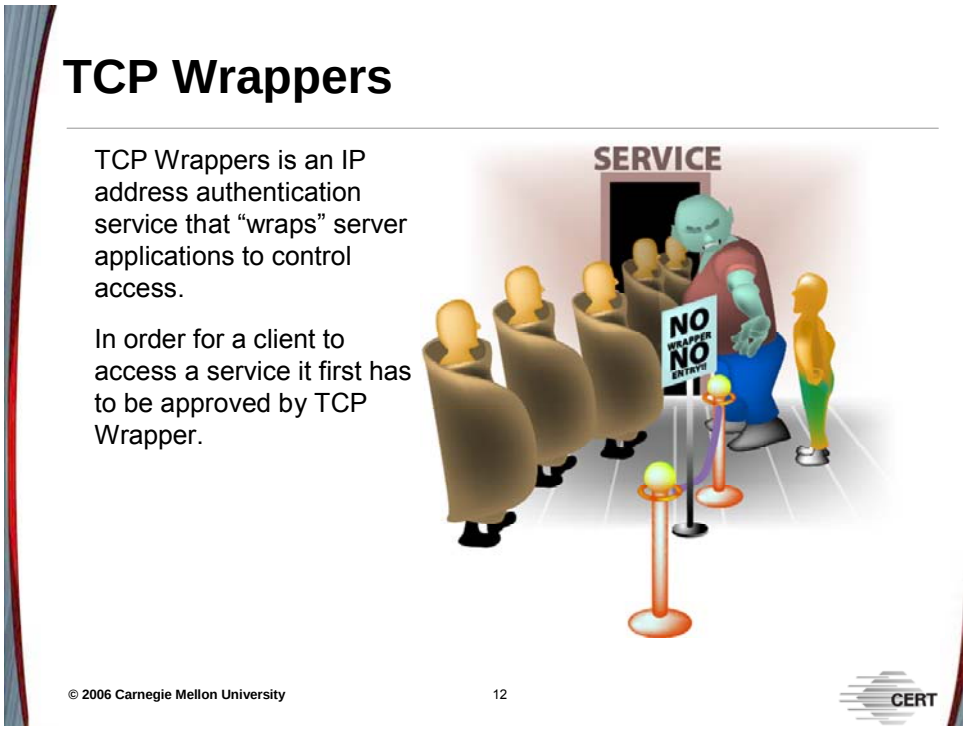
5.2 Application-Layer Access Controls

Authorization management means more than just controlling access by individuals. It also means controlling authorization of traffic. At the application level, we can monitor and enforce rules for traffic flow and integrity. This means we can ensure that packets sent on the network follow the correct technical specification for the communication protocol being used.

The application layer is the networking component that is closest—and most familiar—to end users. It manages their interactions with lower-level network protocols. This layer involves the applications that a user interacts with to read, write, or modify data, including basic applications for Web browsing and sending and receiving mail as well as more complex applications for database management and performing financial transactions.

For example, by using an application gateway, you can make sure all packets being sent to port 25 are actually mail packets and not other types of packets being routed to port 25 just because an attacker suspects he can enter through this commonly available port.

Application-layer mechanisms are often referred to as logical access controls. This type of control is often a part of an application but also can be a separate component added to a system. This layer of security is meant to enhance security beyond that available in the lower levels of the TCP stack (the layered implementation of the TCP protocol).



5.2.1 TCP Wrappers

TCP Wrappers is a filter package[9] for UNIX/Linux systems that will monitor incoming requests for a specific inetd (internetworking) service, such as ftp, telnet, rlogin, etc. Once implemented and configured to protect a service, TCP Wrappers will analyze each incoming request to determine if the host attempting to establish a connection has access privileges. If it does, the request is passed on to the service. If it does not, the connection request is denied.

TCP Wrappers also provides the additional benefit of limiting the information that is permitted to leave your network. This can be important in protecting your network from malicious scanning designed to locate open ports and enumerate the services running on your network. TCP Wrappers allows you to run services while granting access to those users who have already been approved.

TCP Wrappers also provides a variety of other services related to access to hosted services. This includes the ability to filter requests and route them to different hosts based on the origin of the request, as determined by the IP address of the requesting host.

[9] Wietse Venema, the creator of TCP Wrappers, outlines the benefits and functionality in his paper, *TCP WRAPPER: Network monitoring, access control, and booby traps*. Available through http://www.usenix.org/publications/library/proceedings/.

5.2.1.1 Advantages of TCP Wrappers

TCP Wrappers present several benefits:

- There is no need to modify existing software or configuration files.
- The default configuration is such that the software can be installed “out of the box” on most UNIX implementations.
- There is no impact on legal users.
- The wrapper program does not exchange any data with network client or server processes, so the risk of software bugs is extremely small.
- It is suitable for both TCP (connection-oriented) and UDP (connectionless) services that are managed by a central daemon process such as inetd.
- The optional access-control facility can be used to shield off open systems. Network routers can perform a similar function, but they seldom keep a record of unwanted traffic. On the other hand, network routers can be useful in blocking access to ports that normally cannot be covered with wrapper-like programs, such as The Portmapper, NIS, NFS, and X server network ports.

5.2.1.2 Disadvantages of TCP Wrappers

High interaction for Source IP control – Access control via source IP address provides a minimum level of security. An IP address can easily be spoofed (i.e., “fooled” by another client masquerading as a trusted machine). In deciding to trust connections from outside hosts, you should be reasonably certain that those hosts are operated in a secure fashion and have not been compromised. Source IP controls can be more manageable in smaller organizations that have fewer connections from the outside world, but this doesn’t entirely mitigate the IP spoofing problem.

Minimal packet analysis – TCP Wrappers performs a limited inspection of packets, mainly by examining header information and either rejecting or passing on the packet to the destination service.

Increasingly, TCP Wrappers is being replaced with the newer extended internet daemon (xinetd), which provides similar protection of services. It is important to note that performing access control using source IP controls is not the best or most preferred means of achieving security. A better method of protecting services is to use an in-line application gateway or proxy in front of the protected service. Better security can also be achieved by authentication of either a user or host to enforce access controls. This can be done using a variety of mechanisms, but for host authentication, IPsec and Secure Sockets Layer (SSL) can be useful tools.

Application Gateway / Proxy Server -1

A proxy server is a middleman between a client and a service.

A proxy server determines the following:

- Is the request for a valid service/host on the LAN?
- Is the request coming from an authorized host?
- Does the packet meet the specifications for the port/protocol for which it is attempting to establish a connection?
- What is the payload of the packet?
- Is this an authorized command?

CERT

5.3 Authorization Management Implementations

5.3.1 Application Gateway/Proxy Server

A proxy is a device or service that sits between the Internet and the internal LAN. When an incoming request comes into the network, it will be sent to the proxy. The proxy will analyze the request to determine whether it is valid. If the request meets the conditions being monitored, the proxy will pass it to the destination host.

Proxy servers can analyze a request to determine several conditions:

- Is the request for a valid service/host on the LAN?
- Is the request coming from an authorized host?
- Does the packet meet the specifications for the port/protocol for which it is attempting to establish a connection?
- What is the payload of the packet?
- Is this an authorized command?

The term “proxy server” is synonymous with “application gateway” or the variety of other terms used to describe such applications. An application-level gateway, sometimes referred to as an application-level proxy, is considered the most intelligent firewall available. This does not mean it is right for all organizations, however. This issue is discussed later in this module.

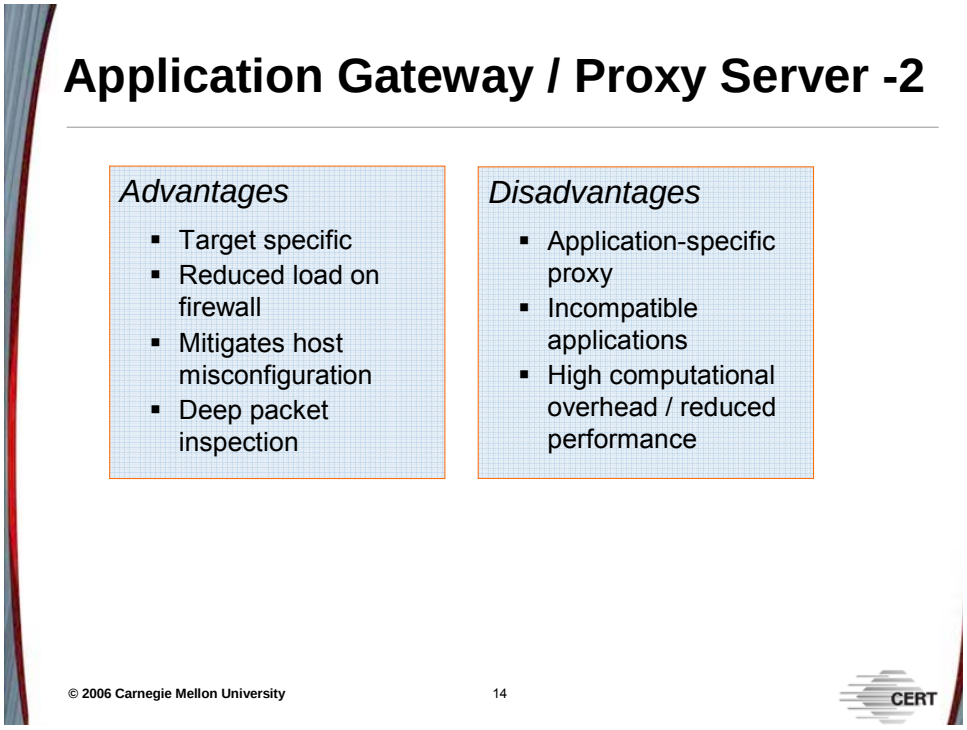
The role of an application gateway is to enforce the integrity of the connection being established and the communication that will ensue once a connection is made. An application gateway can also limit the type of requests allowed in and out of a network. In the instance of an FTP connection, the application gateway could restrict “put” commands. In this scenario, a user who connected to a host via the FTP protocol would be able to download files from the FTP server but would be prevented from uploading files.

5.3.1.1 Examples of Proxy Applications

Below are a few of many popular and widely used proxy server applications:

- Microsoft ISA server – deep packet analysis, proxy server for Microsoft Infrastructures
- Squid – Unix Web proxy application: www.squid-cache.org

A wide variety of vendors provide proxies. The cost can vary greatly, depending on the number of services you intend to support using a proxy.



5.3.1.2 Advantages of an Application Proxy

Target specific –You can perform access controls on an as-needed basis and as defined for a specified host, rather than applying them universally in advance. This allows for increased granularity in defining access controls.

Reduced load on firewall – In a layered defense, an enterprise should have both a firewall and a proxy. When highly critical services are routed through a proxy, the rules on the firewall are much simpler, since it can then forward the protected services to the proxy server in lieu of routing them to individual hosts.

Can reduce effect of misconfiguration on individual host – A proxy between the Internet and critical or publicly available services can minimize the effect of improperly configured individual services.

Deep packet inspection – An application gateway has the ability to examine all seven layers of a packet, including the payload. Traditional packet-filtering mechanisms, by contrast, only look at the header data of a packet. For example, a basic packet filter would look at a packet header and determine the packet's destination port. It would then look at its filter rules and determine whether that was an allowed port. If so, it would allow the packet into the network. Proxies can look at the actual content of the packet.

Support of enhanced authentication – Some proxies are capable of increased levels of authentication. This provides more security than simple source IP controls as implemented

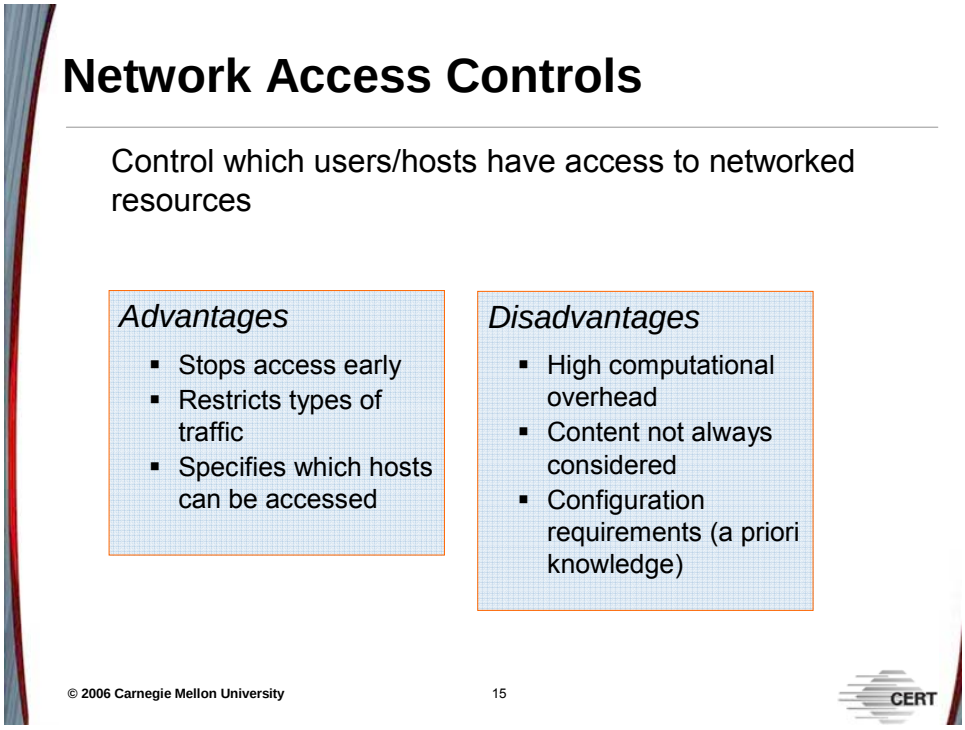
by, for example, TCP Wrappers. One example is Microsoft's ISA Server. It can query the active directory to authenticate and authorize Web requests of individual user accounts.

5.3.1.3 Disadvantages of an Application Proxy

Application-specific proxy – A specialized proxy is needed for each application or service that is to be filtered through the proxy. This can require extra work and configuration, depending on the number and types of services to be protected.

Incompatible applications – It is possible that a service or application may not be compatible with a proxy. In this case, the only options are to allow all traffic from this service or application into the network or deny all such traffic. This can leave you either with a large security hole or without the ability to provide a specific service.

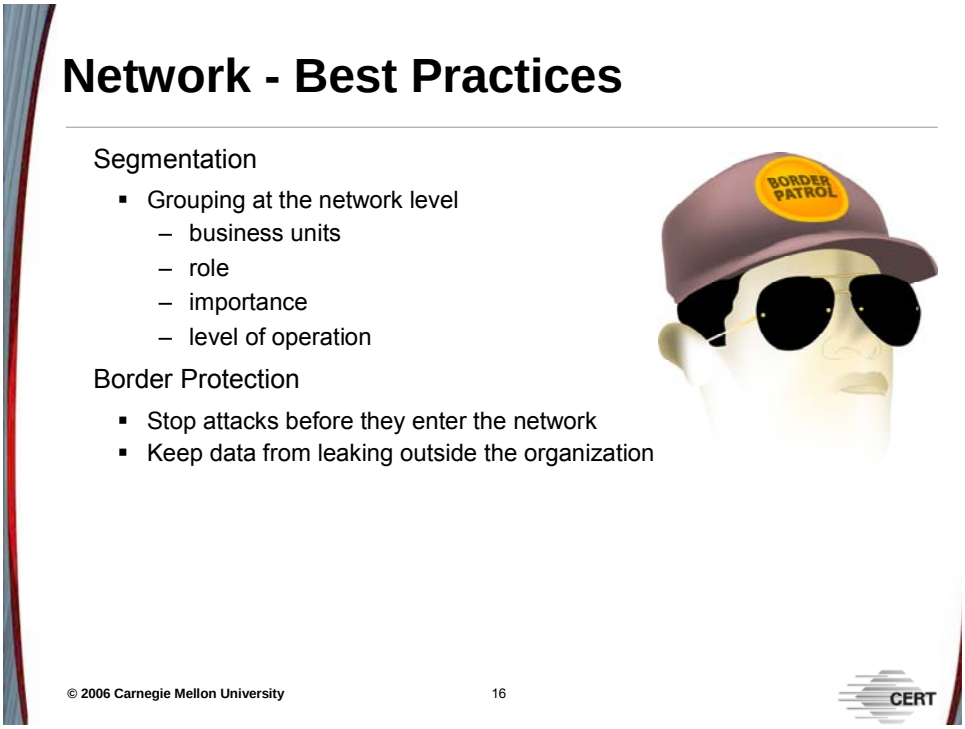
High computational overhead/reduced performance – This can result from the in-depth analysis and filtering involved. The full analysis that takes place in an application proxy can cause delays since the application will be passing communications back and forth between the two parties. This, in effect, creates two connections for every single communication between the inside network and the outside world. For example, running mail transactions through an application relay will create a connection from the client on the Internet to the gateway and another connection from the gateway to the mail server.



5.3.2 Network Access Controls

The goal of network access controls is effectually the same as that of file-system access controls: to limit who and what has access to certain resources. The main difference between the two methods is how you go about achieving this goal. As part of a complete Defense-in-Depth strategy, you should employ multiple mechanisms to ensure security is achieved and maintained even in the event of some system failure.

Network controls are concerned with limiting or controlling traffic flow on the network. Such controls might involve connections between a host such as a Web server and a transaction or database server, for example. Alternatively, they might involve limiting which hosts are allowed to communicate with each other or with machines on a separate segment of a corporate LAN. Network control also includes allowing approved external connections into the network while rejecting those that do not have prior authorization to access internal network resources.



5.3.2.1 Segmentation

Segmentation at the network level can be compared to the use of groups or role-based controls at the file-system level. To minimize unnecessary and unauthorized connections to a host and traffic flow between hosts, you can divide your network into multiple segments. These divisions can be made based on business group, network resource needs, level of operation (i.e., production servers versus testing servers), or importance of information being stored on hosts.

There are several ways to segment your network. The two primary means are routers and firewalls. A router is a device that connects two networks and helps directs network traffic from its source to its destination, finding the quickest path among myriad intermediate relay points. A firewall is a hardware or software device that filters network traffic to ensure authorized traffic reaches its destination while unauthorized traffic is denied.

5.3.2.2 Border Protection

Federal regulations note the importance of boundary protection. NIST Special Publication 800-53, which details the minimum security controls any federal agency should employ, defines border protection as

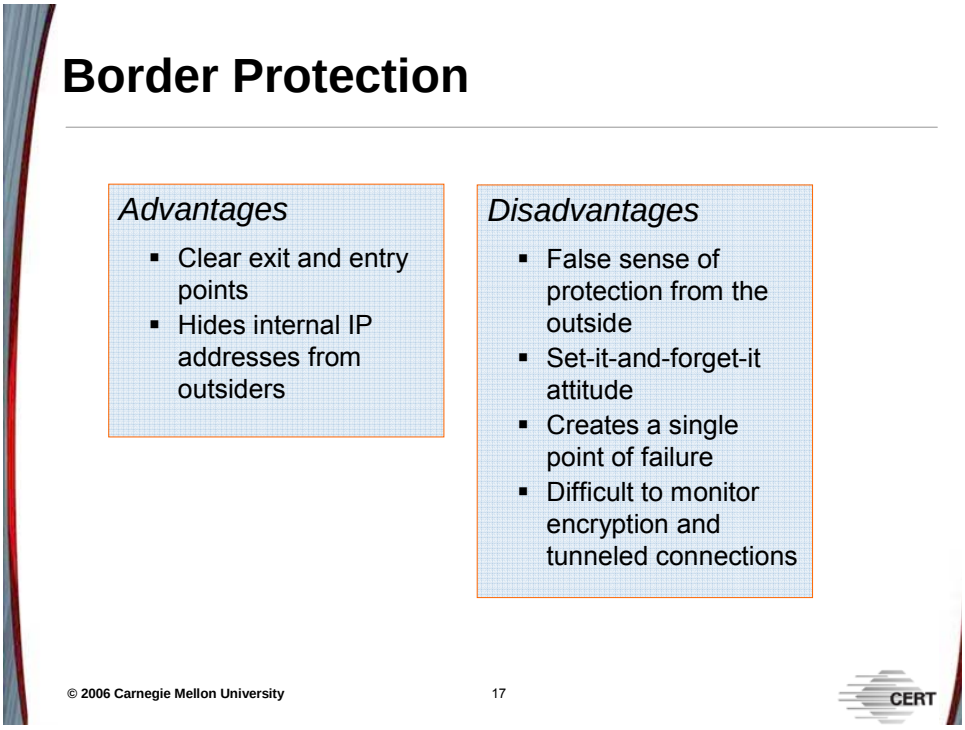
> Any connections to the Internet, or other external networks or information systems, occur through controlled interfaces (e.g., proxies, gateways, routers, firewalls, encrypted tunnels). The operational failure of the boundary protection mechanisms does not result in any unauthorized release of information outside of the information system boundary [NIST 05].[10]

For government agencies, we also suggest the following:

- a multi-tiered defense to mitigate potential failure of primary border security devices
- a "fail closed" implementation in which the failure of a device does not leave holes in the protection of internal information from external exposure

However, border protection is not a panacea. On the next slide, we'll provide more details about the advantages and shortcomings of border protection.

[10] See NIST SP 800-53: *Recommended Security Controls for Federal Information Systems*. February 2005. http://csrc.nist.gov/publications/nistpubs/800-53/SP800-53.pdf.



Border Protection Positives

Clear exit and entry points – Having a firewall and a clearly defined gateway separating the inside from the outside allows for easier monitoring and access control.

Hide internal IP addresses from outsiders – A firewall can be configured as a proxy or can perform Network Address Translation (NAT) to mask internal host IP addresses from outsiders. In this scenario, the only IP address that would be sent to remote hosts would be that of the gateway.

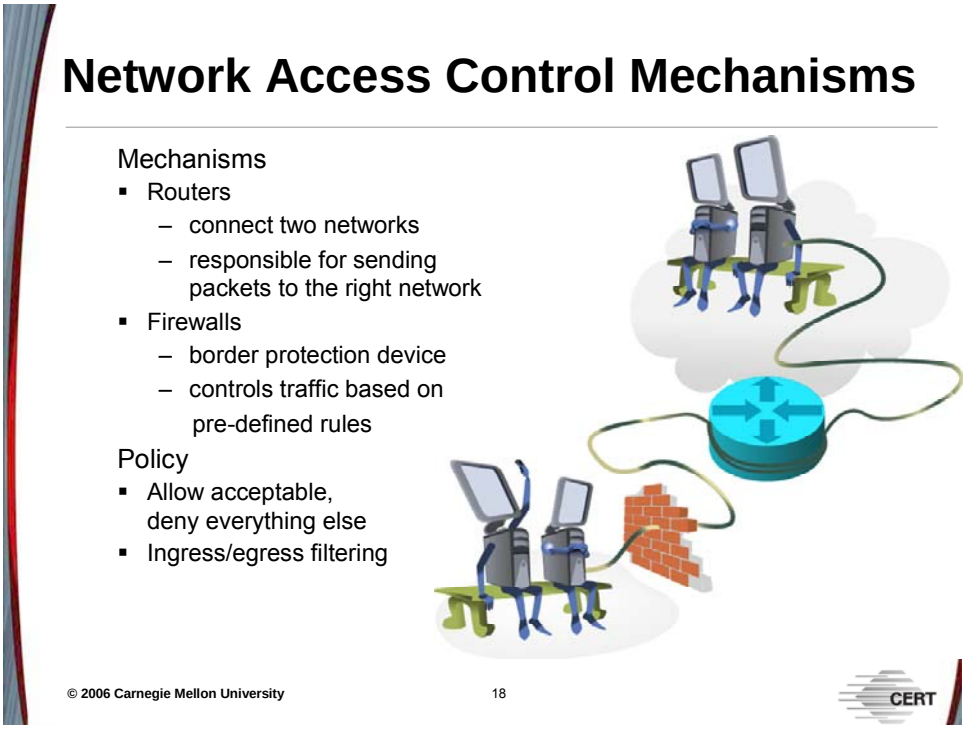
Border Protection Negatives

False sense of protection from the outside – Often, an organization will mistake the presence of a firewall and other network measures for true security. While these network measures are important, they should be reinforced with other mechanisms to ensure complete Defense-in-Depth. Understanding and implementing the other components of Defense-in-Depth will enable an organization to achieve much higher levels of security.

"Set it and forget it" attitude – Many organizations fail to monitor and change their firewall rules to match the changing needs of the enterprise.

Creates a single point of failure – Sometimes an organization has a single connection to the Internet routed through a firewall. If that firewall fails, clients inside the network will be unable to connect to hosts on the Internet, and remote hosts will be unable to establish connections to services that are hosted internally.

Difficult to monitor encryption and tunneled connections – Firewalls and filters are only capable of making decisions based on the data they can process. Encryption obscures data, making decision-making difficult. It is possible to control which ports are available and allow traffic through based on its source and destination ports; however, this approach may restrict the set of various encrypted applications that can communicate across the firewall.



5.3.2.3 Routers

As stated earlier, routers and firewalls are two network access control mechanisms. A router is a "gateway" device, located between two networks; it routes packets from one network to another network until the packets reach their destination. A router is a layer 3 device, which means it looks at the data in the network layer of a TCP/IP packet. A router can serve as a gateway to the Internet or as a means of dividing an internal network into different segments.

Routers also have the ability to perform a limited amount of packet filtering. At layer 3 in the TCP/IP stack, a router will examine the source and destination IP addresses of packets and make routing decisions based on rules defined by the network administrator. This can be a highly intensive activity, so it may not be appropriate for a large network, but for a smaller network it may be suitable. For segments requiring higher levels of security and access control, it is a good idea to place a firewall behind the router so that the router can perform its primary job of routing packets and the firewall can perform the more intensive activity of traffic monitoring. See below for more information about firewalls.

Although trying to enforce firewall-style access control lists on the router is a bad idea, it *is* good practice to implement ingress and egress filtering. These are simple mechanisms and should be utilized in every organization.

Egress filtering refers to filtering packets leaving your network. The goal of egress filtering is to prevent traffic from leaving your network if it does not have a source address that is not assigned to your internal network.

For example, if your network IP address range were 10.0.2.0/24. Egress filtering would block any outbound packet that has a source address other than the addresses in this space. A packet with a source IP of 12.3.4.5, for example, would be dropped by the router before leaving your network.

Ingress Filtering is the monitoring of packets coming into your network. The purpose of ingress filtering is to disallow packets that have a source IP address within your internal network.

For example, if your network IP address space were 10.0.2.0/24, Ingress filtering would block any incoming packet that has a source address within the 10.0.2.0/24 address space. A packet with a source IP of 10.0.2.5, for example, would be dropped by the router before it was allowed onto your internal network.

Disadvantages of using a router for packet filtering include the following:

- There is no user authentication. (A router can typically only filter at the network and transport Layers, whereas a firewall can filter up to the application Layer.)
- Source IP filters or “blacklists” over time are difficult to manage.
- Filtering of many packets by the router will degrade the overall performance of the device, since it is responsible for routing all packets entering and leaving the network.

5.3.2.4 Firewalls

A firewall is a device to monitor incoming or outgoing traffic and block packets based on rules defined by the administrator. Firewalls can be used to provide different types of security. The two main functions of firewalls are 1) internal segmentation and 2) protection from and control of external connections to the internal network. Both functions are important in developing an authorization management strategy and an overall Defense-in-Depth posture.

Internal segmentation serves to separate portions of the internal network from the rest of the network. This should be done when a host or set of hosts has stricter security requirements than the rest of the network.

For example, say a transaction server has a critical database stored on it. The server contains no information that must accessed manually by an employee; instead, its primary purpose is to manage transactions passed to it by a Web-based application server. Therefore, you could place the transaction server behind a firewall and only allow connections from the Web server.

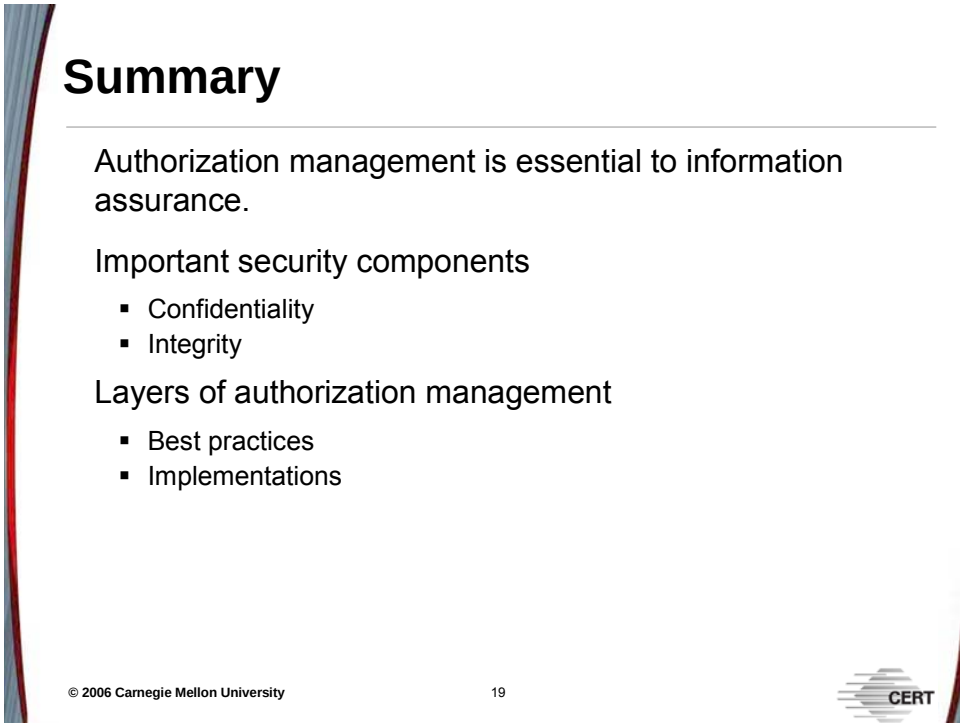
Another example involves a set of hosts on a subnet responsible for viewing, editing, and storing sensitive information. This could be a group of systems responsible for transaction data in a financial institution or a group of systems responsible for working with sensitive government information. In this case, a firewall could enforce network access controls and

support a limited one-way flow of data to the more secure subnet, while also prohibiting data transfer in the opposite direction so that secure machines could not send data to the less secure portion of the network.

It is important to remember that a firewall is only as good as the rules on which it is based. These rules are defined by the system administrator and should be reviewed and monitored with regularity. In reviewing firewall configurations, it is a good idea to look at the logs that are generated.

Various levels of logging are possible and should include at a minimum [Allen 01]:

- packets that are denied upon arrival at the firewall system
- packets that are denied upon departure from the firewall system



Summary

It should be evident by now that authorization management is an essential component of information assurance, not a stand-alone measure. All of the concepts discussed here are meant to be complemented by additional security measures. For example, authorization management relies on policies to define which elements of the architecture are important and the need for access to those objects. Authorization management is also only as strong as the organization's authentication measures. Strong authentication, when combined with properly implemented authorization measures, increases enterprise security by orders of magnitude. Remember that you can only authorize access to a resource if you can authenticate a client or a user. If strong authentication is not in place, then your security is quite weak.

If you can implement the integrated concepts discussed above into your security strategy, you will have taken a giant step toward true Defense-in-Depth.

Review Questions -1

1. Define the security properties enforced by authorization management.
2. What are the layers of authorization management?
3. Explain how an ACL works. What are the downfalls?
4. Name at least three best practices for file-system-level implementation.
5. What is the main benefit of application-level implementation?

Review Questions -2

6. Explain the downfalls of source IP controls.
7. Explain the benefits of a proxy server.
8. What is the primary purpose of network access controls?
9. What are the primary benefits of controlling access at the network level?
10. List three best practices for network access control.

Module 6: Accountability Management

This module describes accountability implementation methods such as log management, network monitoring, and intrusion detection, as well as best practices for achieving accountability.

Instructional Objectives

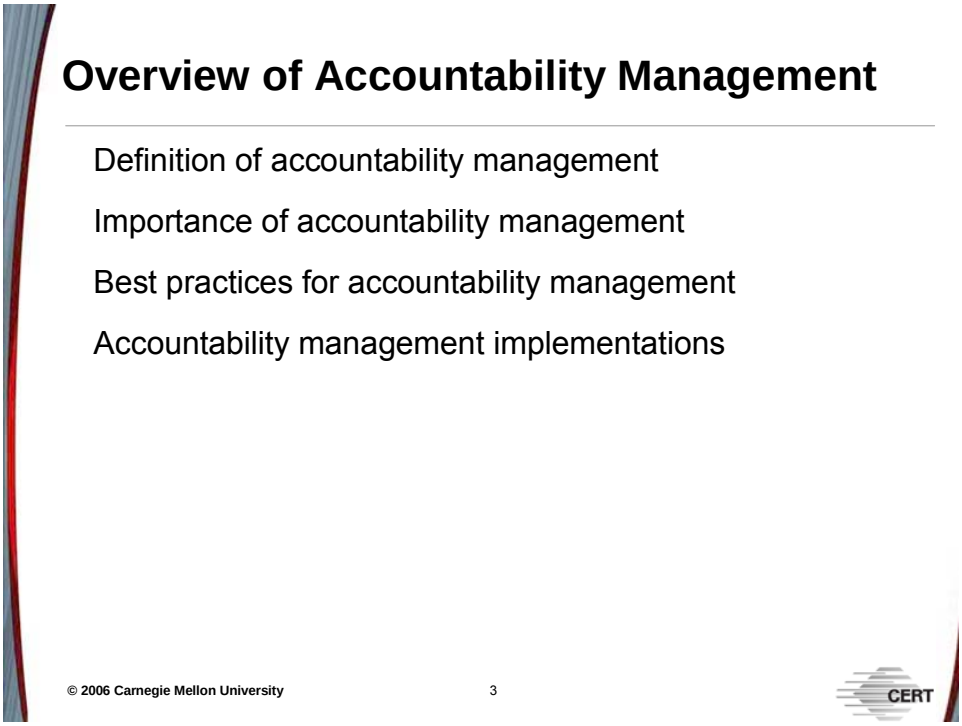
Upon completion of this module, students will be able to

- Define accountability management
- Know the importance of accountability
- Identify best practices for accountability management
- Recognize accountability management implementations

CERT

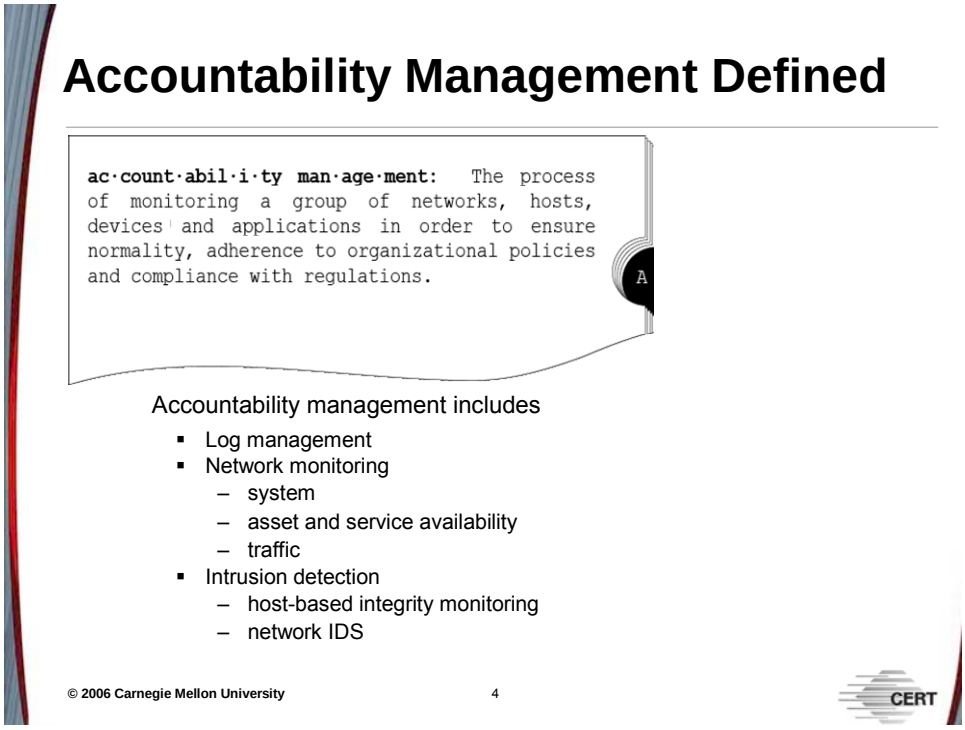
This instructional module will enable students to complete all of the above learning objectives.

6 Overview of Accountability Management



The purpose of this module is to provide a better understanding of accountability management and its effect on an organization's overall information security posture. This module has four main objectives:

1. Accountability management will be defined and the processes it involves explained.
2. The importance of accountability management will be explained. In addition to adding greater visibility to an organization's IT processes and assets, accountability management plays a vital role in keeping organizations in compliance with regulations.
3. Accountability management best practices will be presented. Careful planning must take place before an organization begins to implement accountability measures. By utilizing best practices, an organization can maximize the effectiveness of its implementation.
4. Various types of implementation will be discussed. Each type plays an important role in the overall accountability of an organization's assets.



6.1 Defining Accountability Management

Accountability management can be defined as the process of monitoring a group of networks, hosts, devices, and applications in an attempt to ensure normalcy, adherence to organizational policies, and compliance with regulations. It is difficult to summarize accountability management in a few sentences because of its complexity and its role in seemingly disparate processes throughout an organization. An important aspect of accountability management is its overall enhancement of insight into IT processes, activities, and assets. This in turn allows greater visibility into the organization as a whole.

Accountability management methods consist of both real-time and historical analysis. Real-time analysis enables technical staff to detect anomalous or suspicious activity as soon as possible. Mitigation steps then can be quickly put in place to prevent a small incident from turning into a large one. On the other hand, historical analysis provides an audit trail for prior activities, so that records can be examined to search for events that led to a particular incident. Historical records can also be used to definitively prove whether or not an event, activity, or behavior took place.

The scope of accountability management activities is admittedly broad and includes the five areas discussed below.

6.1.1 Log Management

Most applications and devices in an organization's IT environment produce some type of log messages documenting their activities and the interactions they have with other entities.

Recording, aggregating, and analyzing these messages can provide a highly detailed view of activities that have occurred.

6.1.2 Availability Monitoring

Network monitoring involves observing network components such as hosts, services, and devices. This type of monitoring enables an organization to closely watch key network components and assess the overall health of its networks.

6.1.3 Traffic Monitoring

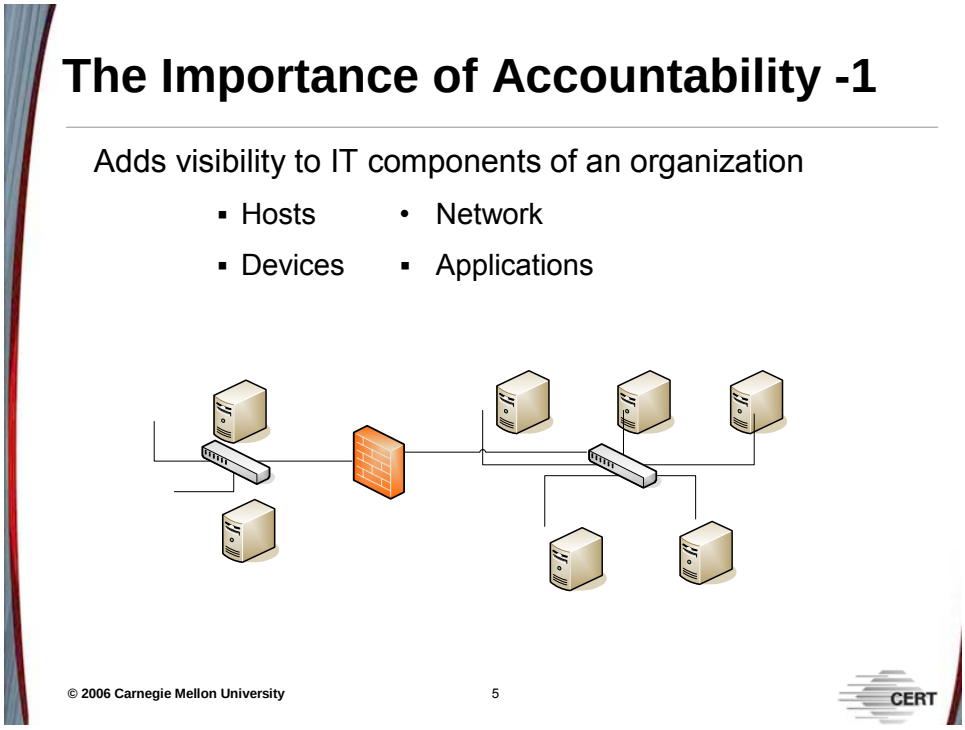
Traffic monitoring consists of examining packets being sent and received on the network. Whereas network monitoring involves observing the senders and receivers, traffic monitoring involves analyzing the actual data flows. Traffic monitoring can also involve watching the amount of data or types of data being sent and received.

6.1.4 Host-Based Integrity Monitoring

This type of monitoring focuses on the integrity of a particular host, ensuring that alterations such as file-system changes will not go unnoticed. Host-based integrity monitoring can provide more granular detail about a host than traffic or network monitoring. For example, if an attack is detected on a network, host-based monitoring enables you to assess the attack's effect on the target host.

6.1.5 Network Intrusion Detection

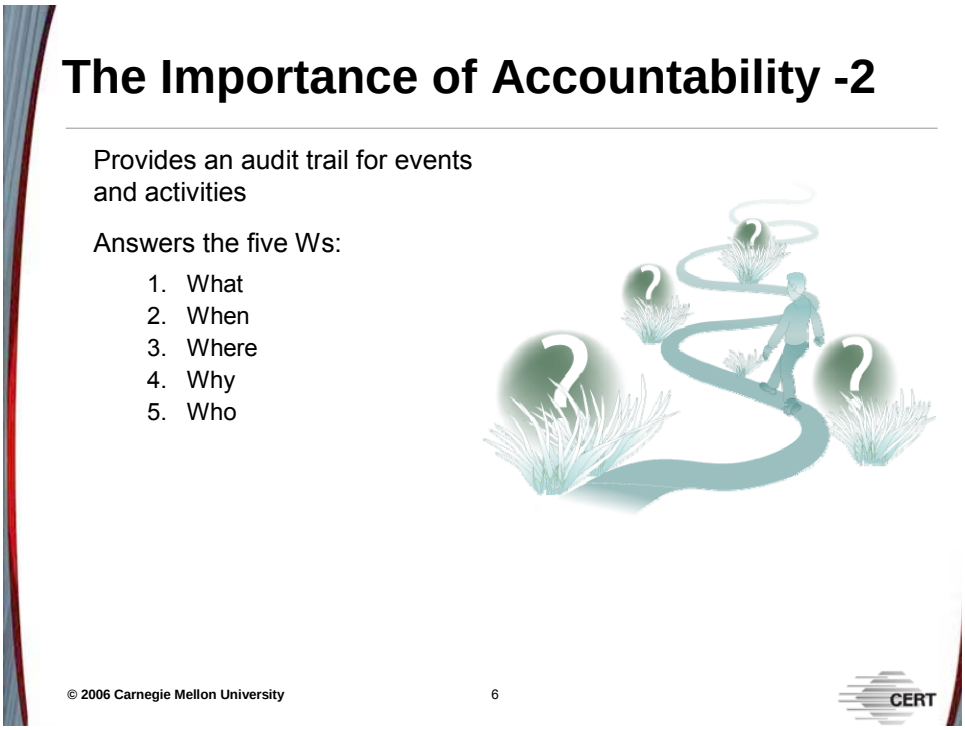
Network intrusion detection involves identifying attacks, suspicious behavior, and anomalous activities on a network. This type of monitoring is typically performed by a sensor that examines activity on a network. Network intrusion detection can be further broken down into signature-based detection, anomaly-based detection, or a hybrid of both. We will discuss intrusion detection in more detail later in this module.



6.2 The Importance of Accountability Management

6.2.1 Greater Visibility

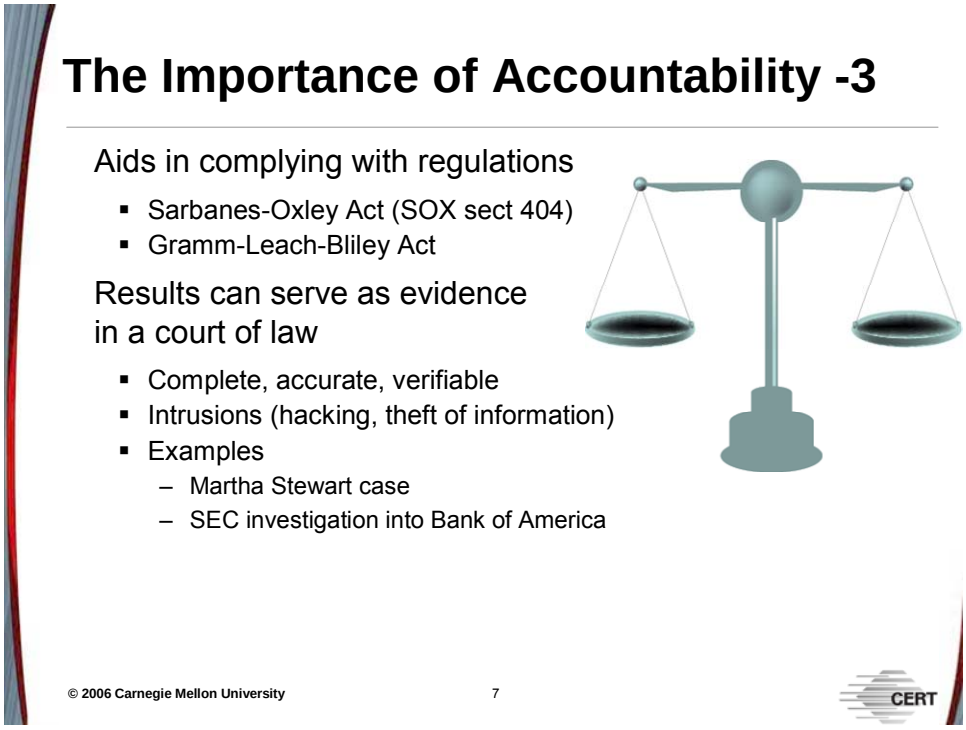
A key aspect of accountability management is its enhancement of visibility into an organization's IT components. Different monitoring techniques add visibility to different components on a network. For example, network monitoring provides insight into the availability of network hosts, devices, and services. Traffic monitoring reveals protocol and application activity on the network. Finally, log file collection provides granular detail about all of the previously mentioned hosts, devices, protocols, applications, and services. Each of these methods alone only paints a partial picture of network activity, but when combined, they provide a complete, well-documented representation.



6.2.2 Accountability Creates an Audit Trail

The visibility that accountability management brings to an organization's IT components makes it possible to uncover audit trails for specific events and activities. When accountability management is implemented correctly, an explanation of what, when, where, why, and whom results. Through accountability management, you can do the following:

- Determine, through logs and monitoring, which components of the IT infrastructure have been affected by an incident. This enables you to understand the scope of the incident.
- Establish at what time an incident occurred, so you can refine your investigation to focus on a specific period of time, and aid in correlating events.
- Determine where the incident might have originated; for example, whether it was caused by an event that occurred inside or outside of the organization's network.
- Understand why a particular incident occurred. Historical information may show events leading up to an incident or a recurring pattern of behavior.
- Identify the party responsible for initiating an action or accessing an asset. It is important to understand that a responsible party may not be a person; it could be a device or a software component.



6.2.3 Regulatory Compliance

6.2.3.1 The Sarbanes-Oxley Act

Within the past few years, accountability management (AM) has become crucial in ensuring compliance with federal regulations. The increasing relevance of AM in this arena is largely due to the fact that business processes are becoming more intertwined with information technologies. The Sarbanes-Oxley Act (SOX) of 2002, which we discussed in the Compliance Management section of this course, requires corporations to have sufficient accountability management measures in place. A number of sections in SOX are directly relevant to accountability management and are reviewed in more detail below.

Section 404: Management Assessment of Internal Controls

Section 404 of SOX pertains to management's responsibility for ensuring that sufficient internal controls for financial reporting are in place. Additionally, management must report on the effectiveness of these controls. Since financial reporting processes are performed using IT systems, managers must rely on accountability management processes to provide evidence that their controls are in place and to quantify their effectiveness.

Section 409: Real-Time Issuer Discloser

Section 409 requires companies to disclose information quickly regarding material changes in their financial condition or operations. In order for companies to make timely disclosures to investors and other relevant parties, they need to ensure the availability of their internal

and external reporting mechanisms. This means reporting mechanisms must be properly monitored, which calls for accountability management. Also, if material changes in a company's financial condition or operations are detected, management must be alerted in a timely fashion. This means a company must be able to aggregate relevant information from its IT systems, monitor it, and create an alert when a certain threshold is met. This is another process best handled by accountability management.

Section 802: Criminal Penalties for Altering Documents

Section 802 of the Sarbanes-Oxley Act addresses the issue of document integrity, which is relevant since a large percentage of companies' documents are now in electronic format. According to the Act, companies are required to maintain their audit and review paperwork for five years. Other IT-related issues that fall under this section are policies for record retention, protection, and destruction, online storage, and audit trails. Ultimately, the organization must be able to prove that its record management processes comply with the regulations outlined in Section 802, which may include the maintenance of log files and monitoring of storage devices. For additional information regarding the Sarbanes-Oxley Act, refer to the Compliance Management module.

6.2.3.2 The Gramm-Leach-Bliley Act

Whereas the Sarbanes-Oxley Act is aimed at public companies in general, the Gramm-Leach-Bliley Act (GLBA), also discussed earlier in the Compliance Management module, specifically focuses on financial services firms. To comply with this law, financial institutions must employ audit and oversight procedures to ensure customer information remains protected from unauthorized access. Processes such as log file collection and analysis, network monitoring, and intrusion detection all can be used to comply with the GLBA. Additionally, the Federal Financial Institutions Examination Council (FFIEC) has published an interpretive guideline for the GLBA entitled *Interagency Guidance on Response Programs for Unauthorized Access to Customer Information and Customer Notice.* In this guideline, the FFIEC specifically cites the importance of using system logs as an audit trail to determine whether information has been inappropriately accessed. If financial institutions are found to be in non-compliance with the GLBA, they may incur civil penalties of up to $100,000 per violation; officers and directors may also be personally liable for penalties of up to $10,000 per violation.

6.2.4 Legal Issues

Another reason why accountability management has become so important is because of its relevance to legal issues; specifically, the use of log files as evidence in a court of law. In 2004, Martha Stewart was brought to trial for conspiracy, perjury, securities fraud, and obstruction of the investigation into the sale of her stock in ImClone. During her trial, computer logs were used as evidence to show that incriminating phone messages were altered and then subsequently restored, which helped prove obstruction of the investigation and

ultimately led to her conviction. In another example, Bank of America was fined $10 million by the Securities and Exchange Commission in 2004 because it failed to turn over evidence, much of it in the form of logs, for an investigation into improper trading by its employees. Whether or not this failure to hand over evidence was deliberate, it showed that insufficient accountability management procedures can be a liability to an organization and can have serious legal implications.

6.2.5 Completeness

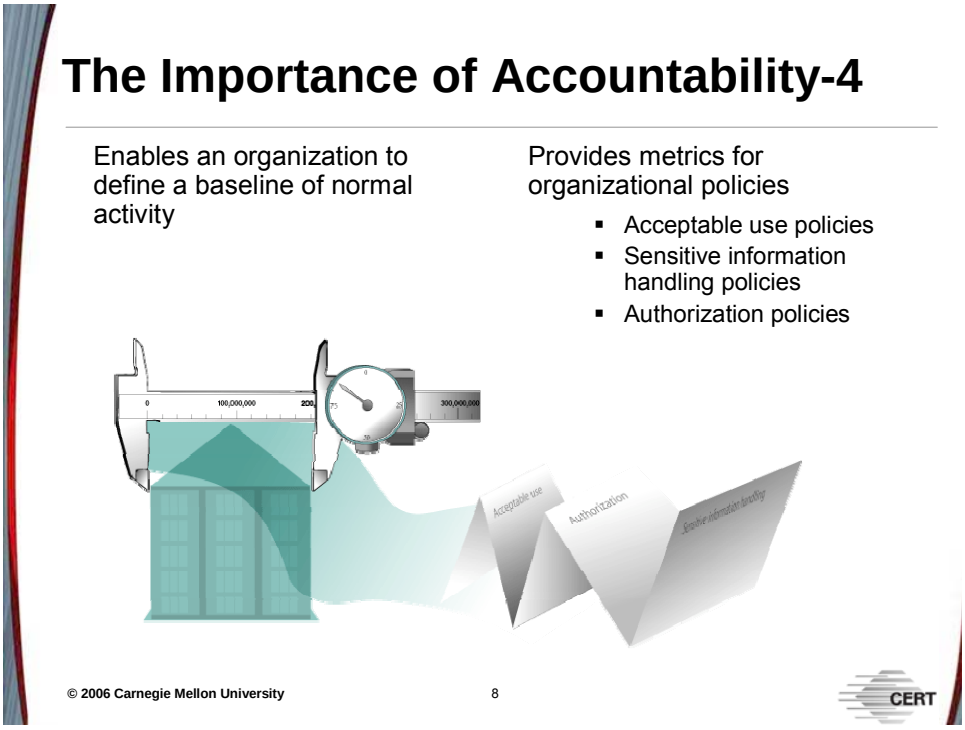
It is important to note that log file collection by itself is not adequate. An organization must be able to ensure that its log files are complete, accurate, and verifiable. Completeness enables the objective investigation of a computer security incident by capturing activities without any gaps in time. Additionally, logs within an organization should be maintained in aggregate form so that they represent the sum total of the entire organization's IT processes. For example, a network scan alone may indicate the presence of a malicious insider; however, correlating this event with other logs may reveal that the suspicious computer is infected by a virus that was propagated via email.

6.2.6 Accuracy

Logs are only useful in legal matters if their entries have not changed from the time they were originally created. This includes time, date, and message information. A log's reliability greatly hinges upon its accuracy.

6.2.7 Verifiability

Although completeness and accuracy are essential, it is just as important to be able to prove that these properties of completeness and accuracy truly apply to the log files. In other words, it is vital to ensure an attacker has not altered log files to cover his or her traces. Methods such as creating hashes of log files (accuracy), correlating events from independent sources (accuracy), and having well-documented log management processes (completeness) can help an organization ensure verifiability.



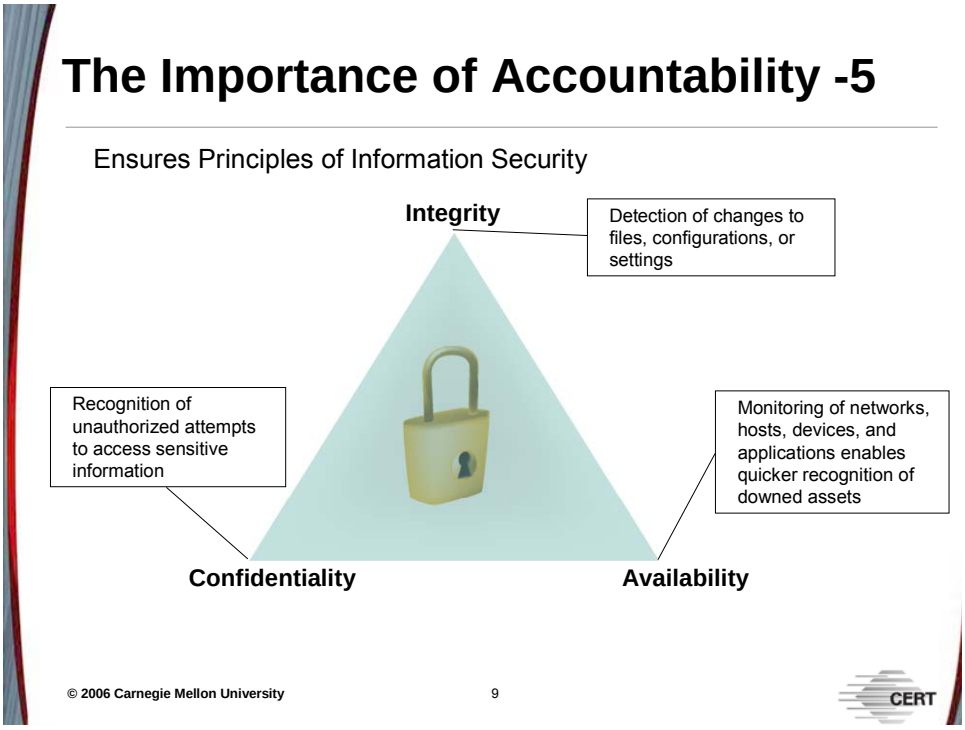
6.2.8 Baselines for Normal Activity

Accountability management enables organizations to create baselines for normal activity. This is a valuable function because it allows you to quantify abstract organizational activities. For example, an organization can use traffic and bandwidth monitoring to determine the average amount of traffic on its network at various times of day. Using this baseline, the organization can detect deviations in network traffic and attempt to determine whether such deviations indicate adverse activity. For an organization that primarily conducts business during the day, network traffic at odd hours, such as 2:00 a.m., may indicate that the network is being attacked. On the other hand, a dramatic decrease in network activity during business hours may be a sign that a component of the network is not working properly.

6.2.9 Policy Metrics

Technology is often used to support and automate policy enforcement and assessment. For example, building access can be enforced by employing Radio Frequency Identification (RFID) card-reader systems that have access-control policies automatically programmed into them. This level of automation ensures more consistent policy enforcement, as it reduces the possibility of human error. Additionally, utilizing technology in accountability management processes allows an organization to develop quantifiable metrics for policies, which by their nature are qualitative. For example, an organization may define in its acceptable use policy a set of applications that are prohibited from use on its network. In this instance, traffic monitoring can identify the presence of rogue applications and determine how well the acceptable use policy is being followed. Management can then use this information to

determine whether the current policy is effective or needs to be amended, or whether additional enforcement measures must be put in place. Policy metrics will ultimately lead to more efficient policies and implementations that better serve the mission of the organization.



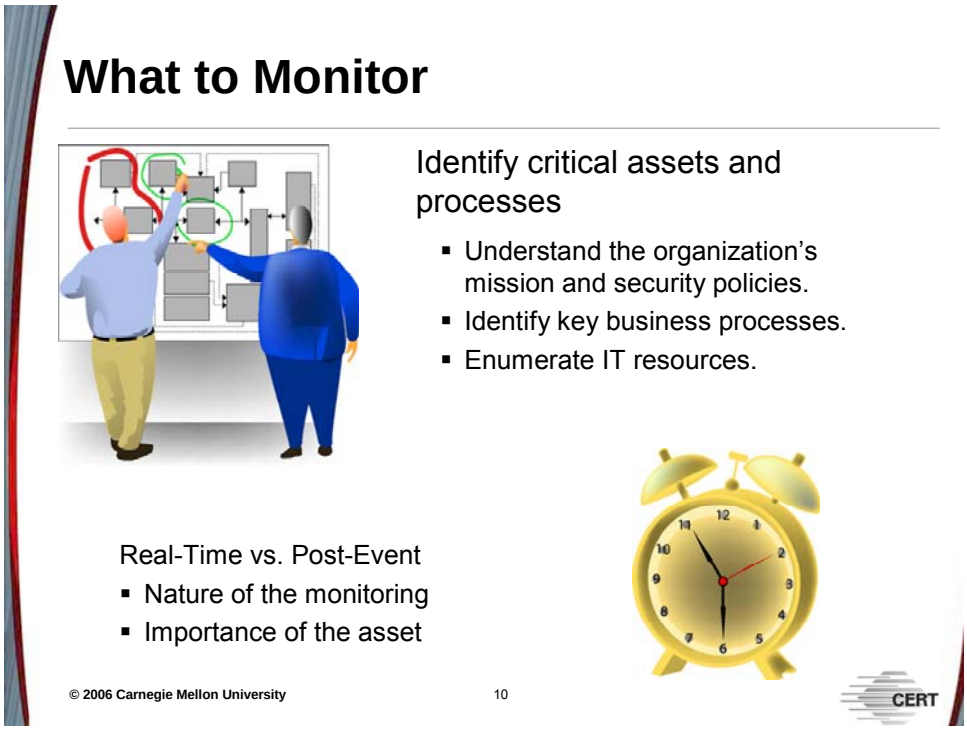
6.2.10 Ensures Principals of Information Security

A key aspect of information security is preservation of the confidentiality, integrity, and availability of the information being protected. Accountability management is important because it helps an organization ensure these properties of information security.

Specifically, confidentiality can be preserved because accountability mechanisms can recognize unauthorized attempts to access sensitive information. These mechanisms may include log file analysis and intrusion detection. Subsequently, an organization can react to any detected illicit behavior and minimize the potential damage. Knowledge gained through this incident will help prevent similar breaches from occurring in the future.

Integrity can be assured through the detection of changes to files, configurations, or settings. Host-based integrity monitoring is a good example of an accountability management implementation that can perform this type of task. Maintaining information integrity is crucial; should an incident occur, malicious or not, an organization will need to rely on the accuracy of its log files and monitoring data to determine the root cause of the incident.

Lastly, accountability management bolsters the availability of an organization's information, services, and IT assets by enabling the monitoring of networks, hosts, devices, and applications. This type of monitoring enables an organization to quickly recognize downed assets and take action to remediate the problem before it becomes a disaster. Availability is not only relevant to security, but also can be vital to the organization's mission, as is the case for many types of service providers (such as ISPs and Web-hosting companies).



6.3 Implementations

The goal of this section is not to teach you how to install, configure, and analyze data from various accountability management implementations. Rather, it is to enable managers and those in decision-making positions to understand at a high level the types of implementations that exist, how they relate to accountability management, and the purpose they serve for an organization. The benefits of this approach are twofold. First, you will be able to interface with technical staff members to understand implementations that are currently in place; make informed decisions for employing new accountability processes; and ensure that the measures already in place are fulfilling their intended purpose. Second, this level of understanding enables managers to effectively communicate with higher-level executives about accountability management's importance to the mission of the organization.

6.3.1 What to Monitor

Before an organization can apply accountability management practices, it must first determine which assets and processes should be monitored. This is a crucial step because an accountability management program is only as useful as the monitored assets are important. The best way for an organization to decide what to monitor is to identify critical assets and processes, which involves three main steps.

First, you must fully understand the organization's mission and become intimately familiar with its security policies. These policies will help you identify critical assets and processes. Identifying key business processes is the second step. A key business process can be

classified as any procedure that must be available and operational in order for an organization to carry out its mission and conduct business. Depending on the size of the organization, one individual alone may not be capable of identifying key business processes. Interviews may need to be conducted to elicit key processes from various groups within the organization. The third step is to enumerate all IT resources related to each key business process. An IT resource could be a piece of hardware or software, or a key employee and his or her knowledge. Again, this can be an extensive task and may require a collaborative effort between IT staff members and system administrators. Once all is done, the organization will have a list of components that tie into key business processes and therefore should be monitored.

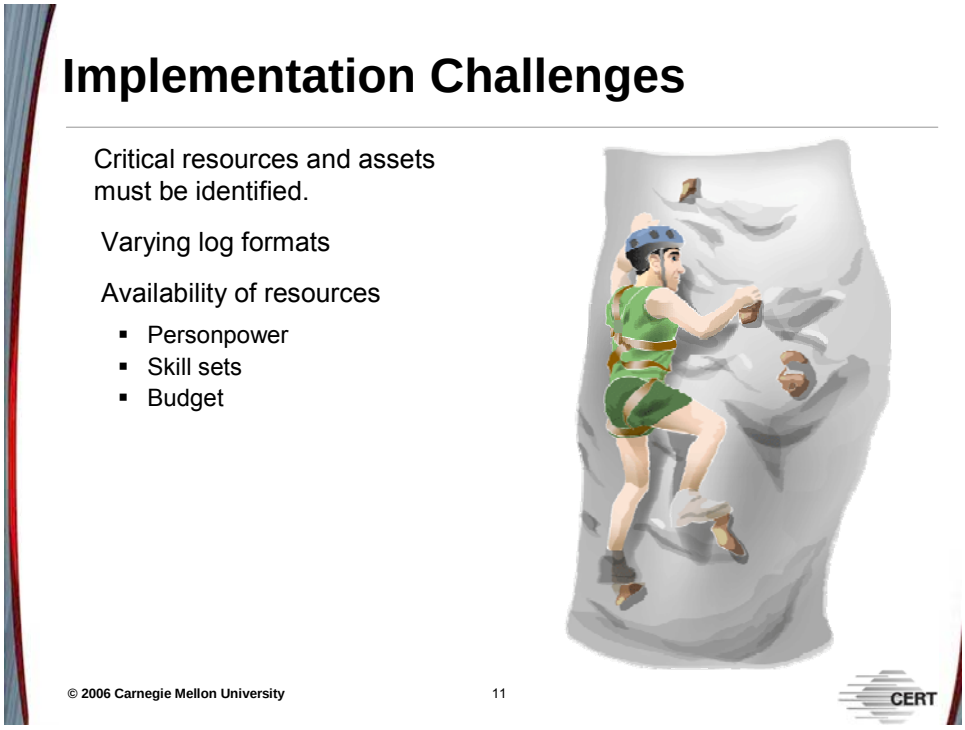
To summarize, the three steps for identifying critical organization IT assets and processes are

1. Understand the organization's mission and security policies.
2. Identify key business processes.
3. Enumerate IT resources related to key processes, including key people and their knowledge.

These steps are illustrated in the following example. Suppose a grocery chain would like to identify in-store components that should be audited. The first thing to do is understand the organization's mission, which in this case is to provide retail grocery services to the general public. For the sake of this example, no security policies will be considered. Next, key business processes need to be identified. A few key processes could be customer checkout, employee shift management, and inventory. The customer checkout process will be further examined to identify IT resources related to its function. Customer checkout may rely on point-of-sale terminals, a network switch, a point-of-sale server, and a service that feeds the POS terminals with pricing and customer information. If any of these devices or applications were to go down, customer checkout could not be performed. This would be in direct conflict with the store's mission. Therefore, these IT resources should be included in any accountability management processes that are implemented by the grocery store.

6.3.2 Real-Time vs. Post-Event

Once critical components have been identified, it is important to consider which monitoring method should be used to oversee them, such as real-time or post-event monitoring. Real-time monitoring is fairly self-explanatory and should be used for assets or processes that need to be constantly watched, such as the uptime of a Web server. On the other hand, post-event monitoring such as log file examination can be used for processes that require an audit trail but need not be closely watched. The advantage of post-event monitoring is its relative simplicity. In general, more important assets and processes should be monitored in real-time, while less important ones can be monitored post-event. The next sections of this module will go into more detail about the different types of accountability management implementations.



6.3.3 Implementation Challenges

6.3.3.1 Identifying Critical Resources and Assets

As was explained in the section "What to Monitor," critical resources must be identified and well documented to ensure that they will be monitored at the appropriate level by accountability management processes.

6.3.3.2 Varying Log Formats

One challenge with regard to log management is the lack of a single standard format for log files. Although there are widely used log formats such as syslog, log file formats ultimately are determined at the discretion of the developer. The problem of asynchronous formats becomes amplified when an organization employs devices from varying vendors and software components from differing platforms. For example, Windows systems use the Eventlog format, while Linux systems use syslog. If an organization deploys both Windows and Linux servers, it faces the problem of aggregating the files into a single logical format. In this case, packages such as Snare Agent for Windows are capable of transforming Windows log messages into syslog format. However, most organizations will have to deal with more than just two types of log formats. Moreover, log formats from various devices and applications may be far from conventional.

6.3.3.3 Security of Log Files

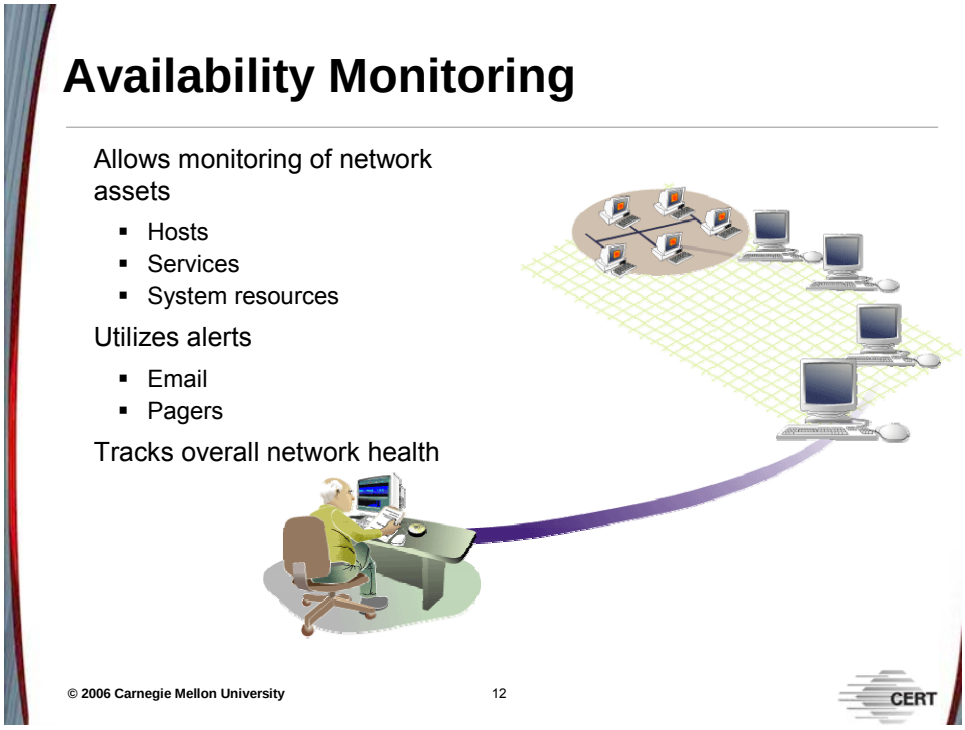
Another challenge is that malicious attackers often can overwrite or alter log files to remove all traces of their presence. Log files exist in the first place to provide an audit trail of system events that can be traced when something goes wrong. If the trustworthiness of log files is in question, their utility falls dramatically. Therefore, securing log files is an important step in the accountability management process. One way to do this may be to save log files to write-only or read-only media, thereby preventing attackers from altering them after the fact.

6.3.3.4 Volume of Log Data

Logging programs also are capable of generating vast quantities of output. The sheer amount of data can seem overwhelming. Finding efficient and effective ways to sort through log data, automate review of such data, or visualize the data can be a significant challenge.

6.3.3.5 Availability of Resources

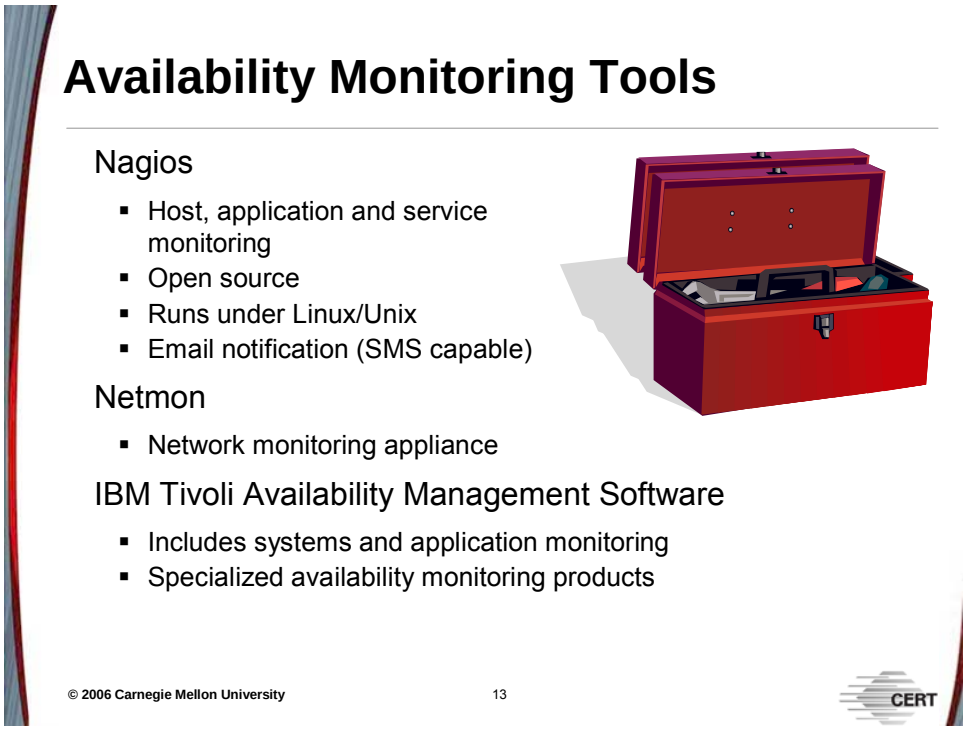
Yet another challenge that organizations face is the ability to procure sufficient resources. In an ideal world, an organization would employ a dedicated accountability management staff with a very high skill set and an unlimited budget for procurement. Realistically, however, most organizations are constrained in the areas of humanpower, skill sets, and budget. Limitations in humanpower and skill sets can be offset by a thorough understanding of accountability management. Imparting this understanding is the main objective of this module. A greater comprehension of accountability management will allow you to make informed decisions regarding implementation methods and gain the most from your available resources.



6.3.4 Availability Monitoring

In contrast to log files, which track the historical record of assets, the purpose of availability monitoring is to track critical IT assets in real-time. An asset needn't be a hardware device; it can be a software component or a system resource such as processor load or memory usage. For example, if an organization's Web site serves a critical function, availability monitoring can be used to ensure the Web servers are running. At the same time, it can also be used to monitor the Web service running on those machines, because physical machine uptime does not necessarily mean the service is functioning properly. Lastly, the processor load can be monitored as well to ensure the Web servers are not being overloaded.

Availability monitoring tools also provide alerting mechanisms beyond the real-time console display. Alerting is very important for two reasons. First, most organizations cannot afford to dedicate resources to constant supervision of an availability monitoring console. Second, unless extremely critical assets are being monitored, it is inefficient for employees to dedicate all of their time to this task. Alerts can be sent via email, voice mail, pager, or cell phone, depending on the urgency of each alert. This approach allows staff members to be more mobile and flexible while at the same time keeping track of overall network health.



6.3.5 Availability Monitoring Tools

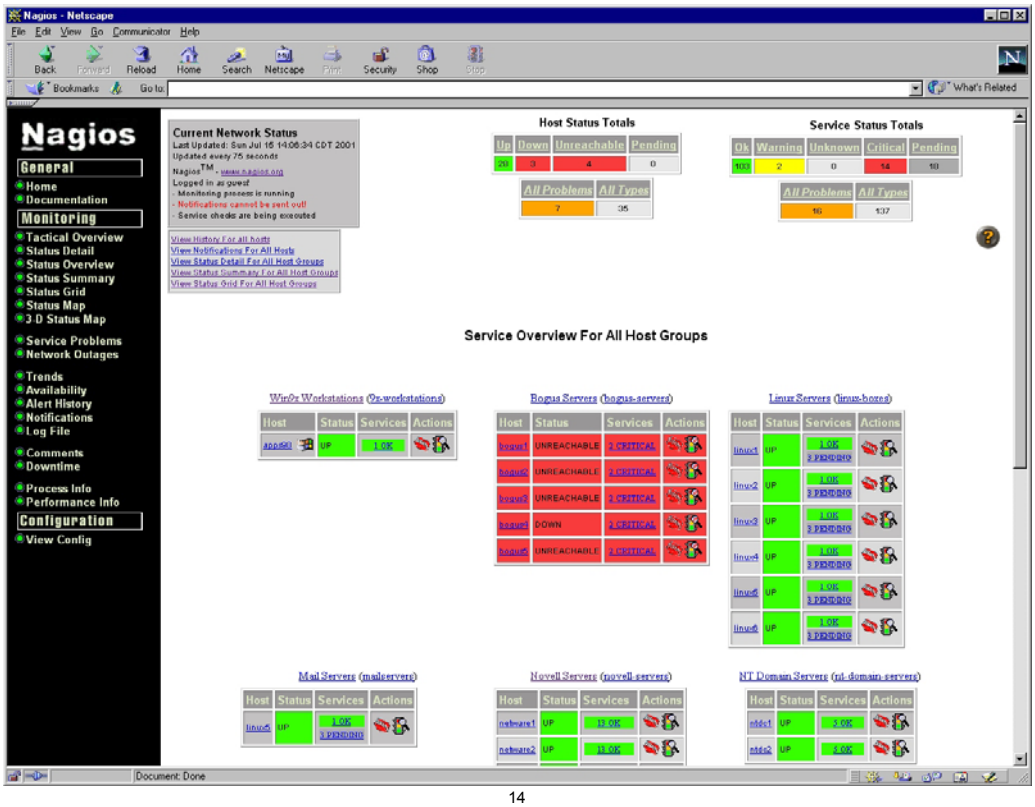
6.3.5.1 Nagios

Nagios, formerly Netsaint, is one of the most well-known availability monitoring tools. One advantage of Nagios is that it is an open-source tool released under the terms of the GNU General Public License, which is ideal for organizations with limited resources. Nagios is designed to run under the Linux operating system and uses a variety of alerting mechanisms such as email, Short Message Service (SMS), and instant messaging. The following excerpt is a descriptive overview of Nagios.[11]

> Nagios® is a host and service monitor designed to inform you of network problems before your clients, end-users, or managers do. It has been designed to run under the Linux operating system, but works fine under most *NIX variants as well. The monitoring daemon runs intermittent checks on hosts and services you specify using external "plug-ins" which return status information to Nagios. When problems are encountered, the daemon can send notifications out to administrative contacts in a variety of different ways (email, instant message, SMS, etc.). Current status information, historical logs, and reports can all be accessed via a Web browser [Nagios 06].

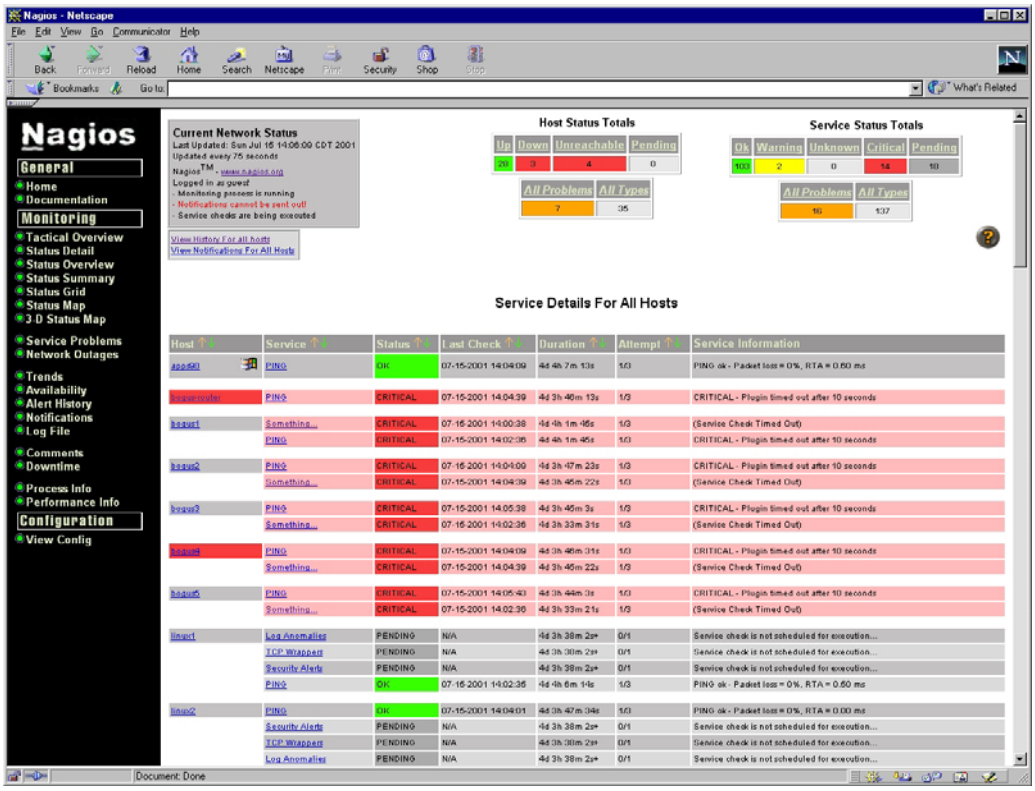
Screenshots from the Nagios console are on the following pages.

[11] See http://www.nagios.org/about/.



14

Figure 2: Status Overview (screenshot from Nagios console)



15

Figure 3: Status Detail (screenshot from Nagios console)

6.3.5.2 Netmon

Netmon is a network monitoring appliance that provides similar functionality to Nagios. Unlike Nagios, however, Netmon is a commercial product. Additionally, Netmon offers the benefit of a dedicated, hardware, appliance running an availability monitoring service. As part of this service, Netmon

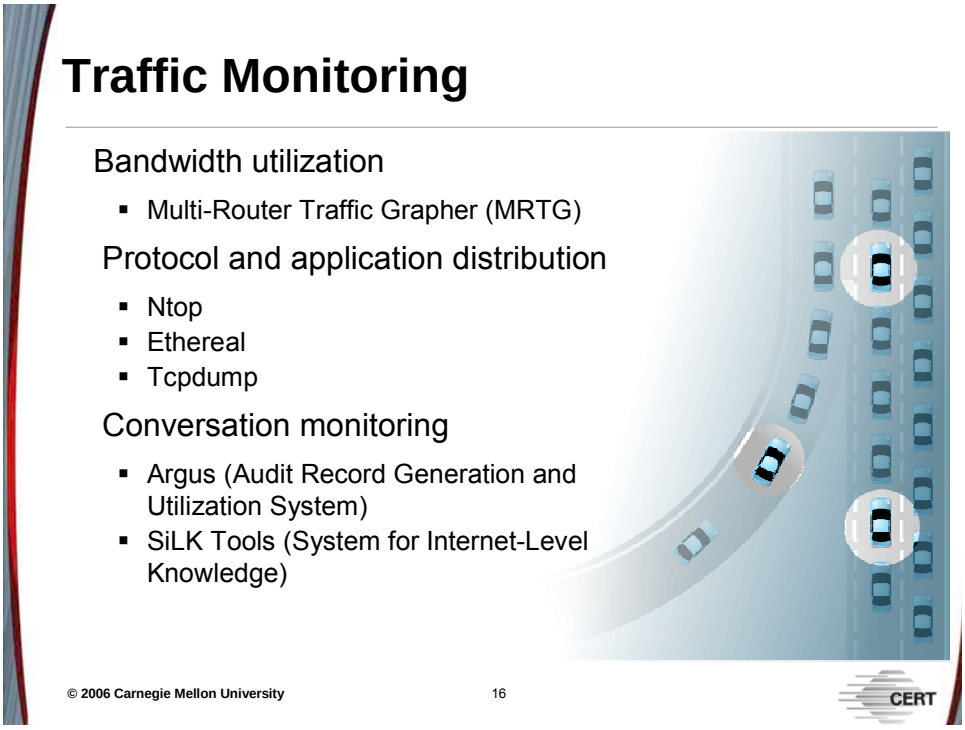
- monitors Internet bandwidth
- sends email or pager alerts for critical events
- keeps track of disk usage
- audits host network security with its own port scanning tools
- locates spyware, adware, worms, and other types of malicious software

It is important to note that some features of Netmon such as virus, spyware, and adware detection are part of intrusion detection, which will be covered later in this module.

6.3.5.3 IBM Tivoli Software

The IBM Tivoli infrastructure management framework is a set of software packages that provide a variety of management services. Included in this framework are availability monitoring capabilities. Like Netmon, IBM Tivoli Monitoring is a commercial solution. It

provides many of the same services as Nagios and Netmon, including host, service, application, and resource monitoring. Tivoli also offers a number of specialized availability monitoring products such as monitoring for .NET and WebSphere Application Server. Further information about IBM Tivoli availability management can be found at http://www-306.ibm.com/software/tivoli/solutions/availability/.



6.3.6 Traffic Monitoring

Traffic monitoring consists of analyzing and observing the actual packets traversing an organization's network. This type of monitoring enables an organization to closely track the volume of traffic as well as the types of traffic on its network.

6.3.6.1 Bandwidth Utilization

The ability to monitor a network's traffic load (bandwidth utilization) is important for two reasons. First, an organization will be able to determine how much of the network's bandwidth is being utilized. Anomalies in the traffic load should be of particular interest. For example, a heavier-than-usual network load could be an indicator of unauthorized applications on the network, such as peer-to-peer file sharing. Second, traffic monitoring enables an organization to define normal behavior patterns for traffic. For instance, traffic monitoring may show that network traffic peaks between 10:00 a.m. and 3:00 p.m. Subsequently, any significant deviations from this behavior, such as a traffic spike at 2:00 a.m., should be a cause for concern, as there should not be much legitimate traffic at that time.

The Multiple Router Traffic Grapher (MRTG) is an example of a bandwidth utilization monitoring tool [Oetiker 06]. The MRTG screenshot in Figure 3 shows bandwidth utilization over the course of one day and one week, while the screenshot in Figure 4 shows a graph of

traffic over the course of one month and one year.[12] The green area reflects incoming traffic, while the blue line indicates outgoing traffic (visible in the online color version of this document).

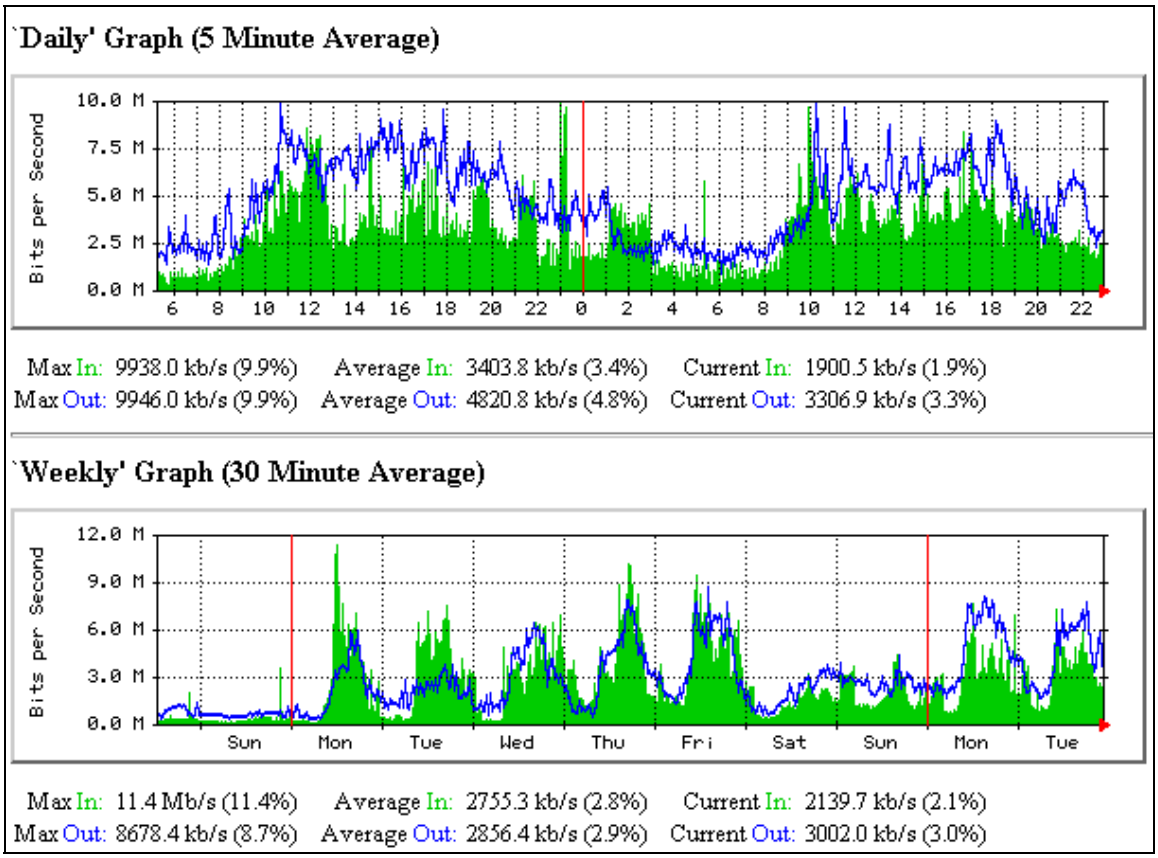


Figure 4: Daily and Weekly Graph (MTRG screenshot)

[12] See http://oss.oetiker.ch/mrtg/.

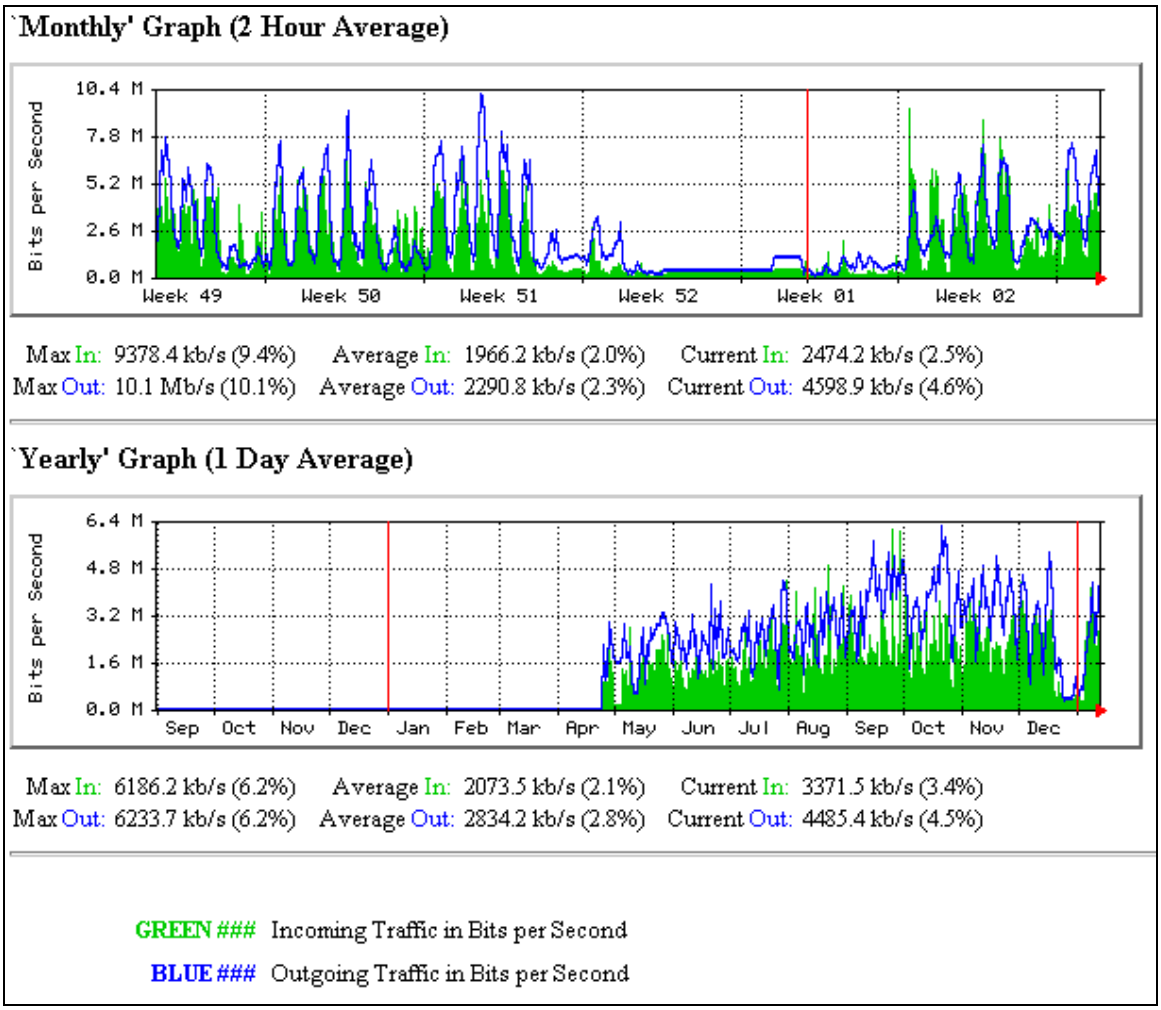


Figure 5: Monthly and Yearly Graph (MTRG screenshot)

6.3.6.2 Protocol and Application Distribution

Traffic monitoring can also involve more granular inspection of packets on the network. This is important because although traffic load monitoring will indicate spikes in traffic, it will not provide further detail. Therefore, protocol and application inspection become necessary to further break down traffic distribution. This approach allows an organization to determine the different *types* of traffic on a network, which helps ensure that network policies are being followed and provides metrics for evaluating the policies' overall effectiveness. For example, if a large number of unauthorized applications are running on the network, this may indicate that the current network policies are not clear enough and need to be reworked.

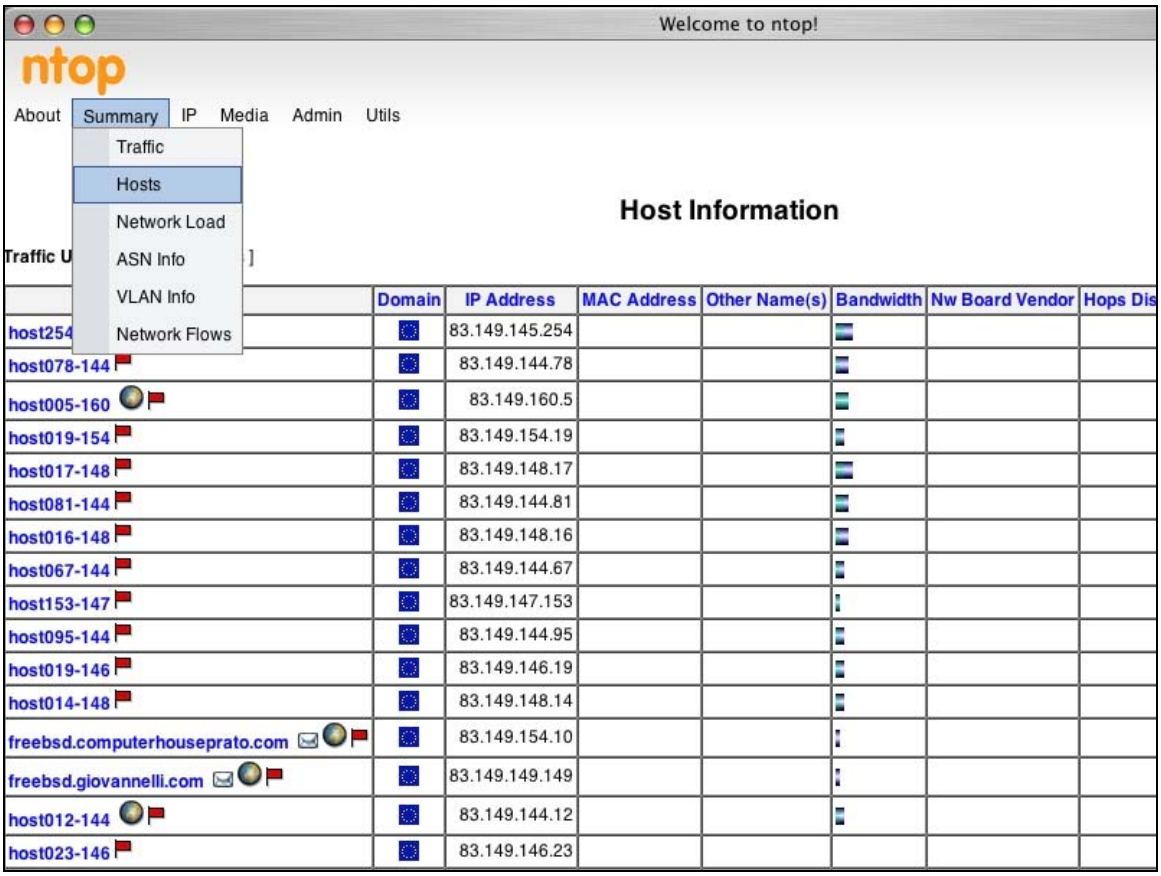
6.3.6.3 ntop

One such monitoring tool is ntop, a network traffic probe that displays network usage, protocol distributions, and traffic statistics [ntop 06]. Ntop is able to[13]

- sort network traffic according to many protocols
- show network traffic sorted according to various criteria
- display traffic statistics
- store on disk persistent traffic statistics in RRD format
- identify the identity (e.g., email address) of computer users
- passively (i.e., without sending probe packets) identify the host OS
- show IP traffic distribution among the various protocols
- analyze IP traffic and sort it according to the source/destination
- display IP Traffic Subnet matrix (who's talking to whom?)
- report IP protocol usage sorted by protocol type
- act as a NetFlow/sFlow collector for flows generated by routers (e.g., Cisco and Juniper) or switches (e.g., Foundry Networks)
- produce RMON-like network traffic statistics

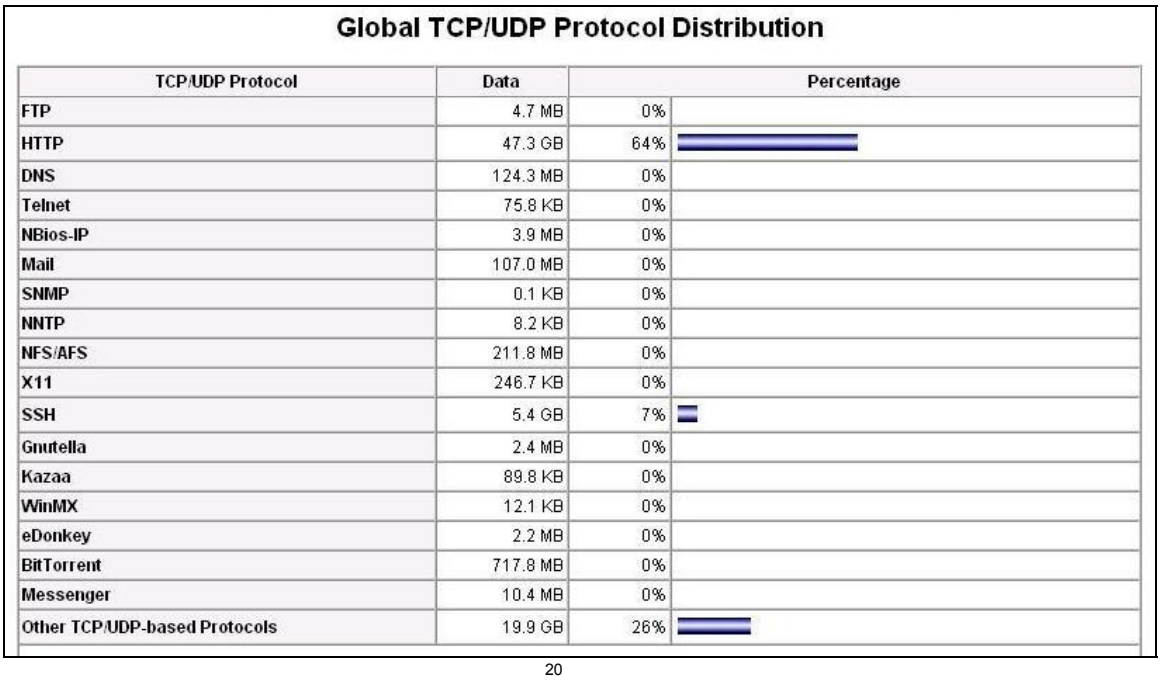
Ntop is distributed under the GNU General Public License and runs on UNIX (including Linux) and Windows platforms. Figure 5 shows the host information screen in ntop, which displays all the hosts on the network, their IP addresses, and bandwidth used. More granular information is available through selection of a specific host. Figure 6 shows the protocol distribution of traffic on the network, which quickly identifies any unauthorized traffic.

[13] See http://www.ntop.org/overview.html.



19

Figure 6: ntop Host Information



Global TCP/UDP Protocol Distribution

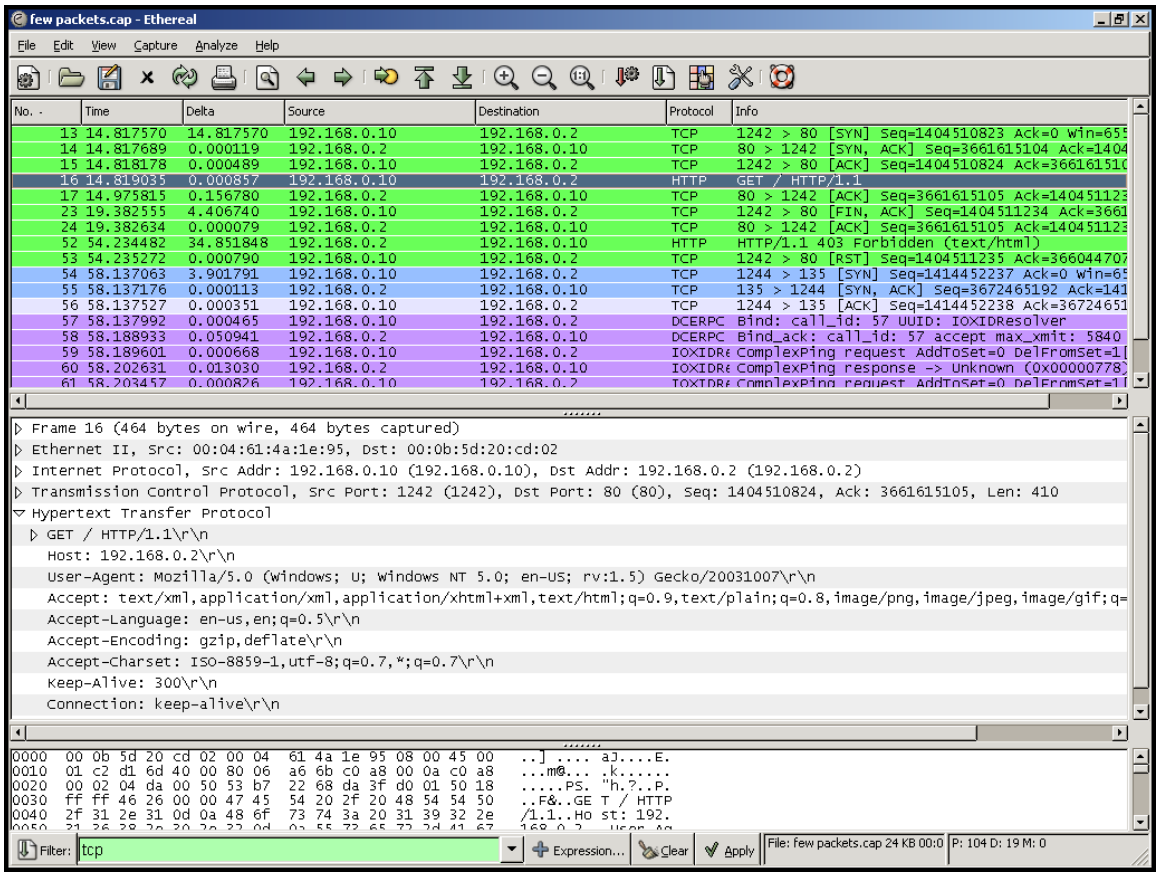
TCP/UDP Protocol	Data	Percentage
FTP	4.7 MB	0%
HTTP	47.3 GB	64%
DNS	124.3 MB	0%
Telnet	75.8 KB	0%
NBios-IP	3.9 MB	0%
Mail	107.0 MB	0%
SNMP	0.1 KB	0%
NNTP	8.2 KB	0%
NFS/AFS	211.8 MB	0%
X11	246.7 KB	0%
SSH	5.4 GB	7%
Gnutella	2.4 MB	0%
Kazaa	89.8 KB	0%
WinMX	12.1 KB	0%
eDonkey	2.2 MB	0%
BitTorrent	717.8 MB	0%
Messenger	10.4 MB	0%
Other TCP/UDP-based Protocols	19.9 GB	26%

20

Figure 7: ntop Protocol Distribution

6.3.6.4 Ethereal/Wireshark

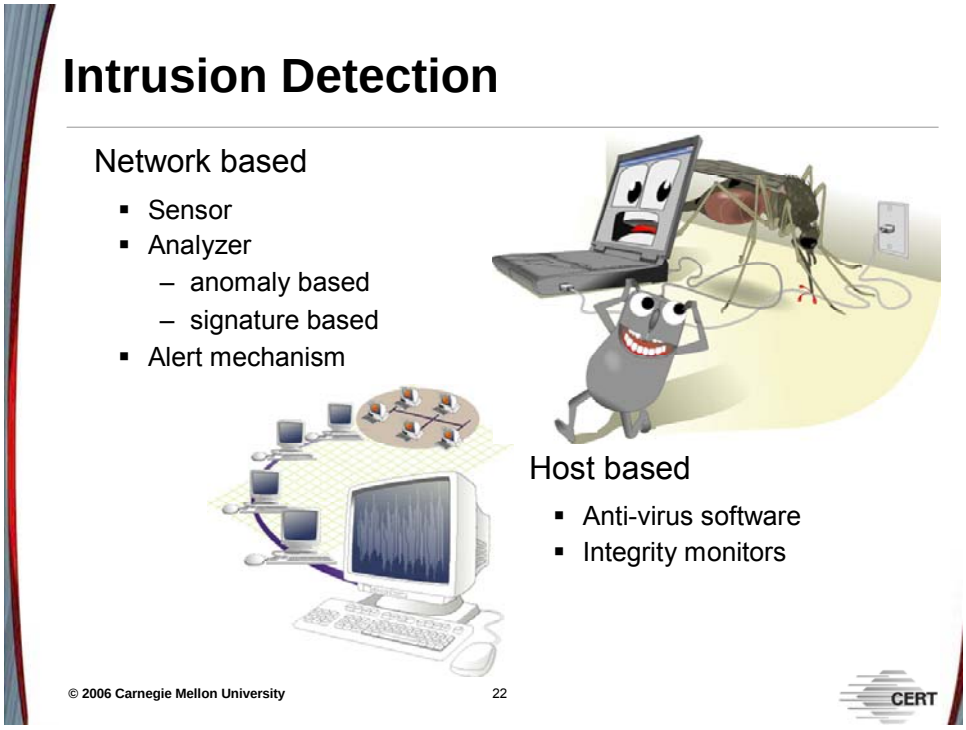
Ethereal, a network protocol analyzer, provides even more granular information than ntop but does not summarize the data as well. Unlike ntop, Ethereal (now called Wireshark) captures the entire packet when it monitors traffic. This means an organization can capture the actual traffic on its network. Figure 7 shows how a traffic capture is displayed in Ethereal [Wireshark 06]. The top window displays the packets that have been captured, and the middle window allows the user to look at detailed packet information.



21

Figure 8: Ethereal Main Window

Each organization's managers will need to determine for themselves which tools should be used for which purposes. Once this legwork is done, a best practice might be to specify the chosen course of action in organizational procedures at the tactical level. This will save time and effort in high-pressure situations, since employees will be able to refer to organizational procedures for guidance.



6.3.7 Intrusion Detection

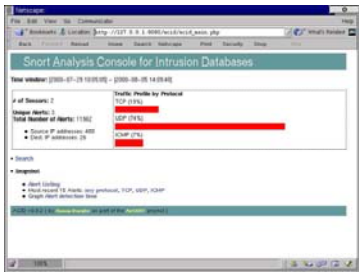
An intrusion detection system (IDS) is a component of accountability management that monitors a system or network for malicious activity. Intrusion detection systems can be broken into two subsets: network-based and host-based. A network-based IDS looks for suspicious traffic on a network. A host-based IDS monitors the host on which it is installed and examines inbound and outbound traffic, file integrity, and/or suspicious processes. Intrusion detection is an important component of accountability management because it serves as an auditing mechanism for attacks on an organization's IT infrastructure. An IDS can be likened to a burglar alarm.

An IDS comprises three main components: a sensor, an analyzer, and an alerting mechanism. The sensor collects data from the network or host, depending on the type of IDS. Data captured by the sensor is subsequently examined by the analyzer, which performs pattern matching to determine whether or not an intrusion has occurred on the network. Analyzers can be categorized as either signature-based or anomaly-based. A signature-based analyzer contains a set of attack patterns that are compared against the data collected by the sensor. If the sensor data matches one of the attack patterns, the analyzer deems it to be an intrusion. On the other hand, an anomaly-based analyzer defines a pattern of normal behavior that is compared against the incoming sensor data. The analyzer concludes that deviations from the defined normal behavior are intrusions. Finally, the alerting mechanism is used to output the results of the analyzer, usually only in the event of a detected intrusion.

Signature- and anomaly-based intrusion detection systems each have advantages and disadvantages. Signature-based intrusion detection systems are generally quicker, less

complex, and easier to implement. However, signatures (attack patterns) must be constantly kept up to date, and a signature-based IDS is not capable of detecting attacks for which it has no signature (e.g., zero-day attacks). Conversely, an anomaly-based IDS can detect zero-day attacks and other unknown exploits. However, it is typically more complex and resource intensive. The signature-versus-anomaly debate is extensive and beyond the scope of this module.

Network Intrusion Detection Tools

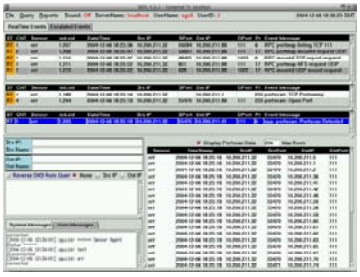


Sguil

- Real-time analysis of Snort events
- Signature based

Snort

- Widely popular IDS systems
- Open source
- Signature based

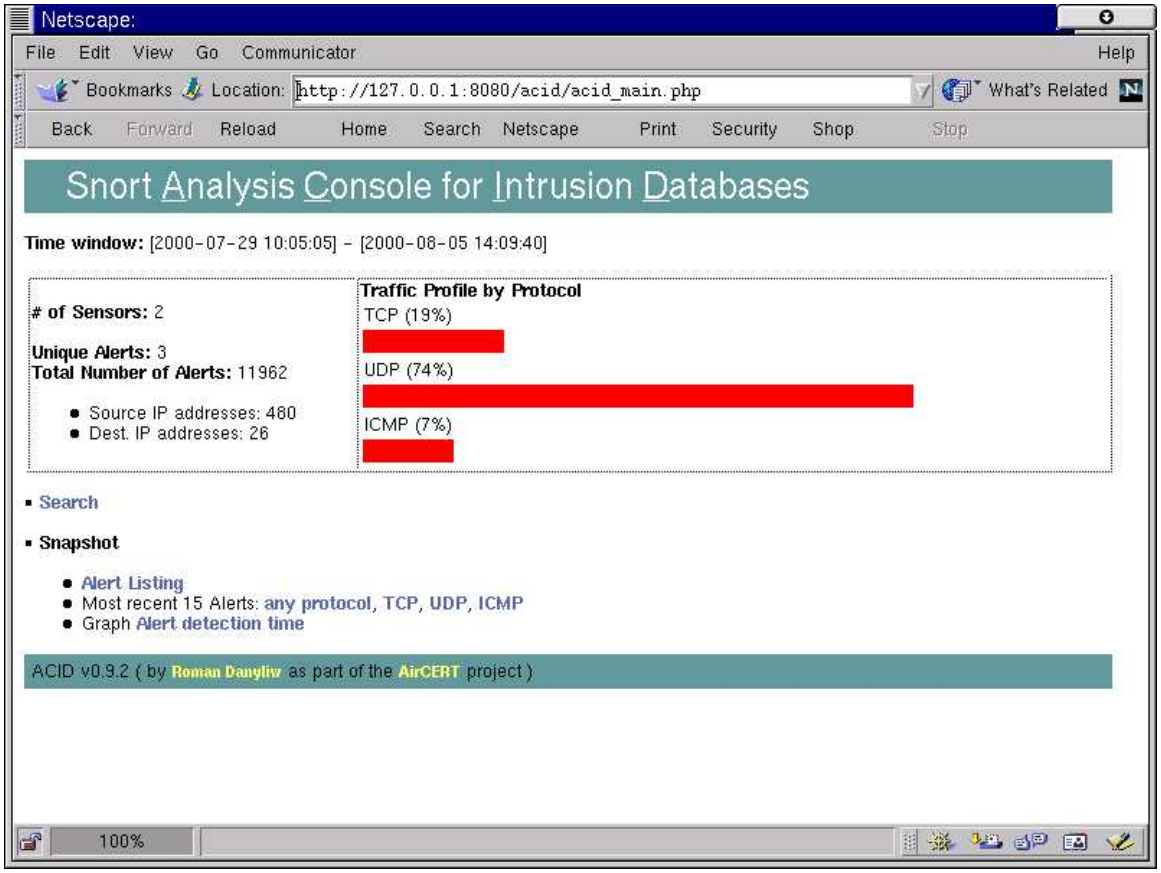


CERT

6.3.7.1 Network Intrusion Detection Tools

As stated earlier, the purpose of a network-based IDS is to passively capture network traffic and examine it for possible attacks. Snort is one of the most popular, widely known network intrusion detection systems [Sourcefire 06]. It is an open-source, signature-based IDS. Snort was originally designed for *nix systems but can now be run on Windows systems as well. Its output is purely textual, but graphical interfaces have been developed such as the Analysis Console for Intrusion Databases (ACID). Figure 8 shows the ACID main screen, which displays, among other things, the number of alerts Snort has detected. More detailed information is available on other ACID screens. Note, ACID is now maintained by the open source community and has changed names to Basic Analysis and Security Engine (BASE). BASE can be downloaded from http://base.secureideas.net/.

Barnyard is a plug-in to Snort IDS that enables Snort alerts to be stored in SQL databases.



23-left

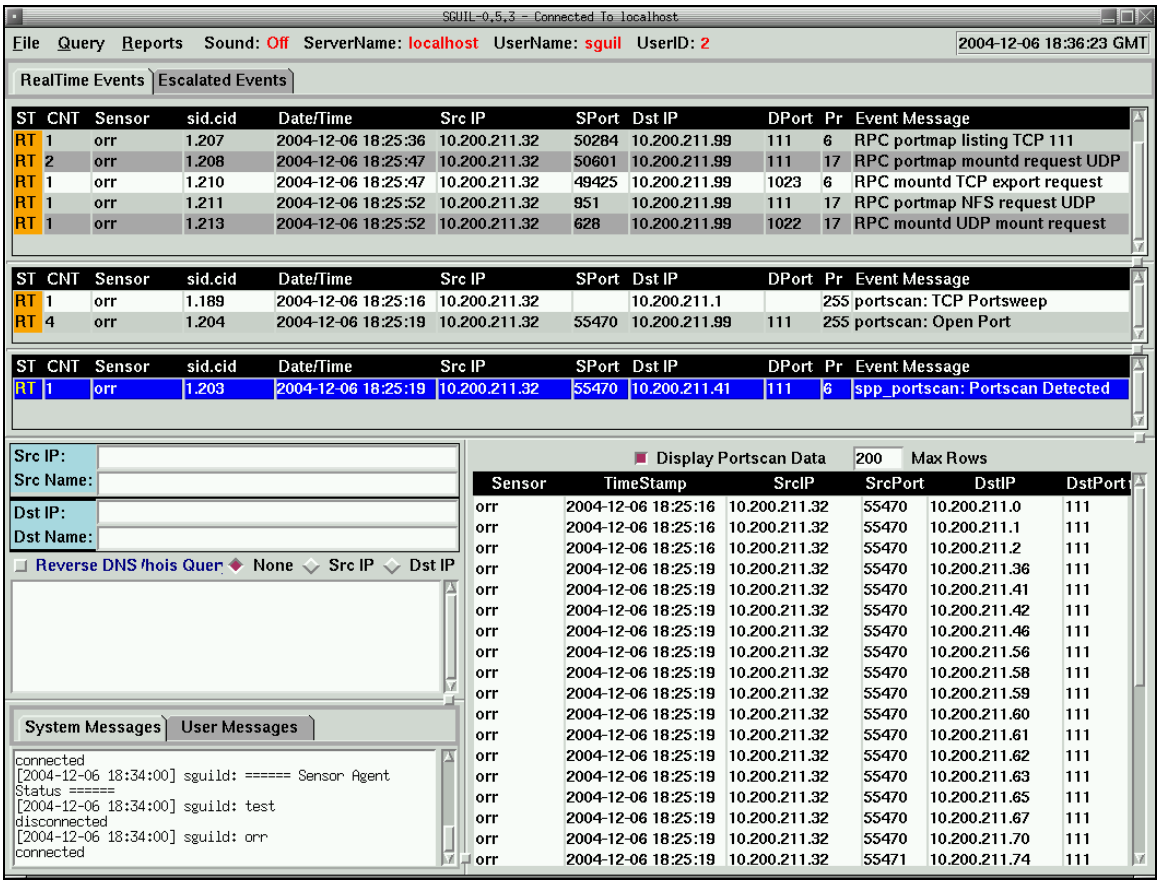
Figure 9: ACID Main Screen [BASE 04]

Another graphical interface that has been developed for Snort is Sguil [Visscher 06]. The following description is from the Sguil Web site:[14]

> **Sguil** (pronounced sgweel) was built by network security analysts for network security analysts. Sguil's main component is an intuitive GUI that provides real-time events from Snort/Barnyard. It also includes other components which facilitate the practice of Network Security Monitoring and event-driven analysis of IDS alerts. The sguil client is written in tcl/tk and can be run on any operating system that supports tcl/tk (including Linux, *BSD, Solaris, MacOS, and Win32).

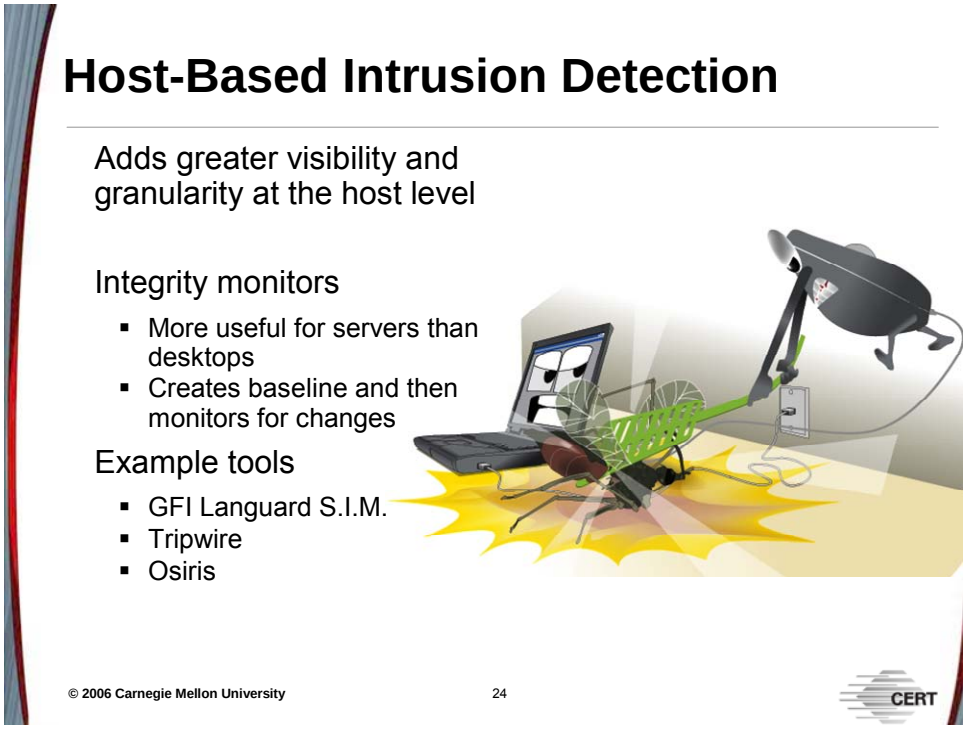
Figure 9 is a screenshot from Sguil that displays the results of a detected port scan.

[14] See http://sguil.sourceforge.net/.



23-right

Figure 10: Sguil Port Scan



6.3.7.2 Host-Based Intrusion Detection Systems

This section will focus primarily on host-based integrity monitoring tools, although it should be noted that antivirus and anti-spyware applications can also be considered host-based IDSs.

The following description of host-based intrusion detection systems comes from the SANS Institute, an organization focused on information security training [SANS 06].[15]

> Host-based ID involves loading a piece or pieces of software on the system to be monitored. The loaded software uses log files and/or the system's auditing agents as sources of data. In contrast, a network-based ID system monitors the traffic on its network segment as a data source. Both network-based and host-based ID sensors have pros and cons, and in the end, you'll probably want to use a combination of each. The person responsible for monitoring the IDS needs to be an alert, competent System Administrator, who is familiar with the host machine, network connections, users and their habits, and all software installed on the machine. This doesn't mean that he or she must be an expert on the software itself, but rather needs a feel for how the machine is supposed to be running and what programs are legitimate. Many break-ins have been contained by attentive Sys Admins who have noticed something "different" about their machines or who have noticed a user logged on at a time atypical for that user.
>
> Host-based ID involves not only looking at the communications traffic in and out of a single computer, but also checking the integrity of your system files and watching for suspicious

[15] See http://www.sans.org/resources/idfaq/host_based.php.

processes. To get complete coverage at your site with host-based ID, you need to load the ID software on every computer. There are two primary classes of host-based intrusion detection software: host wrappers/personal firewalls and agent-based software. Either approach is much more effective in detecting trusted-insider attacks (so-called anomalous activity) than is network-based ID, and both are relatively effective for detecting attacks from the outside.

Host-based intrusion detection adds greater visibility at the host level because it can reveal the *effects* of a network attack. For example, a network IDS may detect an attempted intrusion into a Web server. However, it cannot determine whether or not the attack was successful or the extent of the intrusion. Only a host-based IDS can do that.

6.3.7.3 Host-Based Integrity Monitoring

The purpose of a host-based integrity monitor is to ensure that no unauthorized changes are made to the system. To do this, the integrity monitor usually generates a baseline against which changes will be compared. Hashing can be utilized for this purpose and will be discussed later in this module. However, some integrity monitors use statistical and anomaly algorithms. Host-based integrity monitoring is typically better suited for servers than for desktops, since servers generally contain collections of critical files, settings, and processes that do not change frequently. The importance of ensuring integrity will be illustrated in the next section, which discusses best practices.

Tripwire, GFI LANguard System Integrity Monitor (S.I.M.), and Osiris are examples of integrity monitoring utilities . The following descriptions are not endorsements of these products; their purpose is to help individuals gain a better understanding of the types of integrity monitors that are available.

6.3.7.4 Tripwire

Tripwire is an open-source integrity monitor that has been released under the GNU General Public License. It is designed to run on Linux platforms. The "Red Hat Linux Reference Guide"[16] contains a good description of Tripwire:

Tripwire data integrity assurance software monitors the reliability of critical system files and directories by identifying changes made to them. It does this through an automated verification regimen run at regular intervals. If Tripwire detects that a monitored file has been changed, it notifies the system administrator via email. Because Tripwire can positively identify files that have been added, modified, or deleted, it can speed recovery from a break-in by keeping the number of files which must be restored to a minimum. These abilities make Tripwire an excellent tool for system administrators seeking both intrusion detection and damage assessment for their servers [Red Hat 03].

[16] See https://www.redhat.com/docs/manuals/linux/RHL-9-Manual/ref-guide/ch-tripwire.html.

6.3.7.5 GFI LANguard S.I.M.

LANguard S.I.M. is a freeware utility that runs on Windows platforms [GFi 06]. The following is a description of LANguard S.I.M. from its Web site:

> GFI LANguard System Integrity Monitor (S.I.M.) is a utility that provides intrusion detection by checking whether files have been changed, added or deleted on a Windows 2000/XP system. If this happens, it alerts the administrator by email. Because hackers need to change certain system files to gain access, this FREEWARE utility provides a great means to identify any servers that are open to attack.[17]
>
> GFI LANguard S.I.M. scans your system for important system files, computes an MD 5 checksum for every important system and files this in a database. At scheduled intervals, GFI LANguard S.I.M. scans the list of monitored files, computes another MD 5 checksum and tests the current value against the stored value to determine if the file has been modified. If it detects a change, it notifies the system administrator via email, and logs the occurrence in the security event log.[18]

6.3.7.6 Osiris

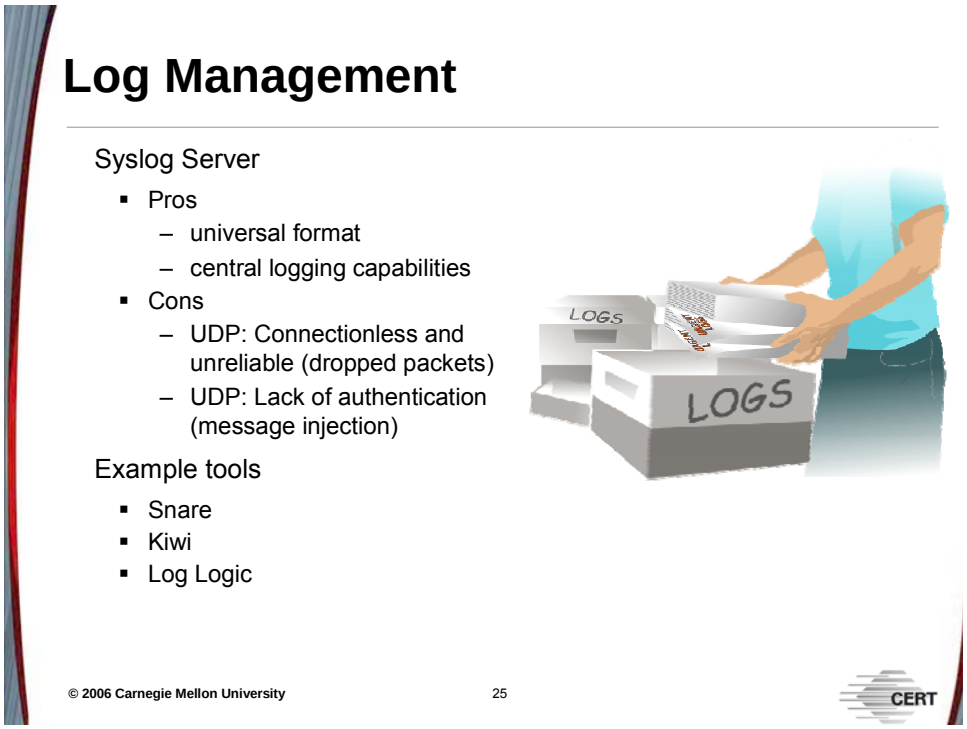
Osiris is an integrity-monitoring system managed by the Shmoo Group, which commits its free time to information security research and development. In addition to monitoring file changes, Osiris can check for changes in user lists, group lists, and kernel modules and extensions. The following description comes from the Osiris User Handbook:[19]

> Osiris is a host integrity monitoring system that can be used to monitor changes to a network of hosts over time and report those changes back to the administrator(s). Currently, this includes monitoring any changes to the filesystems. Osiris takes periodic snapshots of the filesystem and stores them in a database. These databases, as well as the configurations and logs, are all stored on a central management host. When changes are detected, Osiris will log these events to the system log and optionally send email to an administrator.
>
> In addition to files, Osiris has the ability to monitor of other system information including user lists, group lists, and kernel modules or extensions.
>
> Some integrity monitoring systems are signature-based, that is, they look for specific file attributes as a means of detecting malicious activity. This type of approach to host integrity can be very cumbersome to manage and can lead to unauthorized change going undetected. Osiris is intentionally not like this. Osiris will detect and report changes to a filesystem and let the administrator determine what (if any) action needs to take place. There are no complicated assumptions or dumbing-down of information. If a change occurs, it can be detected and reported.

[17] See http://www.gfi.com/lansim/.
[18] See http://www.gfi.hk/lansim/lansimfeatures.htm.
[19] See http://osiris.shmoo.com/handbook.html#part2_chap6.

All Osiris components compile and run on both Windows and common UNIX systems, including BSD, Linux, Mac OS X and Darwin, AIX, IRIX, and Windows NT/2K/XP. This allows for the flexibility to manage all types of platforms from either a UNIX or Windows environment.

Each organization will need to determine for itself the level and scope of intrusion detection it wishes to deploy. Risk assessment, as discussed in the Risk Management module of this course, can help in this determination. What assets must be protected? Where are these assets located? What level of protection is needed? Answers to these questions will guide selection not only of an IDS system, but of all components of a comprehensive security strategy.



6.3.8 Log Management

This section will review different methods that can be used to aggregate log files. These methods will not be covered in great detail, because the goal is not to understand how to perform the actual technical implementation. That task should be commissioned to the IT staff. Rather, the objective of this section is to provide managers with sufficient knowledge to compare multiple logging implementations and determine which ones best suit their organizations.

6.3.8.1 Syslog Servers

One of the most popular log management implementations is a syslog server. Even though syslog has been widely used for event monitoring, it was not standardized until RFC 3164 was published in 2001. The purpose of a syslog server is to collect and aggregate syslog messages from separate devices, systems, and applications in one central location. There are several advantages to employing a syslog server. First, it serves as a central log repository; maintaining such a repository is an accountability management best practice. Second, since all messages adhere to the syslog format, there are no complications stemming from disparate log formats. It is important to note that there are also disadvantages of syslog server implementations. Syslog messages are sent to the server via UDP (User Datagram Protocol), which is a connectionless transport protocol. This means devices do not establish a connection with the syslog server when they send a message. Rather, the message is blindly sent with no guarantee that it will be received. If for some reason the syslog server is down, devices will continue to send log messages to it even though they are not being received.

Furthermore, since UDP lacks authentication, anyone can send messages to the syslog server. This can be problematic because attackers can inject their own messages into the server in an effort to compromise integrity, or they can flood the syslog server with messages to cover their tracks.

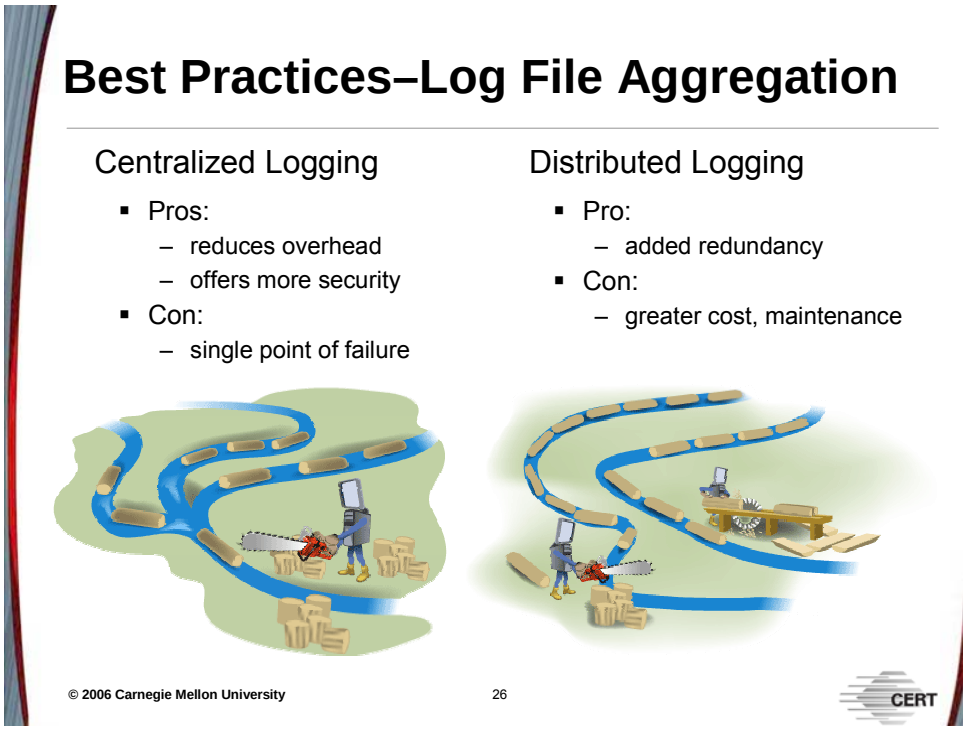
6.3.8.2 Other Tools

It is important to become familiar with log management and log aggregation tools. It is dangerous to become locked in to a single solution because that solution may not always be best for the organization. The ability to compare tools and understand their capabilities enables managers to make informed decisions. Therefore, a few log management tools will be reviewed in this section to serve as a launch pad. The first tool, Snare Agent for Windows, interfaces with the Windows Eventlog subsystem and can transform Windows logs into syslog format and send them to a syslog server. The advantage of such a tool is that it helps alleviate the problem of dealing with varying log formats. Additionally, Snare Agent for Windows is a free tool that has been released under the terms of the GNU General Public License. It along with other tools can be found at http://www.intersectallinace.com.

Another log management tool is the Kiwi Syslog Daemon, which is a freeware syslog daemon for Windows. It receives, logs, displays, and forwards syslog messages from routers, switches, UNIX hosts, and other syslog enabled devices. Many configuration options are available, including rotation and truncation of log files, a customizable real-time interface that can filter high-level events while logging the remaining events to a file, and the ability to operate with either TCP or UDP syslog traffic. Kiwi also offers a suite of related products that are free.[20]

You should keep in mind that log management tools do not only come in the form of software packages. The company LogLogic offers a set of log management hardware devices that are capable of aggregating and analyzing log data. One of the advantages of employing such a device is that it is a completely dedicated log management resource, which makes it more reliable and secure. On the other hand, hardware devices are commercial solutions and may not be affordable or cost-effective for smaller organizations. However, for an organization that highly values its log data, such as a financial institution, a log management device may be well worth the cost.

[20] Excerpt taken from *Advanced Information Security for Technical Staff Handbook*, distributed through the SEI course “Advanced Information Security for Technical Staff.”



6.4 Identification of Best Practices

It is important not only to understand the relevance of accountability management, but also to be able to design a robust accountability management program. To do this, you will need to become familiar with accountability management best practices. Taking best practices into consideration is advisable because they tend to be established, successful approaches and implementations. Additionally, use of best practices can save an organization considerable time and money, allowing it to avoid attempts to "reinvent the wheel." After all, it is not necessary to create procedures from scratch if an established set of practices already exists. Also, an organization stands to expose itself to less risk by utilizing tried-and-true practices as opposed to untested ones. However, it is important to keep in mind that best practices are just one component of creating a successful accountability management program.

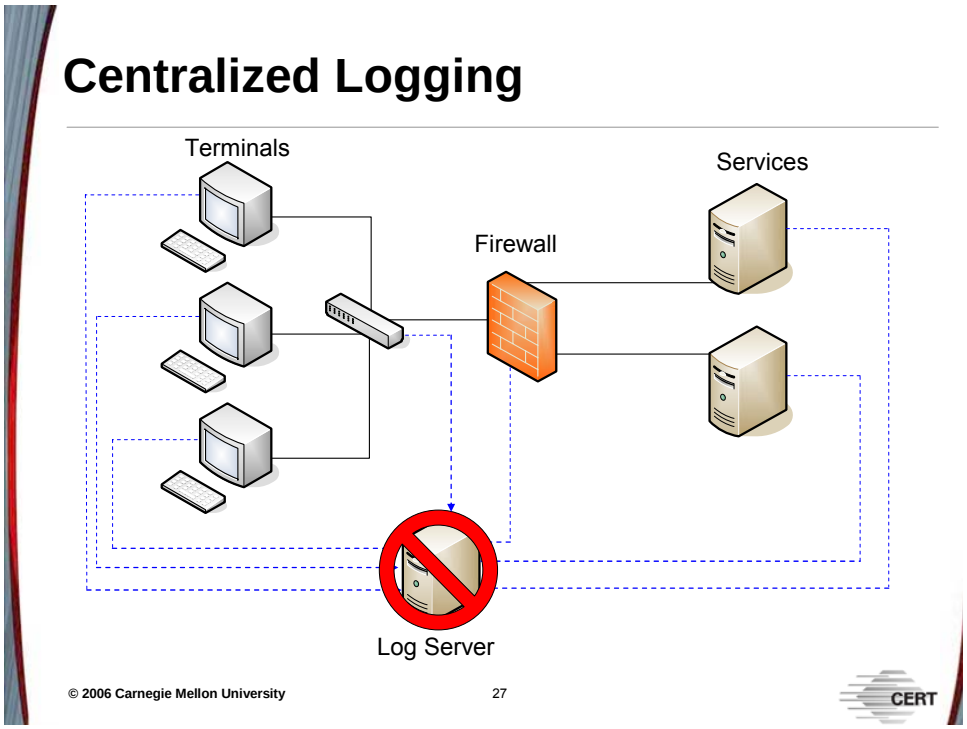
6.4.1 Log File Aggregation

Log files are an essential component of accountability management; however, they are probably the least structured component. As noted earlier in this module, there is no standard format for log files, although some formats such as syslog are more widely used than others.

Another issue regarding log files is their location—usually on the device[21] that generates them. This can make log file analysis a laborious and inefficient task for any organization that has more than a handful of devices. Log file aggregation and consolidation becomes

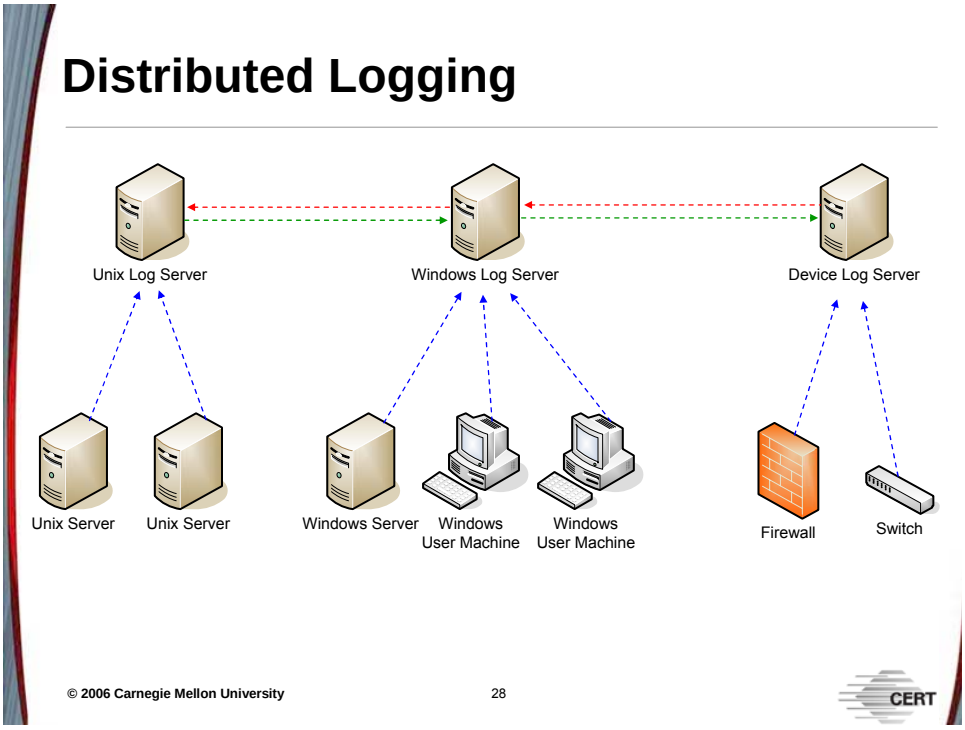
[21] For the purposes of this module, any machine that generates log messages will be referred to as a *device*.

necessary for logs to be useful to such organizations. When developing an accountability management program, an organization should consider two well-established log aggregation conventions: centralized logging and distributed logging.



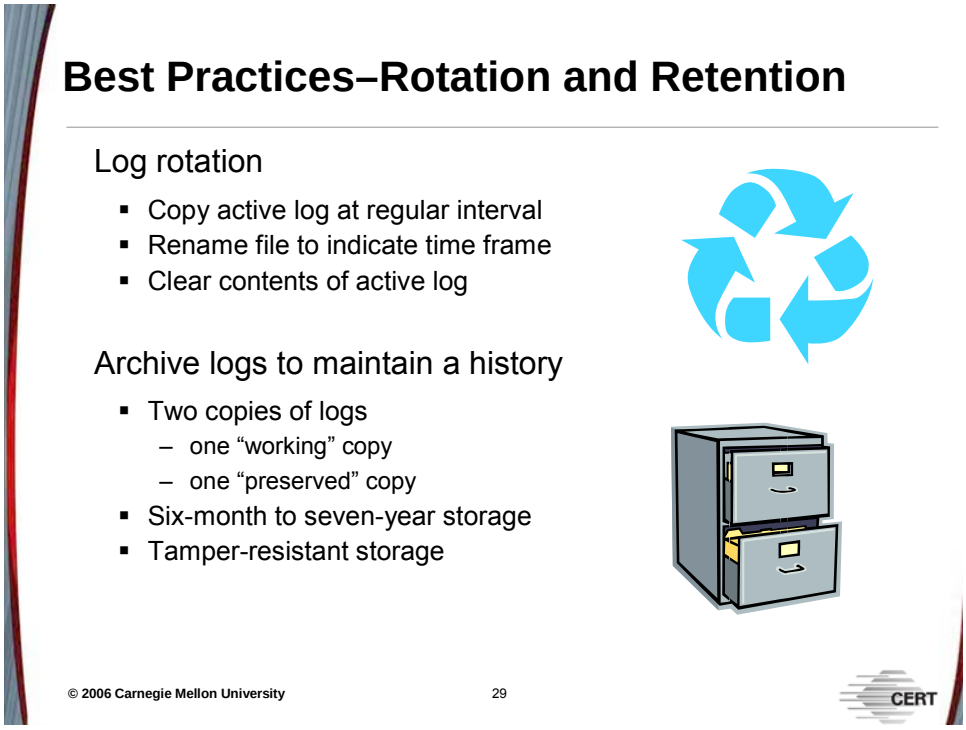
6.4.1.1 Centralized Logging

The diagram above shows a single server that collects logs from all the devices on a network. One of the greatest advantages of centralized logging is that it reduces the amount of overhead required for log file analysis. With all logs centralized in one location, analysts are better able to correlate events from different logs. Centralized logging also ensures better security for log files because IT staff can focus on protecting one machine rather than multiple machines, thereby reducing exposure to security threats. Despite its streamlining of many processes into one, however, centralized logging does involve risks. One distinct disadvantage is that the central log server becomes a single point of failure (SPOF). For example, if the server goes down, no log files will be collected and log file analysis will halt. Additionally, the entire set of logs will be put at risk if the server is compromised. Compromise of all log files would be assumed in such a scenario, rendering them unreliable.



6.4.1.2 Distributed Logging

Distributed logging employs multiple log servers as opposed to a single, centralized server. It is important to understand that distributed logging still involves aggregation of log files, but this aggregation is performed by a set of log servers. The diagram above shows a sample layout of a distributed logging setup. In this example, log file collection processes are divided among three log servers. One server is entirely dedicated to UNIX machines. The second server is for Windows devices, and the third is used for intermediary nodes in the network, such as switches and firewalls. The advantage of utilizing distributed log servers is that it provides redundancy to the organization's log management process. The diagram above shows that the three servers are synchronized in their log collection. As a result, if one of the servers failed, its logs would not be lost. It would also be possible for an organization to configure the servers to act as failovers to each other so that logging would continue as normal even if one server became unavailable. The downside of distributed logging is that it requires the maintenance of considerable resources, in terms of both money and humanpower. An organization may not have sufficient budget or equipment to dedicate multiple machines to the sole purpose of log file aggregation, for example. Furthermore, significant time is needed to configure the servers, monitor them to ensure they are functioning properly, and protect them from compromise.



6.4.2 Log File Rotation and Retention

6.4.2.1 Log File Rotation

Log file rotation involves dividing a log file into smaller, separate files. This is a good practice because it keeps logs logically organized and easy to maintain while also minimizing risk to the log files. If log messages were maintained in a single file, the file would eventually become very difficult to maintain due to its size. Moreover, if something were to happen to that file, all log data could be lost.

Before an organization begins to rotate its logs, it must first determine how often rotation should occur. In a large organization that generates a huge number of log messages, logs probably will need to be rotated every hour or even more frequently. On the other hand, smaller organizations may need to rotate their logs only once a week. This decision must be made on a case-by-case basis. Organizations should consider how much value they place on their log files and how much risk they are willing to incur. The more valuable the log files and the more risk-averse the organization, the shorter the rotation interval should be.

The process of log rotation is straightforward and consists of three main steps:

1. The active log file must be copied at a regular time interval.
2. The copy of the log file should be renamed to indicate the time frame that it represents. For example, if log files are being rotated every hour, then the naming convention 09042005_14.log may be used to indicate that this particular log represents the 14th

hour (2:00 p.m.) of September 4th, 2005. Regardless of what naming convention is used, it must be logical so that log messages from the past can be easily found.

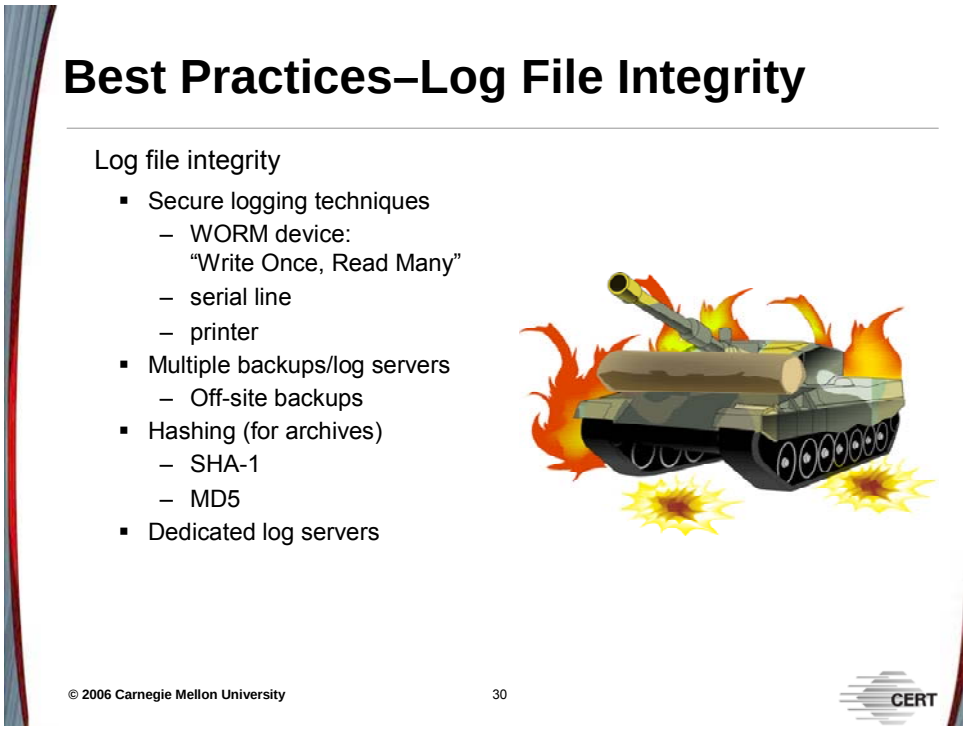
3. The active log should be cleared of its current contents so that it only contains messages within the new time interval.

6.4.2.2 Log File Retention

Organizations also must consider how they will handle retention of log files, which is important for maintaining a complete history of log data. The section "Regulatory Compliance" earlier in this module outlined a number of reasons why log file histories are vital. Furthermore, incidents detected in current logs may call for the examination of archived logs to find correlated events. How long log files should be stored is largely dependent on the organization's mission and available resources. For example, organizations that must comply with the Sarbanes-Oxley Act will need to store some log files for up to five years. Other organizations that are not bound by regulations and that have limited storage space may choose to archive their logs for a shorter period of time. Aside from these considerations, it is recommended that log files be archived for at least six months.

For reliability's sake, organizations should maintain two copies of their logs: a working copy and a preserved copy. The working copy can be used for everyday tasks, such as log file analysis, while the preserved copy serves as an undisturbed archive. If something were to happen to the working copy, malicious or not, the preserved copy would remain intact. Furthermore, maintenance of a preserved copy ensures accuracy, which is one of the three main properties needed for a log file to hold up as evidence in a legal matter.

Lastly, it is imperative that log file archives be tamper resistant. A preserved copy of a log file is only useful if it is well protected. Malicious users often try to alter log files in an attempt to hide their tracks. If attackers are insiders, they can pose an even greater threat to log files. Therefore, safeguards must be put in place to prevent archives from being altered. Specific implementations of such safeguards will be reviewed in the next section, "Log File Integrity."



6.4.3 Log File Integrity

6.4.3.1 Secure Logging

As explained in the previous section, maintaining log file integrity is important because it assures accuracy. This section will focus on a variety of methods that can be used to ensure log file integrity.

First, a write-protected medium can be utilized to prevent log files from being altered; the most popular such medium is a CD-R. A CD-R is a write-once-read-many (WORM) medium, which is ideal for integrity purposes because the file cannot be changed once it is written to the disc. However, keep in mind that a CD-R can be easily damaged if it is not stored properly or is mishandled. There is also some debate over the lifespan of a CD-R, with some people arguing that its lifespan is limited to about 10 years. However, those claims have not been backed up with concrete evidence. The best approach is to acknowledge that the CD-R is a relatively new medium and that its long-term longevity is still somewhat uncertain.

Another logging technique involves use of an isolated machine that is not accessible from the network but has an attached serial line for receiving log data. The advantages of this method are that it allows for a large storage space and is isolated from electronic attacks. However, the machine must reside in a secure location, as it will still be susceptible to physical attacks.

A third alternative for reliable logging is to write log messages to a printer in real-time. This printer would have to be placed in a secure location, and the organization would need

sufficient storage space for the logs, which also could serve as an offline backup if online log data were lost. Analyzing printed logs is not as efficient or effective as analyzing electronic logs. However, if the log data is important enough, this may be a tradeoff that an organization is willing to make.

6.4.3.2 Backups

Maintaining backups of log files is an important aspect of log file integrity because these backups can be used to verify the integrity of the original logs. Ideally, backups should be automated to guarantee that they are generated and to protect against human error. How often log files are backed up is largely an organizational decision. In general, as the value of log data increases, the backup frequency should increase as well.

It is also recommended that multiple backups be created and stored in physically separate locations. Ideally, backups should be stored in at least one off-site location, which will help protect against physical threats such as fire. And if an organization is located in an area that is prone to natural disasters such as floods, earthquakes, hurricanes, tornadoes, or wildfires, it should consider procuring off-site storage in a completely different region. As with backup frequency, the organization's risk tolerance should dictate the total number of backup copies it produces and maintains. Two copies may be sufficient for many organizations, while others may choose to have more. When developing backup policies, it is important to consider log retention policies, and vice versa; the two are very closely related.

6.4.3.3 Hashing

Creating cryptographic hashes of archived log files is an excellent way to assure the integrity of a file. Hashing algorithms such as SHA-1 (The Secure Hashing Algorithm 1) and MD5 (Message-Digest Algorithm 5) take an arbitrary length input and then output a fixed-length message digest. SHA-1 produces a 160-bit message digest, while MD5 generates a 128-bit digest. Hash functions can be used to generate digests for log files that are then stored in a secure location. To check the integrity of a log file, you would create a new digest using the hash function and compare it against the original digest. If the two digests are the same, you can be assured that the files have not been altered. Conversely, if changes have been made to the original log file, the digests will not match.

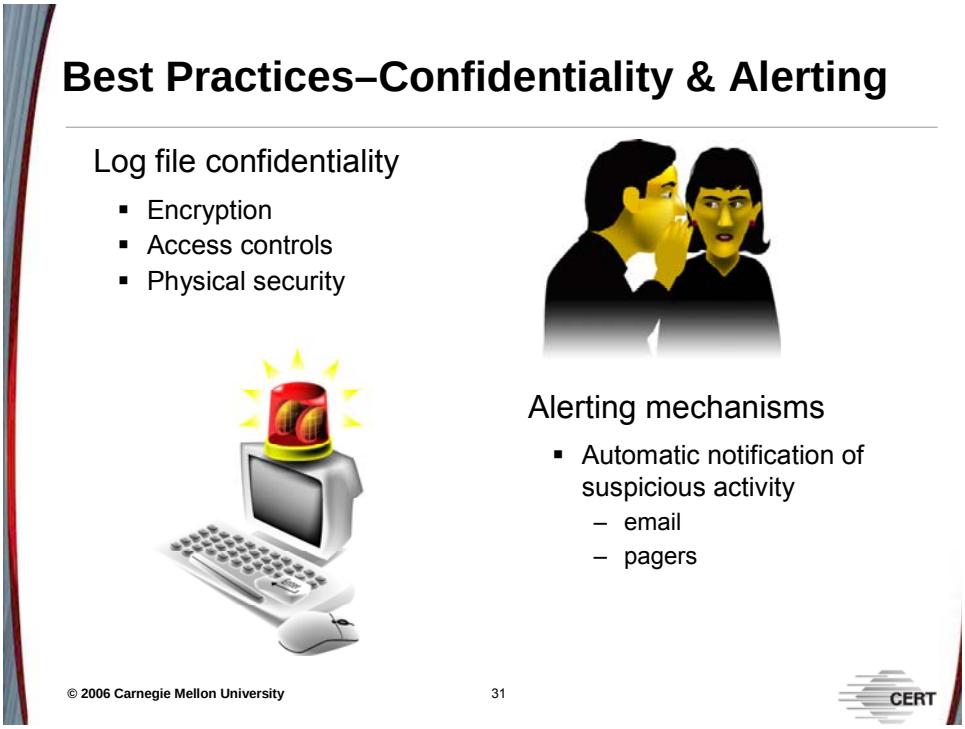
The reason why cryptographic hash functions can be used to provide log file integrity is because they are "weak collision resistant." In this context, a *collision* refers to two separate inputs producing the same output. A weak collision occurs when an input can be found whose digest matches the digest of a different, targeted input. For example, if the message "I owe Johnny $100" had a digest value of A51E1, a weak collision would occur if I determined that the message "I owe Johnny $1" had the same digest value. You can see from this example that if weak collisions are possible, digests cannot be used to ensure the files have

not been altered. Fortunately, cryptographic hash functions are resistant to weak collisions by nature.

There are a few important points regarding the use of hashing for file integrity that should be addressed. First, digests can only be generated for inactive log files. This is because active log files constantly change as new entries are added. Second, log file hashing is a very space-efficient means for ensuring log file integrity. Even if a file contains gigabytes worth of log data, the size of its digest always remains the same. Third, when deciding which hash function to use, you should understand that SHA-1 is considered to be the successor of MD5, which has had flaws discovered in its algorithm. Also, SHA-1 has a maximum input length of 2^{64} bits; any additional input will be ignored. Since no log file should even come close in size to 2^{64} bits, this limitation is not a significant issue.

6.4.3.4 Dedicated Log Servers

Utilizing dedicated log servers and isolating them from the network as much as possible will reduce the number of interactions these servers have with other processes. This is good because the more services and applications that run on a machine, the more entry points there are for an attacker to compromise log files. A dedicated server eliminates the possibility of other services and applications interfering with logs stored on the server. This practice is in keeping with the security concept of segregation of duties—also a best practice.



6.4.4 Log File Confidentiality

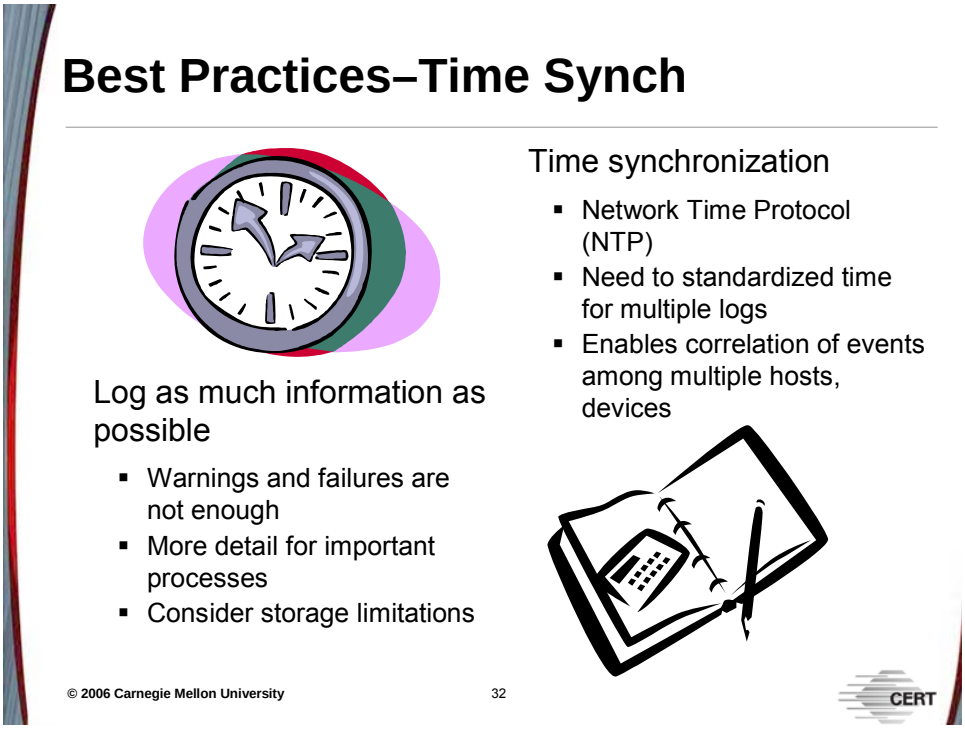
The data in an organization's log files often contains intimate details about the organization's IT structure and business processes. An individual with malicious intent could use this type of information for reconnaissance prior to launching an attack. Additionally, message logs may contain information that is sensitive in nature, and authorization policies may restrict who can view log data. With these considerations in mind, it is important to preserve the confidentiality of log file data. One of the most popular and effective methods for ensuring confidentiality is to encrypt the data, both during transmission and in storage. Assuming the decryption key has not been compromised, this practice guarantees that only authorized parties can decipher and view the data. One way to encrypt log traffic is to utilize Secure Shell (SSH), which creates an encrypted tunnel between two communicating parties. The downside of this type of implementation is that it only provides confidentiality during transmission. The logs will remain unencrypted in their storage facility and will be at risk of exposure. An alternative is to encrypt the log data before transmission, thereby ensuring end-to-end secrecy.

Another practice that can help reinforce confidentiality is to employ access-control measures. Only authorized employees should be able to log in to log servers, electronically access log files, or be granted access rights to log server networks. When access-control policies are being written, management needs to determine who should have access to log data. Proper access controls should then be put in place. This is very important; in a sense, malicious insiders are much more dangerous than outside threats because they have many more resources and opportunities available to them.

Whereas electronic access-control measures aim to prevent unauthorized access to digital resources, physical security is designed to prevent physical interaction with these resources. Storing log servers and archived media in physically restricted areas adds an extra level of confidentiality. None of these measures alone provides a complete solution; each is fallible in its own way. Encryption keys can be cracked, access controls can be configured incorrectly, users can forget to log out, and secure areas can be breached as a result of propped-open doorways or piggybacking. However, when combined, these measures create multiple layers that must all be bypassed for confidentiality to be breached.

6.4.4.1 Alerting Mechanisms

At the end of the workday, IT resources often stay up and running even when the staff goes home. Some business processes are left to run 24 hours a day, seven days a week. Realistically, accountability processes cannot be monitored by staff members at all hours of the day, even during regular work hours. In fact, committing staff power to constantly monitor accountability processes is a waste of resources. It is much more efficient to set up mechanisms that alert IT staff when certain "trigger" activities or messages are encountered. This not only allows staff members to commit their time to other tasks, but also ensures that flagged events will be detected and brought to staff members' attention quickly. Many monitoring processes are capable of sending email alerts when a specific, predefined criterion is met. Similarly, alerts can be sent via pagers and cell phones. It is important for an organization to determine the kinds of events and activities that should be flagged. Weights should also be assigned to give the different flags a hierarchy of importance. Alerts that do not require immediate attention can be sent through email, while more urgent alerts can be sent via pager or cell phone.



6.4.5 Time Synchronization

When logs are being collected from multiple devices, time synchronization becomes a critical asset. This is because time synchronization can create a definitive chronology of events for messages pulled from different devices. If an attack on a host is detected through the host's logs, for example, the next logical step may be to analyze the firewall log to see if the attack's entry into the network can be found. If the clocks of the host system and the firewall are set to different times, the offset must be calculated in order to examine the logs in the correct sequential order. This is a tedious task and becomes impossible to perform manually when dealing with hundreds of separate devices. There are tools available that are capable of correcting timestamp offsets of existing log files; however, there are limitations to what they can do. In essence, time synchronization can be performed using the Network Time Protocol (NTP). The NTP transmits timestamp information via UDP port 123. The NTP servers can be set up to synchronize with external clocks such as the National Institute of Standards and Technology (NIST) time server or a GPS clock.

6.4.5.1 What to Log

When setting up logging capabilities, it is important to record more than just warnings and failures. Doing so provides only a partial picture of what is going on in the network and directly conflicts with the best practice of ensuring completeness of log data. Recording only warnings and failures can be deceptive and can provide a false sense of security. In some ways, normal log messages are more important because they often show successful activities, including those of an anomalous and malicious nature. For example, suppose the audit log

on a Windows machine is set up to record only failures. If an attacker were to compromise an account, successfully log on, and escalate the account's privileges, none of these activities would be recorded in the audit log. This is because all of them were executed successfully despite unauthorized changes. As a result, analysis of the audit log will show no trace of the aforementioned attack. The lesson to be learned is that as much information should be logged as is possible given the organization's available storage capacity. Since storage is not limitless, policies should be created to define the types of logs and log messages that are most valuable to the organization.

Accountability Management Summary

Accountability management

- Essential for a sound information security implementation
- Provides visibility to IT resources and components

Wide variety of implementations–no one method is a complete solution

Understand the big picture–tailor to the organization's mission

- Best practices
- Implementations

Summary

By now it should be evident that accountability management is a key part of any organization's Defense-in-Depth strategy. Accountability management offers the following benefits: it provides visibility to an organization's IT resources; helps preserve audit trails; assists in legal matters and regulatory compliance; and enables organizational policies to be quantifiably measured.

It is also important to understand that careful and meticulous planning is needed to develop a sound accountability management program. Critical assets must be identified, and a wide variety of accountability management implementations must be considered. Lastly, an organization should incorporate best practices into its accountability management program, because these practices have proven benefits and can help conserve resources.

This module was designed to serve as a foundation for the subject of accountability management and to help you develop a solid foundation of knowledge. From here, you should be able to research specific topics in more detail. More importantly, managers should be able to make sound decisions with regard to accountability management that will positively affect the organization as a whole.

Review Questions -1

1. Name one advantage and one disadvantage of having a central log server.
2. Name two federal regulations that require the use of accountability management.
3. What is the purpose of log rotation?
4. What are two methods that can be used to ensure log file integrity?
5. What three properties should a log file possess if an organization wishes to use it as evidence in a court of law?

Review Questions -2

6. Why is time synchronization important for maintaining log files?
7. What protocol is used to transmit messages to a syslog server, and does it have any disadvantages?
8. What type of system is best suited for integrity host monitoring?
9. Why would an organization use bandwidth utilization?
10. What are the two types of network intrusion detection systems, and how do they work?

Module 7: Availability Management

This module covers best practices for ensuring availability, system properties involved, and planning for business continuity and disaster recovery.

Instructional Objectives

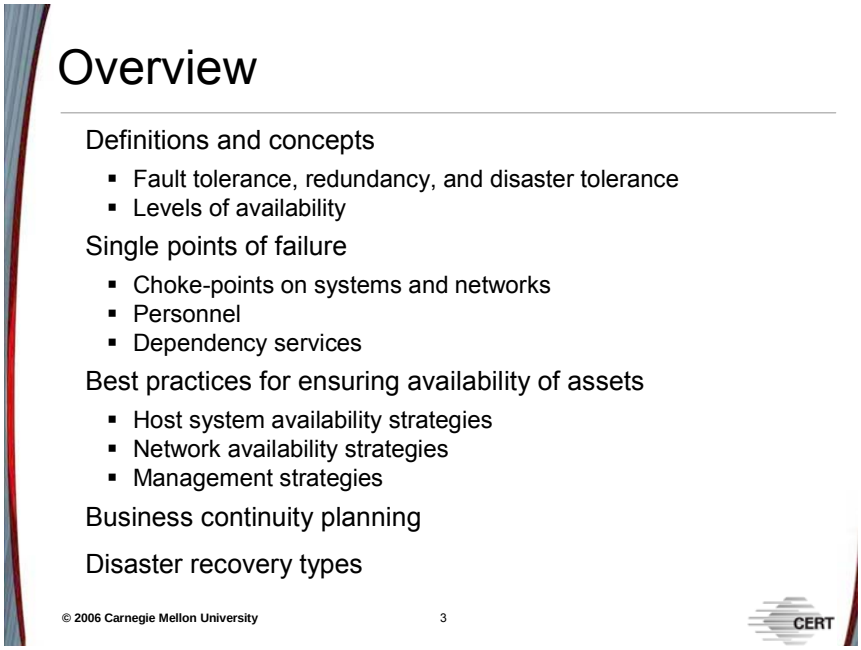
Upon completion of this module, students will be able to

- Define reliability, fault tolerance, and failover
- Identify three levels of availability
- List three potential single points of failure (SPOF)
- Describe two best practices for ensuring the availability of assets
- Name the elements of business continuity planning
- Identify three types of disaster recovery

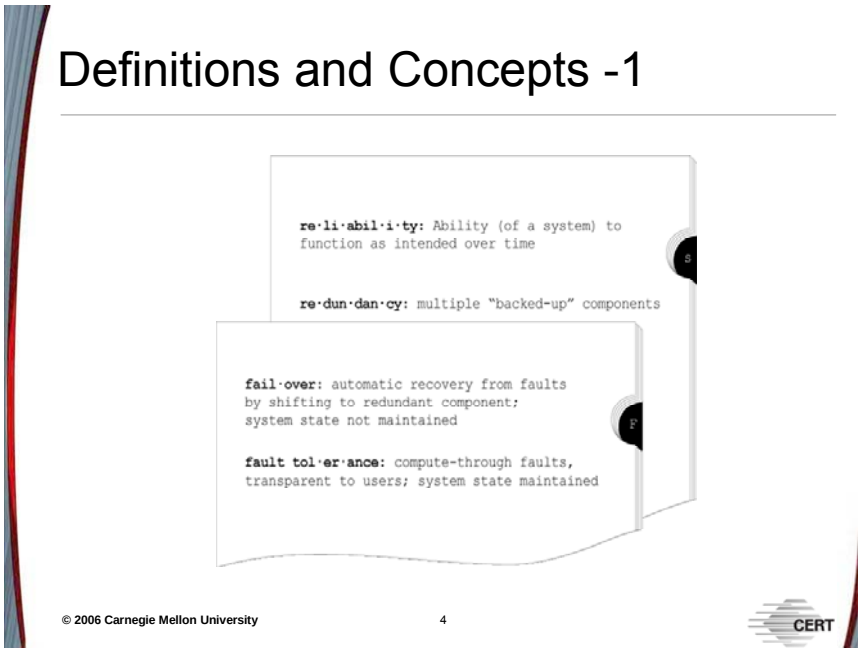
CERT

This instructional module will enable students to complete all of the above learning objectives.

7 Overview of Availability Management



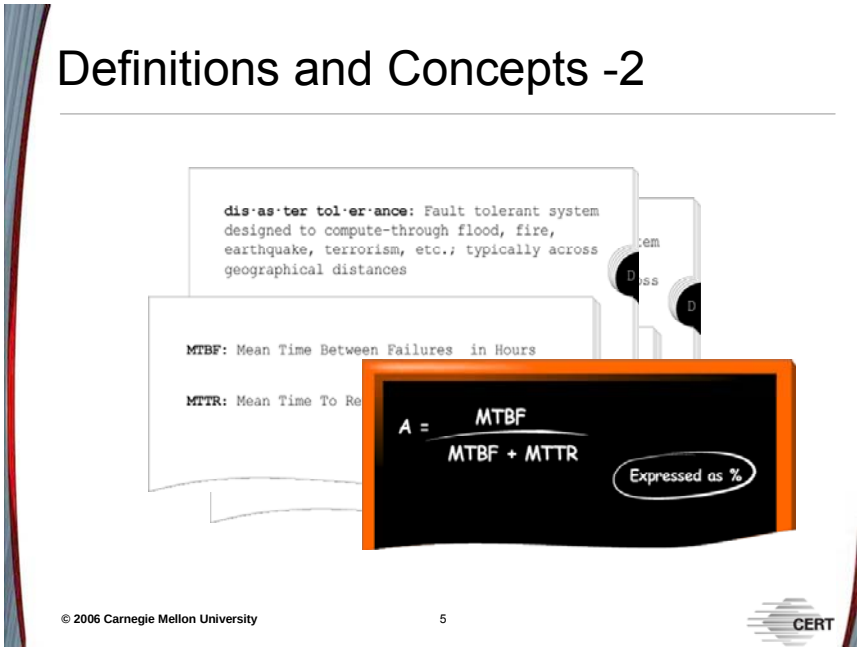
This module focuses on the importance of availability of information assets to most organizations. Downtime is no longer merely an inconvenience. It is associated with loss of revenue, reputation, market share, and even life and limb. IT managers must address real or artificially imposed availability requirements on information assets and are faced with the dilemma of how best to ensure uptime with limited resources.



7.1 Definitions and Concepts

To fully understand the concepts presented in this module, you must understand some basic terminology:

- **Reliability** – the ability of a system or component to perform its required functions under stated conditions for a specified period of time [IEEE 90]
- **Redundancy** – having one or more "backup" systems available in case the main system fails [Newton 06]
- **Failover** – a backup operational mode in which the functions of a system component (such as a processor, server, network, or database) are assumed by secondary system components when the primary component becomes unavailable through failure or scheduled downtime
- **Fault tolerance** – the ability of a system or component to continue normal operation despite the presence of hardware or software faults [IEEE 90]



Disaster tolerance systems are designed with enough redundancy and fault-tolerant features that they can stay resilient despite building-level (at a minimum) disasters. Some systems are so mission critical that redundant replica systems (based on different code and hardware but with the same general requirement specifications) are designed so that the primary system can fail entirely but the mission still survives. Examples include military systems that control weapons systems and hospital systems that control all aspects of patient care.

A few other terms with which you should be familiar are below [Relex 01].

- **Mean time between failures (MTBF)** – the average number of hours that pass before a component, assembly, or system fails. It is a basic measure of reliability for repairable items and a commonly used variable in reliability and maintainability analyses.

  MTBF can be calculated as the inverse of the failure rate for constant failure rate systems. For example, if a component has a failure rate of 2 failures per million hours, the MTBF would be the inverse of that failure rate [MTBF = (1,000,000 hours) / (2 failures) = 500,000 hours].[22]
- **Mean time to repair (MTTR)** – the most common measure of maintainability. It is the average time required to perform corrective maintenance on all of the removable items in a product or system. This kind of maintainability prediction analyzes how long repair and maintenance tasks will take in the event of a system failure.

  MTTR also factors into other reliability and maintainability predictions and analyses. MTTR can be used in a reliability prediction to calculate the *availability* of a product or system.

[22] See http://www.i-mtbf.com/.

- **Availability** – the probability that an item is in an operable state at any time, based on a combination of MTBF and MTTR.[23]

Availability in its simplest form is the time a system is functioning normally. This can also apply to any management component of the IT environment, such as a storage area network (SAN), LAN, WAN, applications, servers, and so forth. Availability also applies to the IT service as a whole or any component thereof, right down to each integrated circuit.

Below is a simple equation to calculate availability, where A is the degree of availability expressed as a percentage, MTBF is the mean time between failures (or uptime) and MTTR is the maximum time required to repair or resolve a particular problem:

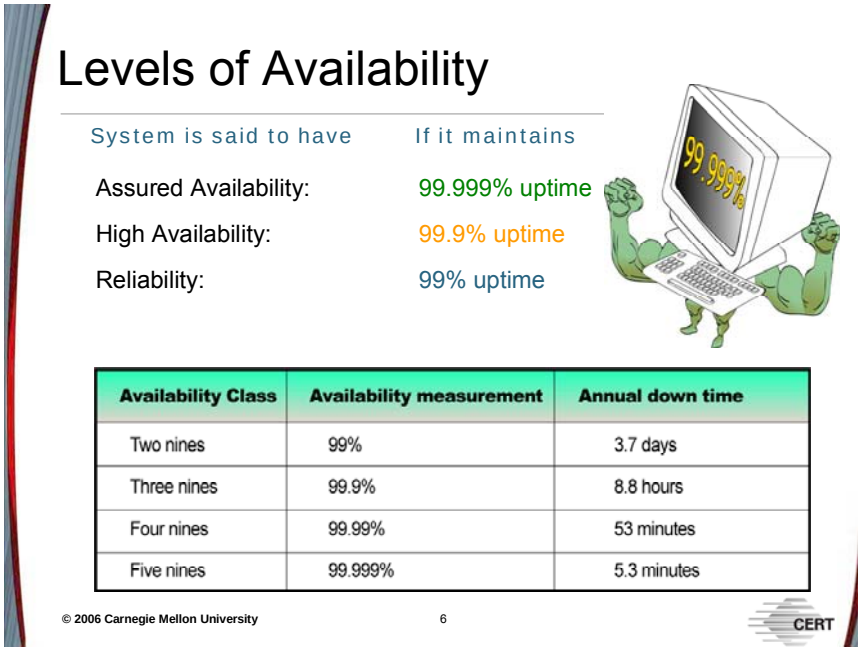
$$A = \frac{MTBF}{MTBF + MTTR}$$

Therefore:

As MTTR approaches zero, A increases toward 100%
As MTBF gets larger, MTTR has less impact on A

For example, if a server has an MTBF or uptime of 100,000 hours and a maximum repair time (MTTR) of 1 hour, then it has a rather impressive availability level of 100,000/100,101, or 99.999%. Reducing the MTTR to 6 minutes, or one-tenth of an hour, increases availability an extra .0009, to 99.9999%. However, keep in mind that a MTBF of 100,000 hours (more than 11 years) is difficult to achieve in reality.

[23] See http://www.mttr.net/.



7.2 Levels of Availability

System availability is generally broken up into classes, known as "the rule of 9s." The number of 9s in the calculated availability percentage corresponds to its class of availability. The following table is excerpted from Dr. Robert Glorioso's high-availability paper [Glorioso 05]:[24]

1. Attribute	1. Assured Availability	1. High Availability	1. Reliability
2. Uptime	2. 99.999%	2. 99% to 99.95%	2. 98% to 99%
3. Recovery Time	3. milliseconds	3. minutes to hours	3. hours to days
4. Fault Handling	4. compute through	4. failover/failback	4. reboot
5. Redundancy	5. no single point of failure	5. possible points of failure	5. multiple points of failure
6. Fault Performance	6. 100%	6. ((N-1) / N)%	6. 0%
7. Human Intervention	7. none	7. coding, scripting, administration	7. administration

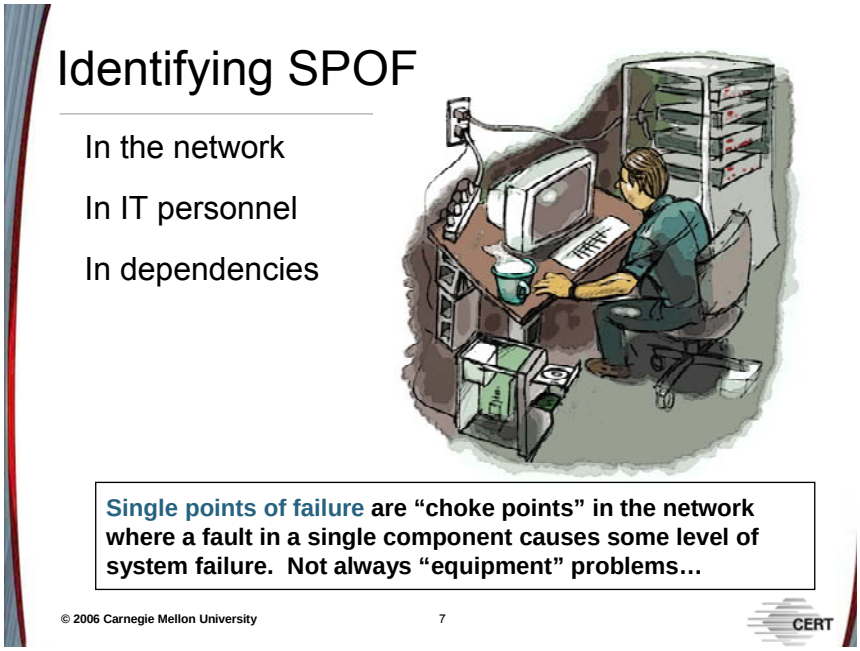
[24] See http://www.disastertolerance.com/aawhitepaper.htm.

It is helpful to translate these numbers into business-case scenarios that describe the rationale for investing in systems designed for high availability. The Harvard Research Group (HRG) has stated that high availability must be defined independently of technologies employed to achieve it. HRG defines systems availability in terms of the impact of system downtime for the business and for the consumer (end user) of the service. HRG's six Availability Environment Classifications (AEC),[25] described below, are a first step toward defining availability in terms of the impact on both the business and the end user or consumer [HRG 06].

- **Disaster Tolerant** (AEC-5) – Business functions that absolutely must be available and for which any failure must be transparent to users. This means no interruption of work, no lost transactions, no degradation in performance, and continuous computing services that are safe even from disasters such as earthquakes, fires, floods, hurricanes, power failures, vandalism, or acts of terrorism.
- **Fault Tolerant** (AEC-4) – Business functions that demand continuous computing and for which any failure is transparent to users. This means no interruption of work, no transactions lost, no degradation in performance, and continuous 24/7 operations.
- **Fault Resilient** (AEC-3) – Business functions that require uninterrupted computing services during essential time periods or during most hours of the day and days of the week year-round. This means that users stay online despite failures, although current transactions may need to be restarted and users may experience performance degradation.
- **High Availability** (AEC-2) – Business functions that allow minimally interrupted computing services, either during essential time periods or during most hours of the day and days of the week throughout the year. This means that users may be interrupted but can quickly log on again. Users may have to rerun some transactions and may experience performance degradation.
- **Highly Reliable** (AEC-1) – Business functions that can be interrupted as long as the integrity of the data is assured. From users' perspectives, work stops and uncontrolled shutdown occurs.
- **Conventional** (AEC-0) – Business functions that can be interrupted and for which data integrity is not essential. From users' perspective, work stops and uncontrolled shutdown occurs. Data may be lost or corrupted.

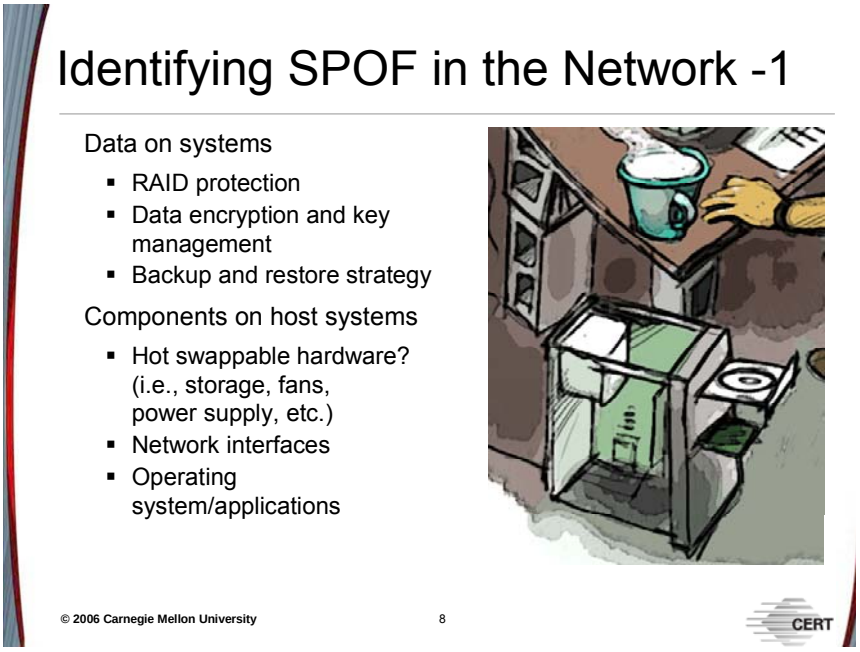
Once you determine which classification your systems require, you can begin considering specific implementation details to achieve the desired level of availability.

[25] See http://www.hrgresearch.com/.
Popular categories can be found at http://tingurl.com/Iv6r2.



7.3 Single Points of Failure (SPOF)

As an IT manager, single points of failure (SPOF) will eventually cause problems. As part of prudent risk management, you should seek out all SPOF in your daily operations, conduct risk analysis, and finally make risk management decisions. It is important to remember that SPOF are not only a hardware and software issue, but also affect management of IT personnel, dependent services, and so on.



7.3.1 Single Points of Failure in the Network

Critical data stored on host systems can be identified as a single point of failure in and of itself. Some examples would be information contained in databases and HTML pages stored on Web servers.

7.3.1.1 RAID Protection

It is advisable to use RAID on systems that contain critical data [RAID 04]. RAID can provide fault tolerance for data, such that end users are not likely to notice if one hard drive fails (although failures can affect performance levels). The following describes RAID and its six levels:[26]

> What is RAID?
>
> RAID is an acronym for Redundant Array of Inexpensive (or Independent) Disks. A RAID array is a collection of drives which collectively act as a single storage system, which can tolerate the failure of a drive without losing data, and which can operate independently of each other.
>
> What are the different RAID levels?
>
> A research group at UC-Berkeley coined the term "RAID", defining six RAID levels. Each level is a different way to spread data across multiple drives--a

[26] See http://www.datarecoveryclinic.com/raid_data_recovery.htm.

compromise between cost and speed. Understanding these levels is important, because each level is optimized for a different use.

RAID Level 0
RAID Level 0 is not redundant, hence does not truly fit the "RAID" acronym. In level 0, data is split across drives, resulting in higher data throughput. Since no redundant information is stored, performance is very good, but the failure of any disk in the array results in data loss. This level is commonly referred to as striping.

RAID Level 1
RAID Level 1 provides redundancy by duplicating all data from one drive on another drive. The performance of a level 1 array is only slightly better than a single drive, but if either drive fails, no data is lost. This is a good entry-level redundant system, since only two drives are required; however, since one drive is used to store a duplicate of the data, the cost per megabyte is high. This level is commonly referred to as mirroring.

RAID Level 2
RAID Level 2, which uses Hamming error correction codes, is intended for use with drives which do not have built-in error detection. All SCSI drives support built-in error detection, so this level is of little use when using SCSI drives.

RAID Level 3
RAID Level 3 stripes data at a byte level across several drives, with parity stored on one drive. It is otherwise similar to level 4. Byte-level striping requires hardware support for efficient use.

RAID Level 4
RAID Level 4 stripes data at a block level across several drives, with parity stored on one drive. The parity information allows recovery from the failure of any single drive. The performance of a level 4 array is very good for reads (the same as level 0). Writes, however, require that parity data be updated each time. This slows small random writes, in particular, though large writes or sequential writes are fairly fast. Because only one drive in the array stores redundant data, the cost per megabyte of a level 4 array can be fairly low.

RAID Level 5
RAID Level 5 is similar to level 4, but distributes parity among the drives. This can speed small writes in multiprocessing systems, since the parity disk does not become a bottleneck. Because parity

data must be skipped on each drive during reads, however, the performance for reads tends to be considerably lower than a level 4 array. The cost per megabyte is the same as for level 4.

Which RAID level should I use?

The right choice depends on the application. The following table summarizes the RAID levels with some of their possible uses.

RAID Level Uses

Level 0 (striping)
Any application which requires very high speed storage, but does not need redundancy. Photoshop temporary files are a good example.

Level 1 (mirroring)
Applications which require redundancy with fast random writes; entry-level systems where only two drives are available. Small file servers are an example.

Level 4 (parity)
Applications which require redundancy at low cost, or with high-speed reads. This is good for archival storage. Larger file servers are an example.

Level 5 (distributed parity)
Similar to level 4, but may provide higher performance if most I/O is random and in small chunks. Database servers are an example.

Often, it makes sense to use more than one level. For instance, in a two-drive system, one partition could use level 0 to offer the highest performance for temporary files; another partition could use level 1 to offer security for important data or applications. In a three-drive system, a partition for temporary files could use level 0, the boot disk could use level 1, and large data files could be stored on a level 4 partition.

There are hardware and software RAID systems available. Typically, hardware systems provide better performance and features, (like hot-swappable hard drives) but are more expensive than software RAID systems. Windows Server 2000/2003 comes with RAID levels 1 and 5 built into the capability of the operating system.

7.3.1.2 Data Encryption and Key Management

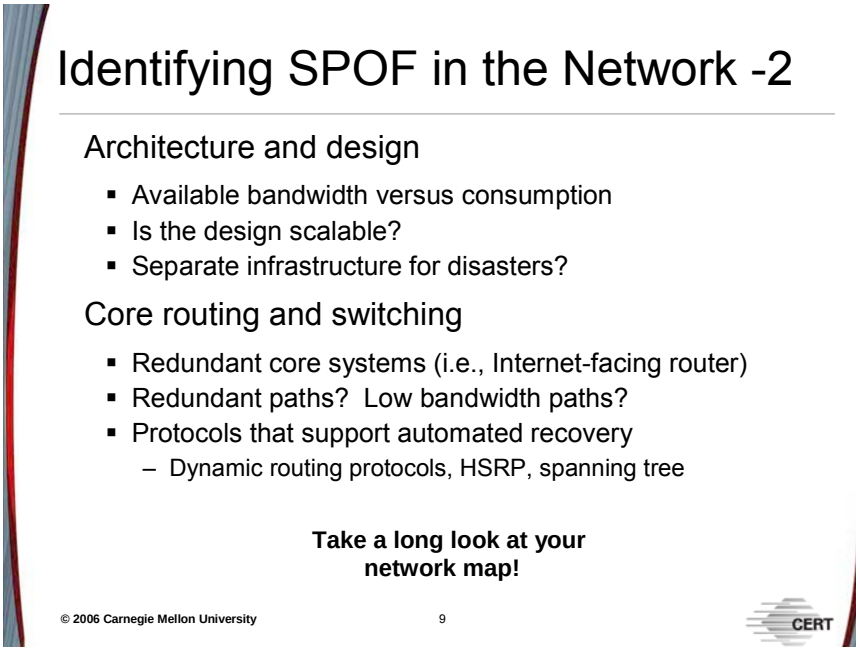
As mentioned in the Identity Management module, cryptographic key management is very important. If data is encrypted and the key is not escrowed (backed up and protected) then the key itself becomes a SPOF. Unless you're careful and follow good key management practices, your data may be so secure that it will be unavailable even to you. Therefore, a best practice is to escrow cryptographic keys either with a trusted department within the organization or with a trusted third party.

7.3.1.3 Backup and Restore Strategy

If you experience a failure or incident that renders your data unavailable, you likely will have to rely on the effectiveness of your restore capability. SPOF exist if you don't have multiple copies of backed up data or if all of your backup tapes are stored in the same location. Offsite storage in disaster-proof containers is a good practice. Your backup data is only as good as your ability to restore it. Therefore, conduct practice restores regularly.

7.3.1.4 Hot-Swappable Hardware, Network Interfaces, Operating Systems, and Applications

To eliminate SPOF further, host systems can have redundant processors, fans, power supplies, storage, and network interfaces. The ability to remove and replace failed components while the system is still running (components are "hot-swappable") is a great benefit.



7.3.1.5 Architecture and Design

One of the most effective tools for an IT manager can be a printed map of the organization's network. Maps are great for identifying potential SPOF and other potential problems. Data flows can be visually traced throughout the network and can be used to further illustrate issues with the architecture as a whole.

Take a look at your bandwidth availability at peak and off-peak usage times. New applications (e.g., Kazaa, IPtv) and usage characteristics can change your bandwidth availability. If your bandwidth is more than 90% utilized on a regular basis, you may have SPOF issues if utilization spikes occur. Tools such as Multi Router Traffic Grapher[27] can help you visually track bandwidth utilization over time [Oetiker 06]. Additionally, traffic management utilities such as Packeteer[28] and Linux Advanced Routing and Traffic Control[29] allow you to throttle dedicated portions of your bandwidth to various users, protocols, and applications [Packeteer 06, Linux 06].

You will also want to monitor your host's systems to ensure they are the correct size for the load/usage to which they are subjected by applications, services, and users. This monitoring should be conducted throughout your architecture so that potential SPOF can be avoided. For example, if you have a router connecting very busy network segments (maybe traffic has increased substantially since its initial installation), you may eventually have a SPOF due to the now-undersize router. Windows and UNIX systems have extensive system monitoring capabilities built in that can issue alerts if thresholds are exceeded. Most routers and

[27] See http://www.mrtg.org.
[28] See http://www.packeteer.com/.
[29] See http://lartc.org/.

switches have similar capabilities, and many network management systems can centrally monitor multiple network systems (i.e., HP Openview, Tivoli, etc.).

These considerations raise some questions:

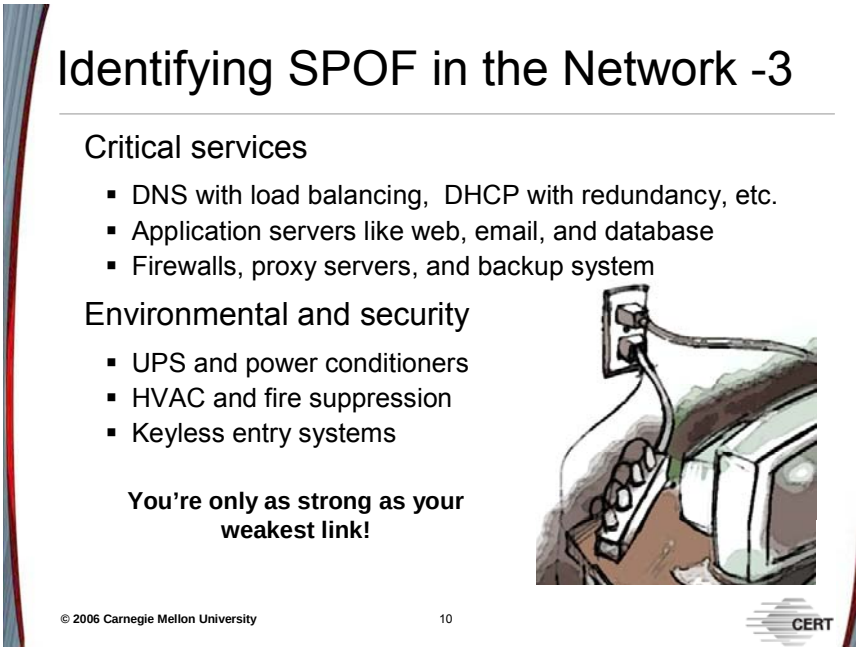
- Is the overall design of your network and its components scalable to meet future growth/usage projections? Has this even been considered?
- Is any portion of your network so critical to the mission of the organization that it requires a completely separate standby infrastructure? (See the management best practices presented later in this module for examples.)

7.3.1.6 Core Routing and Switching

Core routers and switches certainly have SPOF potential, since they are central to almost all network communications. Most organizations' connection to the Internet is via a border or "gateway" router. This device is frequently a SPOF and is often overlooked by IT managers. Single paths to destination networks (leased lines, ISDN, Frame Relay, and dial-up) are potential problem areas as well. Be sure to evaluate bandwidth utilization on these links and consider the criticality of any traffic that traverses them. Could your organization tolerate an extended outage of any one of these links?

Protocols that help promote availability of systems and networks are certainly worth investigating. The spanning tree protocol (STP) is designed to eliminate broadcast storms in switches. However, it also provides redundancy if there are multiple switches and network paths. If a switch fails, after a minute or so STP will automatically failover and select other redundant switches to use. Dynamic routing protocols (OSPF, BGP, EIGRP, RIP, etc.) have built-in failover mechanisms that will calculate a new route to the network (if one is available) when a router or destination network fails. Cisco's Hot Standby Router Protocol is designed to failover to a standby router if the primary router fails. Other vendors and open source solutions are available as well.

Ask yourself whether your organization can tolerate the downtime involved during the failover process. This question may lead you to perform a business analysis to determine the answer.



7.3.1.7 Critical Services

Critical services are often the most important systems to protect with fault-tolerant solutions. Consider potential SPOF in these systems and make prudent risk management decisions, using the tools and techniques discussed in the Risk Management module of this course.

The Domain Name System (DNS) is one of the most critical services for many organizations because it translates domain names into IP addresses, and vice versa (e.g., http://www.cert.org = 192.88.209.14). This functionality is essential because it is much easier for human beings to remember domain names than IP addresses. Use of multiple DNS servers to provide redundancy is common, but DNS load balancing for other services, such as Web and email servers, may not be implemented as frequently despite its status as a best practice. Load balancing is accomplished by having multiple IP addresses (multiple servers) all resolve to the same domain name in DNS. For example, "www.mywebsite.com" can have multiple records in DNS ([10.1.1.1], [10.1.1.2], [10.1.1.3]), with each address corresponding to a redundant Web server. When name resolution requests for www.mywebsite.com are handled by a DNS server, corresponding IP addresses will be returned in serial order (also called round-robin DNS), thereby easing the burden on each of the three Web servers. If one Web server fails, the user need only resolve the hostname through DNS again to be directed to one of the available servers.

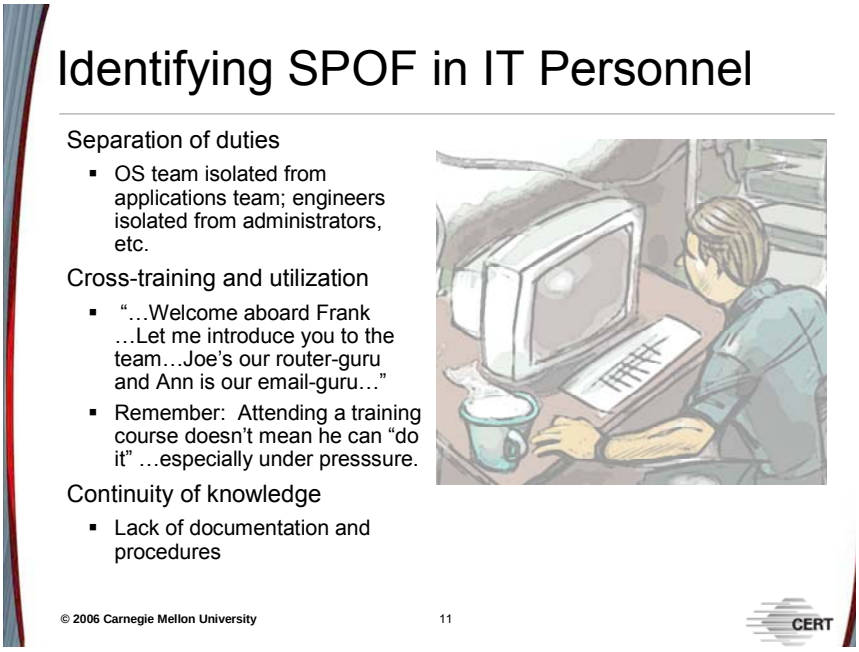
Redundancy and fault tolerance in other network services like Dynamic Host Configuration Protocol (DHCP) is advisable, too, since DHCP performs the critical task of issuing IP addresses and other network configuration information to host systems. Lastly, firewalls typically filter all inbound and outbound network traffic and are very commonly SPOF. Consider a high-availability solution for this critical service.

7.3.1.8 Environmental and Security Systems

One of the most common areas for SPOF involves supporting systems like environmental and security systems. Consider these systems carefully when conducting risk assessments, by asking the following:

- Do you have backup air coolers for your data center?
- Do these coolers run off a different power source than the primary ones do?
- Is the physical security of your datacenter protected by a keyless entry system?
- If this system crashes or loses power, does it fail open (unlock the doors)?
- Do you have plain old locks and keys on the doors for redundancy?
- Do you have multiple sources of power redundantly configured to support your critical systems?
- Do you have battery backup (uninterruptible power supply) capabilities for these systems in case of a catastrophic failure or natural disaster at the power company?

In the end, the takeaway point is that IT managers have many responsibilities and tasks, and effective planning is the best practice for them all. By asking the difficult questions up-front, you can prevent SPOF situations and deal effectively with any problems that do arise.



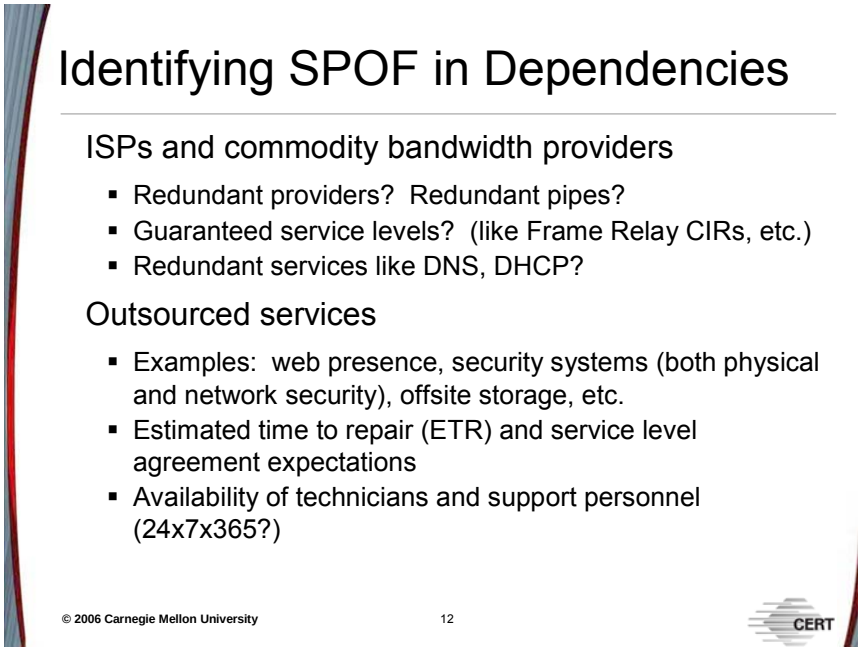
7.3.2 SPOF in Dependencies

As mentioned previously, SPOF can occur throughout IT operations. One of the most overlooked SPOF involves IT personnel. IT managers should treat their personnel as key (in some cases critical) information assets and should make risk decisions based on their availability.

7.3.2.1 Separation of Duties

Separation of duties can be both good and bad. From a security perspective, it presents fewer risks because not all IT personnel have total access to and control of all networked systems. From an availability perspective, however, problems can arise, especially in smaller organizations with relatively few IT personnel. IT managers need to ensure that they promote cross-utilization and training so that SPOF in personnel knowledge and experience are minimized or eliminated. It is equally important to have documented procedures for administration, troubleshooting, and maintenance of networked systems. If IT managers follow these practices, they will likely be better prepared if one of their key personnel is on vacation or otherwise unavailable during a system failures or incident.

We've talked previously about bandwidth and dedicated links to destination networks. It's very important to consider the relationship an organization has with its Internet service provider (ISP) or commodity bandwidth provider (CBP), as we will discuss next.

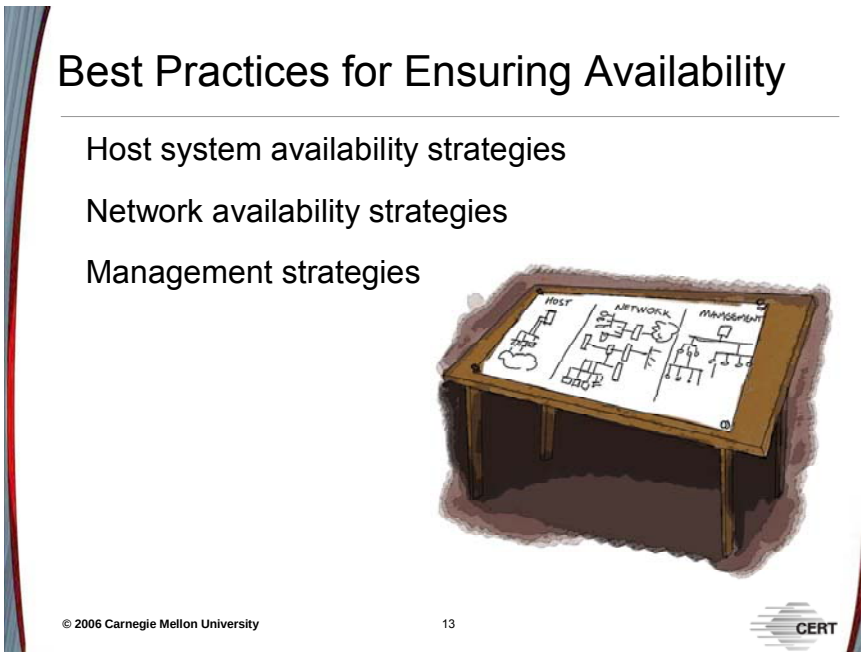


7.3.2.2 ISPs and Commodity Bandwidth Providers

If an organization's mission is reliant upon its Internet link(s) to the degree that outages equate to lost revenue (via e-commerce), reputation, or market share, it should consider fault-tolerant or redundant solutions. One survivability best practice involves having more than one Commodity Bandwidth Provider providing Internet links (i.e., separate T-1 lines) and ensuring that those links are separate physical lines, not just multiplexed circuits on the same medium. Another best practice involves establishing service-level agreements (SLAs) so that providers will commit to certain availability parameters (i.e., the Committed Information Rate in Frame Relay). IT managers need to have some level of confidence that, for example, their T-1 leased line will be up and available at the full 1.54 Mbps speed 99.9% of the time. If the ISP is also providing DNS and DHCP services to an organization, it should be doing so in a fault-tolerant or at least redundant manner, and this should be stipulated in the SLA.

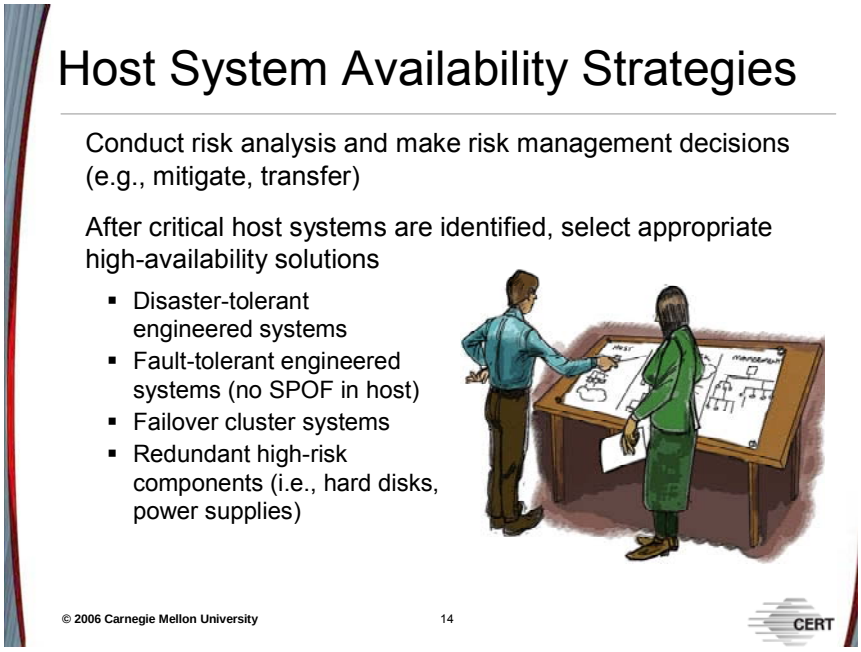
7.3.2.3 Outsourced Services

Many organizations (including the United States Navy and Marine Corps) have determined that it is more cost effective to outsource some or all of their IT services. The notion of residual risk was discussed earlier in this course in terms of organizations outsourcing services and the associated risks (to availability, etc.). However, the responsibility (and moreover, the consequences) for those risks to critical services is still retained by the organization. Specific expectations regarding levels and quality of service should be part of contracts and SLAs that outsourced service providers agree to. Examples would be availability of support personnel, whether on-site service and maintenance will be covered 24x7x365, and so on. If you are an IT manager with dependent outsourced services, do you know that your specific service expectations will be met?



7.4 Best Practices for Ensuring Availability

Now that we've established that risks to IT service availability are significant, how can we effectively manage these risks? The three general strategies listed on this slide and discussed in the following pages are not meant to be comprehensive; however, they offer a good starting point for investigating effective means of managing risks to availability.



7.4.1 Host System Availability Strategies

The crux of survivability is effective risk management. By following the risk management guidelines discussed earlier in this course, IT managers can determine the relative importance and criticality of their key host and networked systems to the mission of the organization. In a related vein, we've in this module on ways to mitigate risks to availability, including disaster/fault tolerance, redundancy, and failover. So, what are some specific technical implementations that can provide adequate levels of availability? Let's use an example to answer this question:

> Clyde's Military Surplus, Inc. has grown in 10 years from 3 area stores to 15 stores across the region. What's more, an e-commerce initiative was adopted early on, and Web sales now total more than 10 percent of all revenue. The company manages its Web infrastructure internally and recently had a system failure on its lone Web server (IIS 5 on Windows 2000) that caused 1.5 days of downtime. The CEO (Clyde Clemons) has told Tim, the IT manager, to "fix the problem—but these are hard times, so be prudent with expenditures."
>
> With this as his guideline, Tim evaluates his options.
>
> He defines the Web server as a SPOF and, after conducting risk analysis and availability calculations, determines that the server must have five 9s of availability (99.999% uptime). However, he doesn't have the personnel and budget resources to re-architect the software portion of the e-commerce solution at this time. He investigates disaster-tolerant solutions but decides that the organization's relatively low risk of a building-level disaster does not currently warrant this expense. However, Tim foresees a time in the future when this capability could provide

enough return on investment to justify the expense. He would like to consider a fault-tolerant solution that can be scaled to provide disaster tolerance without requiring re-engineering of the entire system.

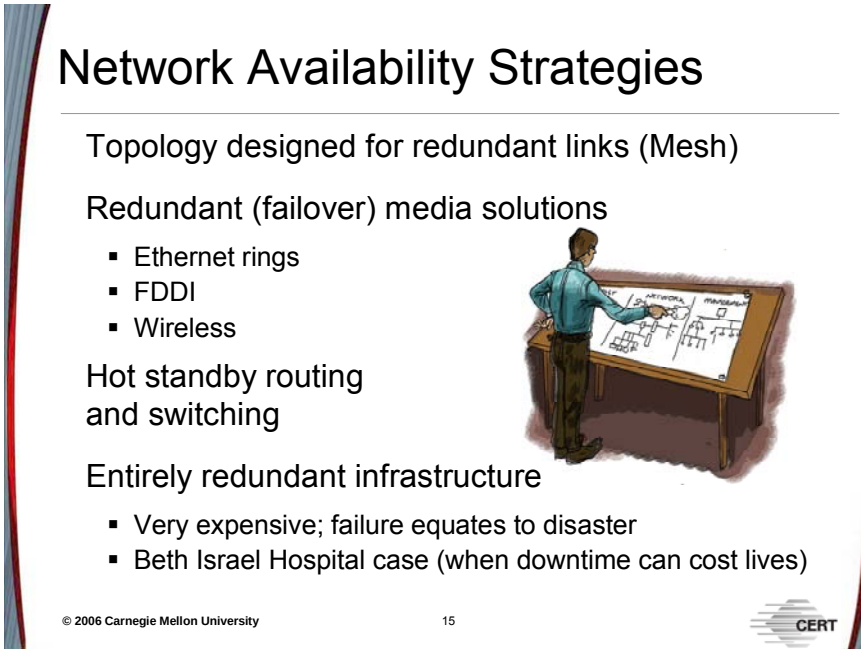
Tim next investigates various Windows 2000 cluster solutions. However, clusters are redundant servers that generally provide only fault-resilient (failover) availability services and therefore don't meet his availability requirements. And even though hardware failures partially caused the 1.5-day outage, Tim quickly dismisses the idea of merely increasing the redundancy of the server's hardware components because a solution engineered to be fault tolerant by design couldn't have any SPOF.

As stated, Tim cannot currently afford to change his Web server solution or the underlying Windows 2000 platform. If possible, he wants to avoid proprietary hardware server platforms because they tend to have higher total costs of ownership. He does market research and determines that a solution is available that meets his requirements.

Tim selects a fault-tolerant solution from Marathon Technologies[30] called the Marathon Assured Availability server [Marathon 06]. It provides Clyde's Military Surplus with the necessary five 9s of availability without requiring reengineering of the Windows 2000/IIS e-commerce solution. It can be scaled up to a fully disaster-tolerant system called SplitSite that can isolate each half of the redundant server at a distance of up to 55 kilometers apart. This solution will not incur reengineering expenses and runs on standard, vendor-provided, Intel-based servers (i.e., Dell, Compaq, IBM, and Hewlett-Packard). Because the system is designed to be fault tolerant, there are no SPOF within it and no failover downtime to be concerned with. Additionally, since this solution uses standard industry servers and shrink-wrapped Windows operating systems and applications, the time and expense required to implement the solution into Tim's infrastructure is minimal.

This is only one example and attempts to illustrate the steps and considerations involved in attempting to increase the availability of host systems. There are other solutions available for Linux and other UNIX-based platforms as well as proprietary systems (like IBM's AS400 series servers) that provide similar levels of availability.

[30] See http://www.marathontechnologies.com/.

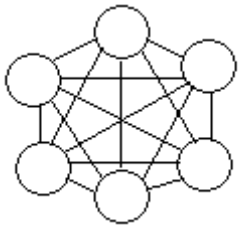


7.4.2 Network Availability Strategies

As seen in our discussion of single points of failure, simply making data and host systems fault tolerant may not always provide the expected levels of availability. SPOF in the network and other service dependencies affecting the fault-tolerant host can decrease availability levels.

7.4.2.1 Topology Designed for Redundant Links (Mesh Topology)

When the cost of network failure is too high to be tolerated, strategies exist that can increase network availability to avoid SPOF. For example, the network can be designed with what is termed a *partial* or *full mesh topology*. This means that all core Internet work systems (including routers, switches, hubs, and some critical servers) have redundant links to each other. In a full mesh topology, every node has a connection to every other node in the network [Webopedia 06].



Full mesh is very expensive to implement but yields the greatest redundancy, because if one node should fail, network traffic can be directed to any of the other nodes. Full mesh is usually reserved for backbone networks. Partial mesh topology is less expensive to implement and yields less redundancy than full mesh topology. With partial mesh, some nodes are organized in a full mesh scheme, but others are only connected to one or two other nodes in the network. Partial mesh topology is commonly found in peripheral networks connected to a full meshed backbone.[31]

[31] See http://www.webopedia.com/TERM/M/mesh.html.

Other topologies like Ethernet rings and Fiber Distributed Data Interface (FDDI) can provide redundancy in network connectivity using the same physical medium. In some cases, it is wise to build network redundancy by utilizing multiple media. For example, having both a wired and wireless capability within the same network can potentially increase availability if partial network failures occur.

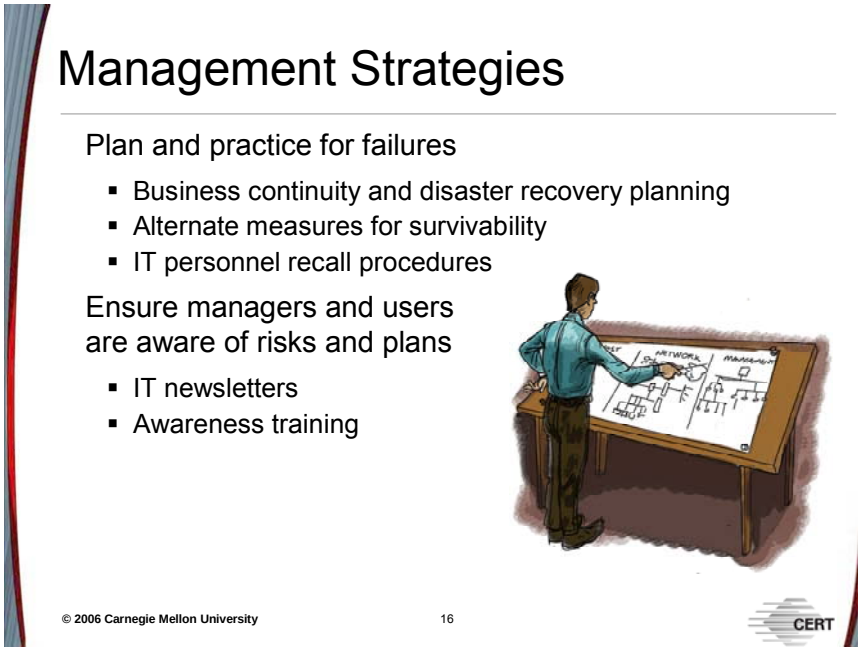
7.4.2.2 Hot Standby Routing and Switching

"Hot standby" routing and switching technologies allow for automatic failover and recovery if an individual router or switch fails on a network. Some modular switches and routers (e.g., a Cisco Catalyst 5500 multi-layer switch) are designed to contain redundant switch or router modules (blades inserted into slots in the switches' chassis) that implement hot standby protocols, thereby providing redundant routing and switching all within a single black box.[32]

7.4.2.3 Entirely Redundant Infrastructure

As mentioned previously, some networks are so mission critical that a completely separate and redundant standby network infrastructure is required. One real-world example of this situation involves Beth Israel Deaconess Medical Center in Boston. The hospital experienced a multi-day network outage that severely impacted hospital operations and even put patients' lives at risk. The hospital has since invested in a completely separate infrastructure and fault-resilient solution for maintaining continuity in the event of another major network outage [Caffrey 04].

[32] See http://www.cisco.com/en/US/products/hw/switches/ps686/index.html.



7.4.3 Management Strategies

Technology is only part of the solution when it comes to increasing the availability of IT operations. Effective management practices are equally essential.

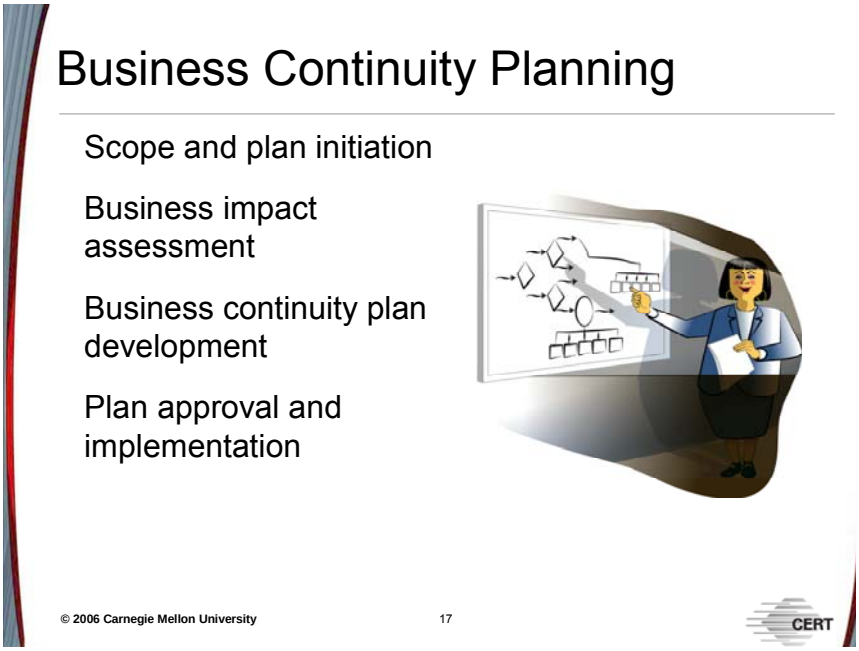
Systematic planning, including business continuity and disaster recovery planning, is possibly the most important practice of all for ensuring availability; unfortunately, systematic planning is also frequently dismissed by IT managers whose time is already filled by operations and maintenance issues. However, without systematic planning, an organization is unlikely to achieve true Defense-in-Depth.

Alternate measures should be considered and planned for in situations where risk has been accepted. For example, if the email server fails and must be rebuilt (within the organization's risk-tolerance level), are there procedures in place to failover to paper-based and voice-based operations?

It is advisable to practice data restores and system failures and recoveries. It is also a best practice to create detailed procedure documents that instruct IT personnel on how to respond in emergency failure situations.

All of this planning and practice will do little good if the organization as a whole is unaware of contingency plans and procedures. It is a good idea to keep personnel informed through monthly newsletters and/or awareness training sessions. These sessions also can help build confidence and rapport between the IT department and the rest of the organization's personnel.

Lastly, it is important to test contingency plans before they are actually needed. In this way, any shortfalls or holes in the plans can be identified and corrected in a non-emergency situation.



7.5 Business Continuity Planning

Although business continuity is a wide-ranging topic that could merit its own training class, we will provide a brief overview here. In essence, business continuity plans are designed to avoid or minimize interruptions to normal operations. An *interruption* is defined as a failure that could result in the loss of capital due to the inability of the organization to operate as it usually does. Interruptions may range from natural disasters to computer break-ins. Regardless of the type of interruption, business continuity plans should aim to minimize the disruptive effect and allow for prompt return to normal business operations.

When determining what types of interruptions it may face, an organization should consider the following critical information processing areas:

- local and wide area networks
- telecommunications and data communications links
- workstations and workspaces
- applications, software, and data
- media and records storage
- staff duties and production processes

There are four elements involved in business continuity planning:

1. scope and plan initiation
2. business impact assessment (BIA)
3. business continuity plan development
4. plan approval and implementation

We will look at these four elements in more detail.

7.5.1 Scope and Plan Initiation

Business continuity planning begins with scope and plan initiation. Participants in this phase should include personnel from various departments within the organization, especially those performing critical duties. They should define the scope of the plan, detail the organization's operations, and identify the resources needed to carry out the plan. The plan should address how to recover from a disruptive event and should identify specific mitigation strategies. After defining and creating the plan, participants should then implement and test it.

The plan cannot be successful without senior management's support and involvement. Senior management should monitor not only development of the plan, but also its execution in the event of a disruption. It is in management's best interest to be involved in this process. After all, if an organization should suffer financial losses as a result of a disruptive event, stockholders might hold management personally responsible if management did not demonstrate due care in development of the business continuity plan.

7.5.2 Business Impact Assessment (BIA)

The business impact assessment (BIA) process involves creating formal documentation describing the impact various disruptions would have on the organization. The details of this documentation include potential financial or quantitative loss, potential operational or qualitative loss, and vulnerability assessment.

There are three primary goals of the business impact assessment:

1. Criticality prioritization – identify and prioritize every critical business/operations unit process, and evaluate the impact of a disruptive event.
2. Downtime estimation – estimate the maximum tolerable downtime that the business/operation can tolerate while still remaining viable (i.e., what is the longest period of time a critical process can remain interrupted before the organization can never recover?).
3. Resource requirements – identify resource requirements for critical processes, allocating the most resources to time-sensitive processes. [Krutz 01]

The business impact assessment has four steps:

1. gathering the needed assessment materials
2. performing the Network vulnerability assessment
3. analyzing the information compiled
4. documenting the results and presenting recommendations

Gathering assessment materials, the first step of the BIA, can be as trivial as designing an organizational chart to show the organization's departments and their relationships. This step enables the organization to identify operational dependencies and determine priorities.

The vulnerability assessment is conducted on a smaller scale than a regular risk assessment, concentrating only on continuity and disaster recovery planning. The primary function of this assessment is loss impact analysis covering both quantitative and qualitative losses. Quantitative losses include loss of revenue, increased capital expenditures, increased operating expenses, and losses from contract or regulatory requirements. Qualitative losses include loss of competitive edge or market share and loss of credibility among the public.

Critical support areas are identified in the vulnerability assessment. A critical support area is defined as a business unit or function that must be present to sustain continuity of the business processes, maintain life safety, or avoid public relations embarrassment. Some examples of these areas include telecommunications, data communications, physical infrastructure, accounting, payroll, and transaction processing. Elements that support these areas (i.e., personnel and resources) must be identified as well.

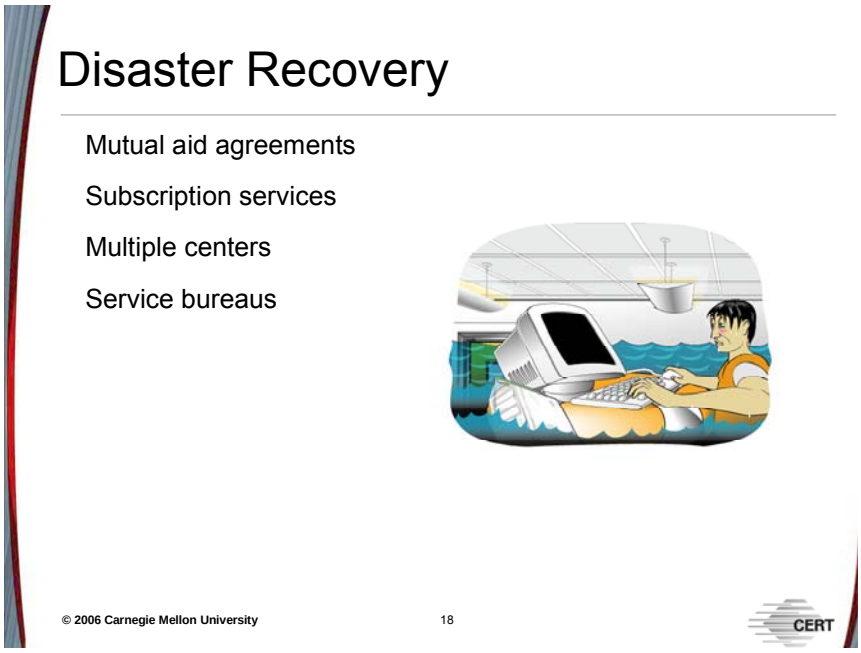
Finally, all of this information must be organized into a formal report and presented to senior management with recommended recovery priorities based on the results of the analysis.

7.5.3 Business Continuity Plan Development

In business continuity plan development, the information collected in the BIA is used to create a recovery strategy plan to support these critical business functions [Krutz 01]. In this step, we begin to outline a strategy for developing such a plan. First, the continuity strategy will be defined. In this step, all organizational elements such as computing, facilities, personnel, and supplies and equipment must be included, and their role in the continuity strategy noted. Results of this strategy definition phase should then be documented in the complete recovery strategy plan.

7.5.4 Plan Approval and Implementation

Once the recovery plan is documented, the final step is to implement the overall business continuity plan with management support and approval. Management must make the entire organization aware of the plan and routinely review and update it to keep it current and usable.



7.6 Disaster Recovery

Disaster recovery is the process of managing the impact of a realized risk. A disaster recovery *plan* can serve as a preventive measure—a strategy for keeping your computer equipment and information assets available to legitimate users in case of an emergency. Having a disaster recovery plan in place may spell the difference between a problem and a catastrophe [Russell 91]. The plan should detail actions for personnel to take during and after disruptive events that may affect the organization's information assets. These actions include responses to the event such as providing backup operations and managing the recovery process.

A disaster recovery plan also should include instructions for protecting the organization from computer services failures, minimizing delays in operations, executing operations on standby systems, and minimizing confusion during the disaster.

Having an alternate means of continuing operations is essential in the event of a disaster. The details of this alternate processing arrangement are important elements of the disaster recovery plan. Some of the most common types of alternate operations arrangements include

- mutual aid agreements
- subscription services
- multiple centers
- service bureaus
- other data center backup alternatives

7.6.1 Mutual Aid Agreements

A mutual aid or reciprocal agreement is an arrangement with an organization that has similar daily operations and environment. The two organizations agree to support each other by sharing facilities in the event of a disaster. The advantage of this type of arrangement is obvious—the costs to continue operating during a disaster are minimal. However, there are also disadvantages associated with this type of agreement. It is unlikely that each organization will have enough unused resources available for the partner organization in the event of a disaster. Also, if the disaster is large enough, it may affect both companies so that neither would have an alternate site to continue operations. This plan clearly has limitations.

7.6.2 Subscription Services

Subscription services are probably the most common alternate-site solution. In this approach, a third-party service provides the organization with backup facilities so that it can continue operations even if its own facilities are unavailable. There are three types of subscription services: hot site, warm site, and cold site.

7.6.2.1 Hot Site

A hot site is a fully equipped alternate site, including functioning printers, servers, and workstations. The idea is that in the event of a disaster, personnel can report to this site and resume operations almost immediately. There are many advantages to a hot site. For starters, the site is available 24/7 and is exclusive to the organization. No one else will be using it. Should a disaster occur, the site will be ready immediately and can be utilized for short-term or long-term periods of time.

However, one of the disadvantages of a hot site is cost. Full redundancy of all services can be expensive, and a hot site requires constant maintenance and upkeep. The cost of maintenance and operations of all hardware and software (including applying the same patches, backups, upgrades, etc. that are applied at the primary site) adds to overhead costs and can be a strain on organizational resources. Another costly item is security. Since a hot site is essentially a mirrored site with identical data, it requires the same security controls (including physical security) that are used at the primary location.

7.6.2.2 Cold Site

A cold site is the most common type of alternate site. It is simply a room with power ready for use in the event of a disaster, but it contains no servers, workstations, or other equipment. In the event of a disaster, it will be necessary to bring in equipment and set it up. The main advantage of a cold site is its low cost; the main disadvantage is that it takes a long time to get the site up and running so that the organization can commence operations. This disadvantage may mean that a cold site is an inadequate resource for disaster recovery at some organizations. In addition, there is no way to determine whether a cold site will suffice in the event of a disaster, as there is no way to test it until disaster strikes.

7.6.2.3 Warm Site

A warm site is a cross between a hot site and a cold site. Like a hot site, it is an equipped off-site facility ready for use in the event of a disaster; like a cold site, it takes a while to bring it online. A warm site does not have full redundancy of services. There are power and communication links and some servers, but usually only a few workstations—most workstations must be brought in and loaded with data from the primary site. This option is less expensive than a hot site since there are minimal overhead and maintenance costs, but it may not be viable for organizations that need to ensure continuity of critical operations with minimal downtime.

7.6.3 Multiple Centers

Multiple centers, or dual sites, represent another alternate-site approach in which organizations distribute their operations across several locations. These locations may be owned by the same organization or may be part of a reciprocal agreement. The advantage of this type of site is low cost and the opportunity for resource and support exchange across locations. Disadvantages include the risk that a disaster could take out more than one location, as well as the potential difficulty of managing multiple configurations, people, and processes at different locations. Coordination also may be difficult, in terms of determining which locations should bring which services back online first.

7.6.4 Service Bureaus

Organizations rarely contract with service bureaus for alternate backup services. Although this option features quick response and testing ability, it is extremely expensive.

7.6.5 Disaster Recovery Plans

7.6.5.1 Updating the Disaster Recovery Plan

Disaster recovery plans may quickly become obsolete due to reorganization or changes in the computing infrastructure. The plan should be reviewed, audited, and updated often. Updated versions should be distributed throughout the organization, and any previous versions should be discarded.

7.6.5.2 Testing the Disaster Recovery Plan

It’s important to develop a testing program for any disaster recovery plan. The testing program should include full backup from tapes and be performed on a regular basis. The main purpose of these tests is to (1) verify the actions detailed in the plan and (2) identify failures that should be corrected. The tests will also help personnel become familiar with and train for their assigned duties in the event of a disaster.

To get the most out of testing, an organization should develop a document that outlines the reason for the test, sample scenarios, functions to be tested, and the objective or desired outcome of the test. The document should include the test timing, extent of the test, specific steps to be taken, participants and their assigned tasks, and resources or services required. Testing should not disrupt normal business operations.

The table below shows five types of disaster recovery plan tests. They are listed in order of increasing detail and complexity. The organization should start with the checklist approach and progress through the table, with the eventual goal of completing a full-interruption test.

Level	Type	Description
1	Checklist	Copies of plan are distributed to management for review.
2	Structured walk-through	Business unit management meets to review the plan.
3	Simulation	All support personnel meet in a practice execution session.
4	Parallel test	Critical systems are run at an alternate site.
5	Full-interruption test	Normal production is shut down; real disaster recovery processes are used.

The procedures in the disaster recovery plan describe what tasks must be done to recover, what roles specific personnel must play, and how the organization must communicate to external groups in the event of disaster. The primary elements of the disaster recovery process can be classified as follows [Krutz 01]:

- recovery
- salvage
- resumption of normal operations
- other recovery processes

7.6.5.3 Recovery

The disaster recovery plan should define a recovery team. At the declaration of a disaster, this predefined recovery team will implement the recovery procedures. This team's primary duty is to get critical business functions up and running at the backup site. The team will need to bring the materials needed (e.g., tapes, workstations) to be operational to the backup site and perform any necessary installations so that critical operations may resume.

7.6.5.4 Salvage

A salvage team, defined in advance by the disaster recovery plan, must return to the primary site as soon as the disaster has ended and there is no risk of personal danger. Its task is to

determine the viability of the primary site and try to salvage or repair equipment. This team supervises the building cleanup, which includes such activities as water removal. This team also declares the site resumptive or not after examination and testing.

7.6.5.5 Resumption of Normal Operations

Once it has been determined that operations can move back to the primary site, procedures must be in place to ensure minimal disruption or risk. The least critical business functions should be brought back to the primary site first.

This entire process requires well-coordinated plans and resources. The disaster is not over until all operations have been returned to their normal location and function. A large window of vulnerability exists when processing returns from the alternate backup site to the original production site. When all areas of the enterprise are back to normal in their original location, with tests certifying all data as accurate and all operations as functional, you can officially declare the disaster at an end.

7.6.5.6 Other Recovery Considerations

Other considerations during a disaster that should be addressed by the disaster recovery plan include

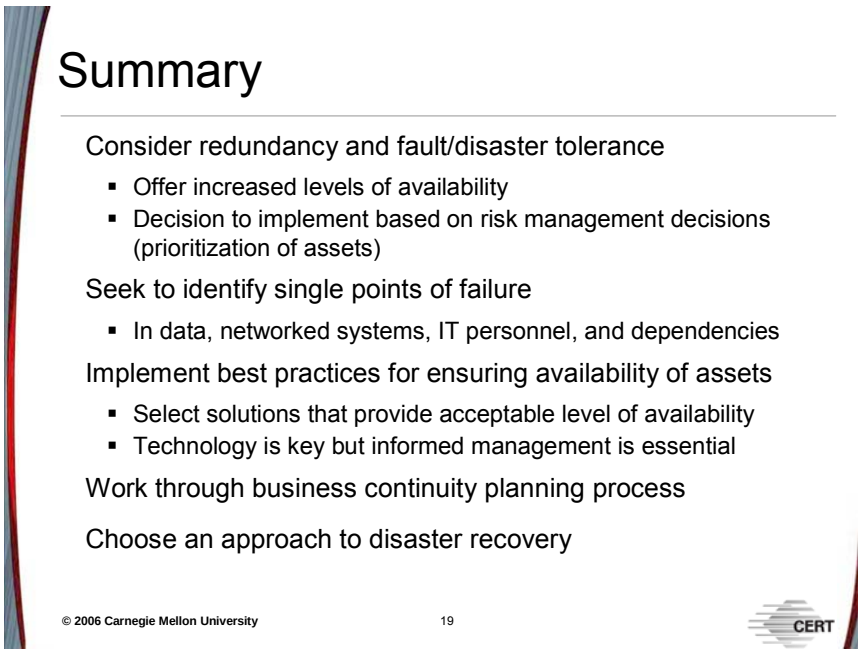
- interfacing with external groups – The recovery plan should discuss how to communicate with external groups during a disaster. These groups include municipal emergency departments, civic officials, utility companies, customers, the media, and shareholders. These relationships should not be neglected.
- employee relations – The disaster recovery plan should include management of employees and families. During an event that may involve major safety dangers, the organization should be prepared to continue paying salaries even if business production has ceased, since its employees are critical assets and paying them is therefore a critical business process. The organization's insurance should be able to continue paying the salaries for an extended period of time. In major disasters, relocation or other living expenses may need to be provided to employees as well.
- fraud and crime – Competitive organizations or other external parties may look to benefit from a disastrous event by exploiting security vulnerabilities and opportunities for fraud. Other possibilities for fraud and crime include vandalism and looting. When resuming operations after a disaster, IT personnel should stay alert and monitor host and network activity for any signs that systems may have been compromised or exploited.
- financial disbursement – It is possible that expenses incurred during a disastrous event will surpass the event manager's authority, so disaster recovery plans should address expense disbursement. For example, signed, authorized checks should be available for financial reimbursement.

- media relations – As mentioned earlier, the plan must explain how to handle the media. A credible, informed spokesperson should be appointed to address the media so that the media doesn't seek other sources and potentially report false information. Failing to make oneself available for the media may lead to rumors of a cover-up. The organization should always report its own bad news quickly and honestly to avoid skepticism and rumors.

7.6.5.7 Case Study: Beth Israel Deaconess Medical Center

In November 2002, the Beth Israel Deaconess Medical Center in Boston could have used a good disaster recovery plan that included alternate computing resources. Its network got caught in an endless loop until it eventually came to a halt. The disaster lasted four days and forced the hospital to revert to an old paper-based system that required hundreds of thousands of sheets of paper to be hand-delivered across the campus. Many residents had to be taught how to write orders and fill out flow sheets.

The hospital kept backups of data, but in this case it was the network that became clogged and crashed. Many experts had to be flown in to help remedy the situation.



Summary

Availability of IT operations is of great concern in terms of the survivability of an organization. IT managers therefore should investigate technological and management solutions to manage and ensure availability of their organizations' critical information assets, following a thorough risk assessment that prioritizes those assets.

Review Questions

1. How is availability calculated using the MTBF and MTTR?
2. Describe three ways to mitigate single points of failure (SPOF).
3. What is the difference between partial and full mesh topology?
4. Name the four elements of business continuity planning.
5. What are three types of mutual aid agreements, and how do they differ?

Module 8: Configuration Management

This module introduces the process for configuration management, discusses its role and the benefits it offers to an organization, and provides best practices to follow for achieving it.

Instructional Objectives

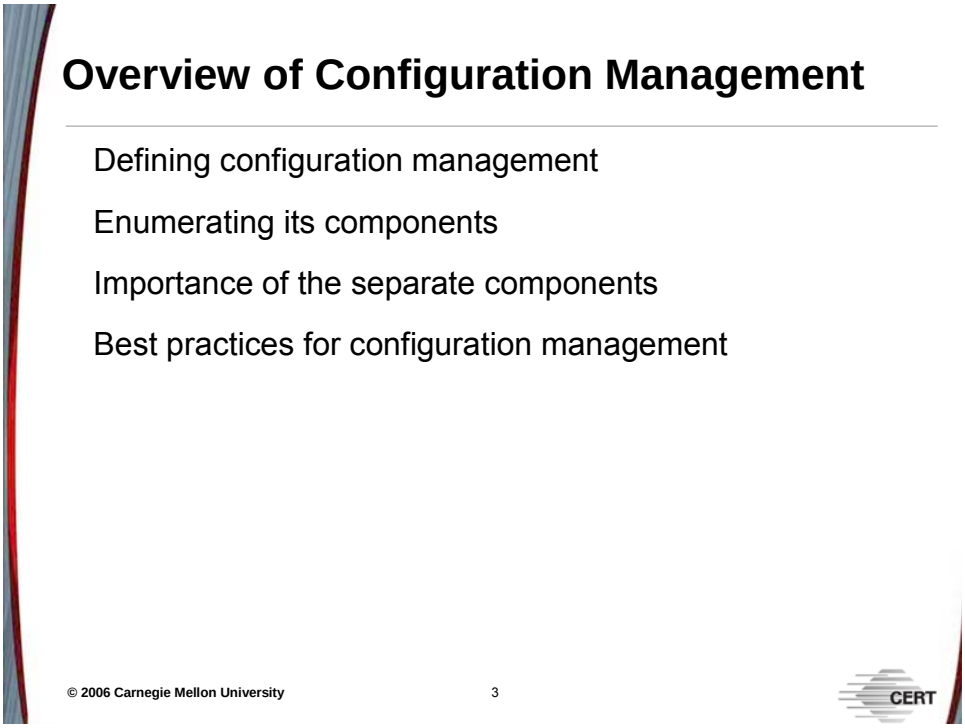
After finishing this module, students will be able to

- Define configuration management
- Know the benefits of configuration management
- Describe the individual components of configuration management
- Characterize the best practices for configuration management

CERT

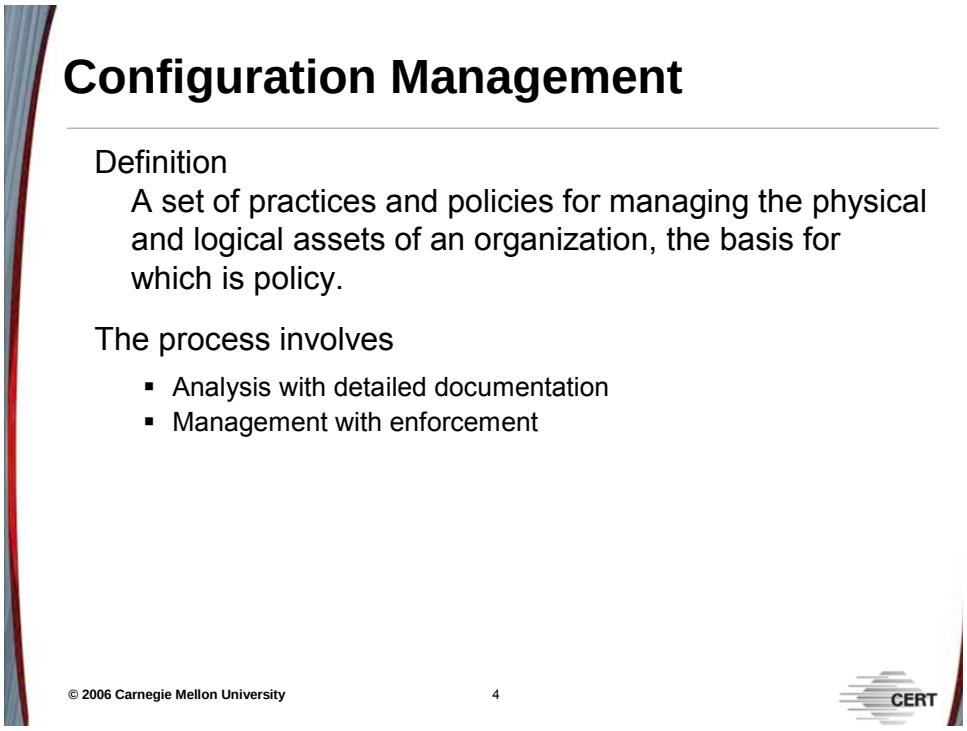
This instructional module will enable students to complete all of the above learning objectives.

8 Overview of Configuration Management



This module attempts to lay a foundation covering

- the holistic concept of configuration management
- the role and benefits of configuration management in an organization
- best practices for configuration management as a component of Defense-in-Depth



8.1 Defining Configuration Management

Configuration management is a set of processes for managing the enterprise architecture of an organization. Indeed, it can be seen as a tool for implementing and managing the other components of the Defense-in-Depth strategy, which are discussed in other modules of this course. This does not mean it is unimportant; rather, configuration management is essential to achieving security and accountability in any organization. After all, organizations often manage many different systems and devices that interact with each other as a single network. Configuration management can help IT administrators make sense of this interaction process.

Configuration management begins with organizational policy. This policy should state the goal of the configuration management effort – what is to be achieved. After a network's initial configuration is developed, implementers of configuration management then can merely put policy into practice.

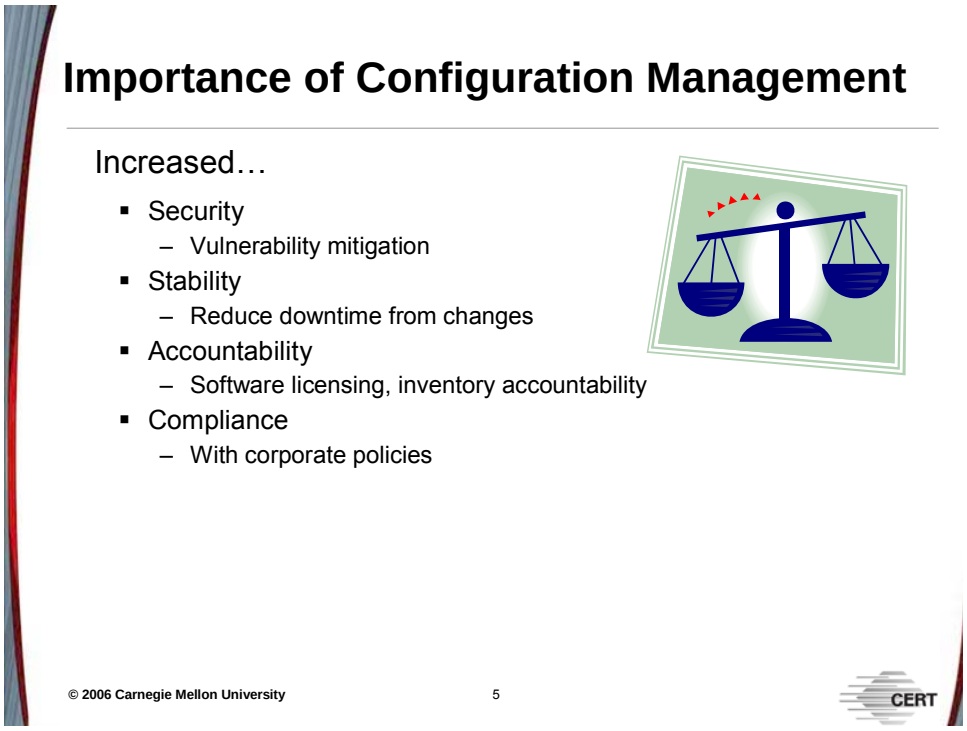
The most important component of each step of configuration management is documentation. Documentation must accompany every step of the configuration process, from the definition of the actual management process to be implemented, to the planning and evaluation stages of individual configuration events, through the review and monitoring of necessary changes.

8.1.1 Why Configuration Management?

While it has always been good practice to have policies and procedures for managing information and network security, this is now becoming an integral part of being in business. With the increase in state and federal regulations specifying minimal standards for

information assurance, it has become even more important to understand the role that technology plays in maintaining compliance.

Sarbanes-Oxley, HIPAA, GLBA, and FISMA are just a few of the new regulations being imposed on business. Further discussion of regulation and policy is available in the “Compliance Management” module of this course.



8.2 The Importance of Configuration Management

8.2.1 Why Is It important?

Through proper configuration management, an organization can strengthen the properties listed below.

Security
Increased security can be achieved through control of assets, including location, installed software, and the individual configurations of machines and devices. For example, if an organization knows precisely what is supposed to be installed on its machines, it is likely to notice unauthorized software much more quickly. Internal assessments also will enable an organization to analyze and test its security measures prior to a security incident.

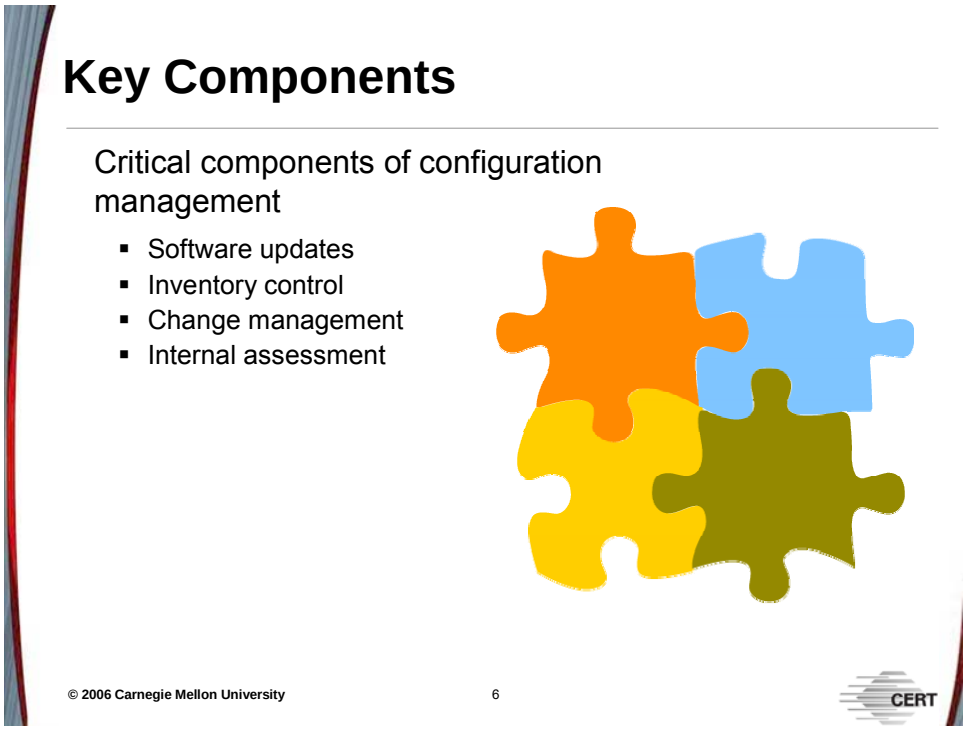
Stability
By developing and following specific change management procedures, an organization can reduce the downtime resulting from updates and changes to system components. Performing adequate testing and analysis prior to deploying changes should result in increased stability. For example, if the change management process calls for a patch to be rolled out and observed on a test network before it is deployed in a production network, unexpected consequences such as system instability can be minimized.

Accountability
A natural result of procedural control is increased awareness and accountability. Through configuration management, an organization can achieve greater accountability regarding

the physical and logical status of assets. This applies to both hardware and software. For example, by tracking who makes changes to its systems, an organization can trace any problematic changes to their source and then use that person's knowledge to achieve faster resolution.

Compliance

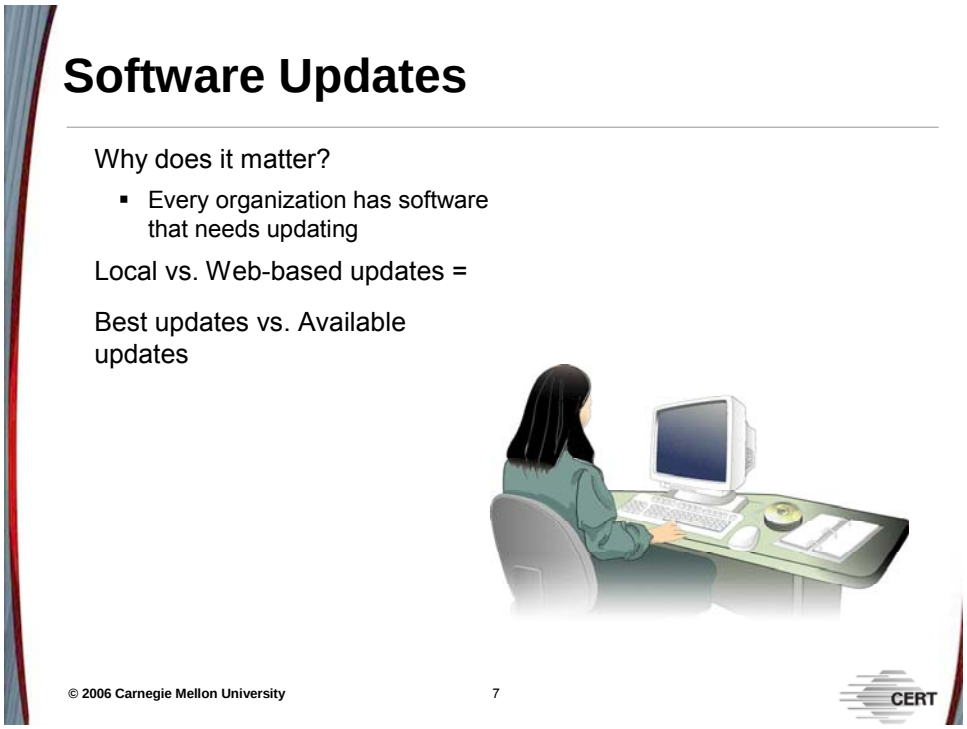
Organizational policy is a prerequisite for information assurance and should be the foundation for managing a network. Increased control over systems and network components increases an organization's ability to apply the principles of configuration management and comply with corporate security policies. Configuration management provides an excellent mechanism for enforcing compliance in a networked environment. For example, if policy specifies what a particular system must do and how it must be configured, configuration management can ensure reality matches intent.



8.2.2 The Components of Configuration Management

Configuration management has four key components. When combined, these can significantly improve the security posture and preparedness of an organization. The key components of configuration management are

1. software updates
2. inventory control
3. change management
4. internal assessment



8.2.3 Update Management

Updating applications is a necessary activity in every organization. Every day, new bugs and vulnerabilities are found in software. To mitigate risk, administrators find themselves continuously trying to keep up with evolving software environments. So, if there is an application running on a machine, there is a good chance that during its lifespan it will require updates. How an organization chooses to handle updates and patches will determine the ease and effectiveness of this activity.

Methods for managing the update process should be carefully considered. Many organizations address update management with a laissez-faire attitude. This approach leads to trouble, such as updates on one Web server but not on another, and should be avoided. Updating is not a one-time operation, but rather a process that must be repeated with regularity. Thus it should be viewed as any other business-critical process.

Anyone who has updated more than one or two systems at a time knows it can be a convoluted process. This section is geared toward explaining the “why and how” of effectively managing this process.

8.2.3.1 Defining the Process

It is important to regard update management just as you would regard any other business process. As a component of your organization’s Defense-in-Depth strategy, it should rely on set procedures and methods to ensure it is handled in an efficient and effective manner. With this in mind, each business should develop a set policy for update management that can be

implemented by members of the technical staff. The following activities are key components of the update process.

Assess

This initial phase determines what assets you are responsible for that will require updating. Decisions made at this stage will enable you to make further decisions about measures your organization can implement to manage the update process.

Considerations during the assessment stage include

- What operating systems will be managed?
- What devices will be managed?
- Where are the devices located?
- How do they connect?

Identify

During the identification stage, you will want to determine what patches and updates are available for the products you identified in the assessment stage. This stage also involves some analysis to determine which updates are relevant to your organization's environment.

For example, if there is an update for IIS Server on Microsoft Windows 2003 Server, but your organization does not use IIS to host Web content, you may choose not to apply this update.

Evaluate and Test

During the evaluation phase, you will make final decisions about which updates to apply. However, these decisions should be made only after testing the updates in a simulated production environment to ensure they do not adversely affect important business applications. You should also consider whether updates can be removed after they have been installed.

Plan

After testing, you should plan for deployment of the updates. This plan can vary depending on the tools to be used for deployment, but should involve careful consideration of the business environment and needs of the organization. In some cases it may be easiest to deploy a patch to all hosts, while in other situations you may want to deploy in stages. For example, you might deploy to a specific department or group or deploy during planned downtime. A primary factor should always be the importance of the update and the likelihood that *not* deploying quickly could lead to exploitation of a vulnerability.

Implement

The implementation phase is the actual updating of hosts in your network. If the previous phases are complete, this phase should go relatively smoothly. However, as with all major projects, there may be problems along the way. You cannot always predict and prepare for every scenario you might encounter. This is why you will need to follow up your deployment with a review.

Review

This is an assessment of the deployment. You should be able to determine if all hosts were successfully updated or if there were problems. In addition, this stage may involve assessing the user experience with important applications. Even when tests are performed prior to implementation, there is always a possibility of unforeseen side effects.

8.2.3.2 The Options

You can either make your update decisions locally or have a software company make the decisions for you. There are attractive qualities to each of these options, depending on the size of your organization. We will see, however, that choosing to make decisions locally provides much more control over infrastructure.

8.2.3.3 Web-Based Updates

This is a very common means of updating systems. If you have ever configured or even used a computer running on a Microsoft Windows operating system, you are probably familiar with how Web-based updates work. When you take a machine out of the box, there are probably a few settings to choose from, but by default your system will 1) attempt to make contact with a Web-based update server hosted by the software vendor, and 2) download any updates applicable to the software running on the machine. This is the *available updates* model. You are exercising very limited control over the updates being installed on a machine.

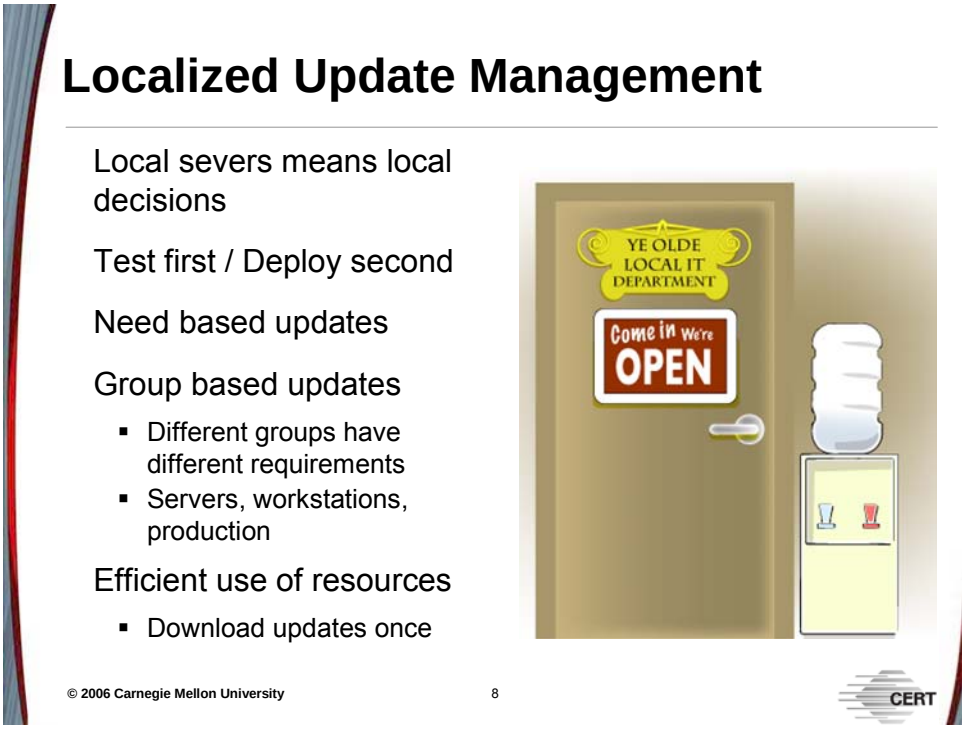
Why Is This Bad?

When using the available updates model, you are choosing to download updates that are applicable to the specific software running on a host. This model does not take into consideration any other software installed on the host. If there are business-critical applications installed on a host, they may not be compatible with the blanket updates provided by the vendor of an operating system or other applications.

Even though this is the easiest model to choose, it is not the easiest model to manage. In fact, this model can cause headaches for a system administrator, especially when it comes time to be accountable for the status of machines.

The difficulties associated with this choice include

- limited control over which updates are installed
- uncertainty about the following issues
    - Have all updates successfully installed?
    - Have users interfered with updates?
    - Have users declined to install some updates?
    - Have users changed the update settings?
    - Do the updates cause problems with critical applications?
    - Can the updates be uninstalled?



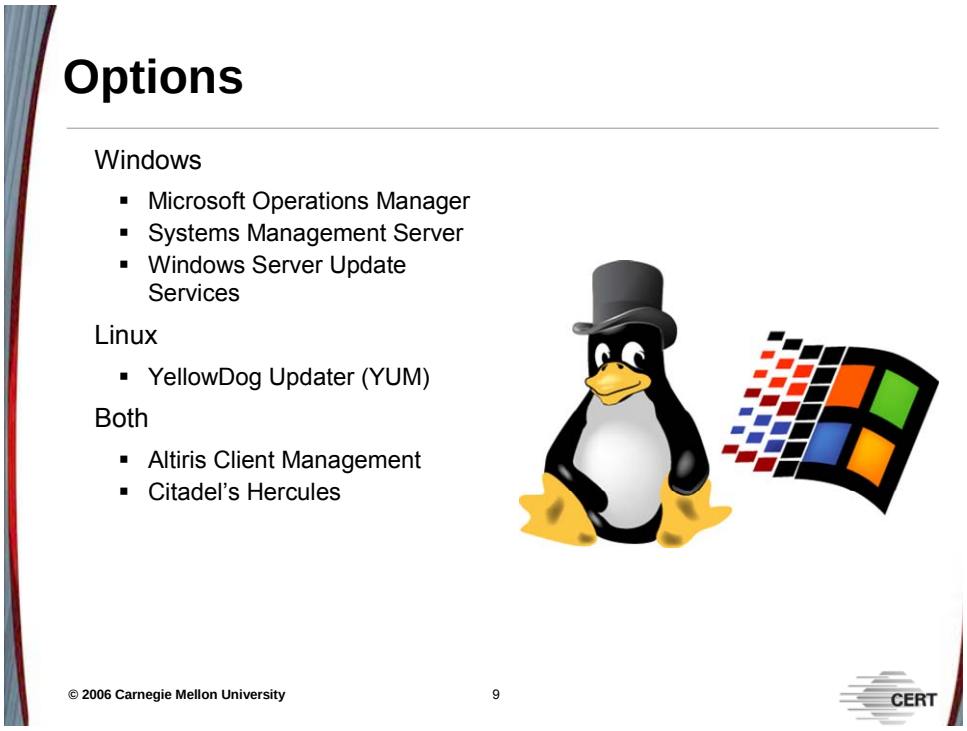
8.2.4 Local Updates Server

There are many ways to implement the local updates model. Most solutions that are available offer a variety of options for managing updates.

Since updates often pertain to the operating system, there are several solutions offered by Microsoft, for example Windows Server Update Services (WSUS). Some of these solutions are applicable to the operating system only, while others can manage updates and deployment of third-party applications. There are also third-party solutions to managing updates across an enterprise.

One application that is available from a third party is the Altiris Client Management Suite, which enables patch management across an organization with heterogeneous operating systems and varied applications.

In addition, Citadel software sells a product line called Hercules that was developed to provide enterprise-wide security and vulnerability management. With an application of this type, you can address a multitude of security and configuration management tasks. The Hercules line includes products to inventory and audit hardware and software resources. Based on the audit results, you can then identify necessary patches and updates to be applied either within defined groups or across the entire network.

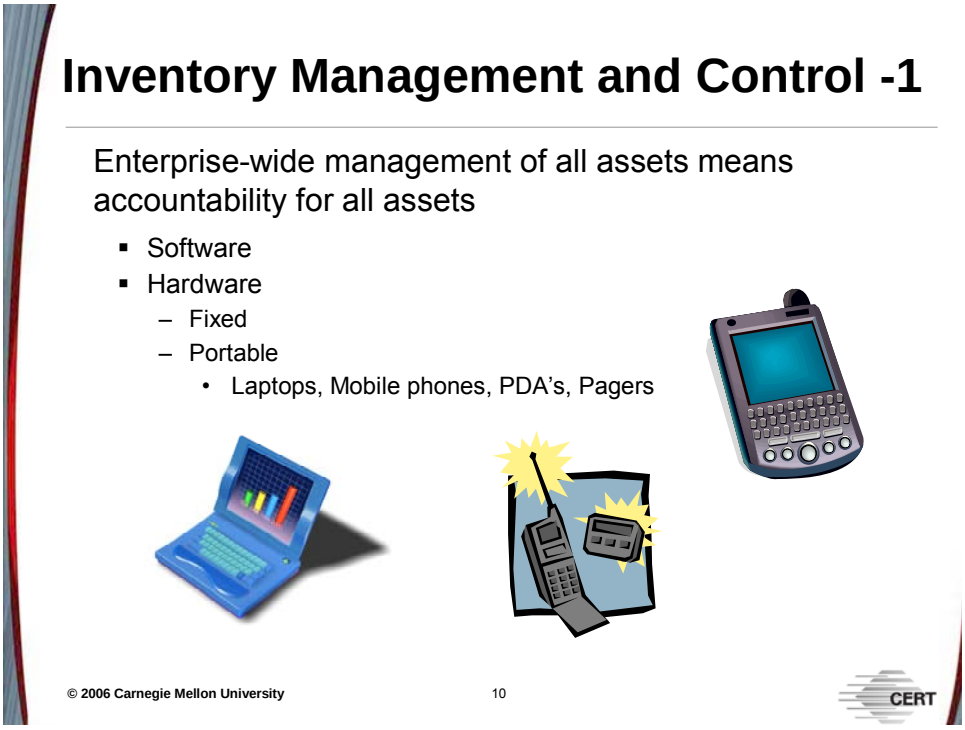


8.2.4.1 Issues to Consider

Before choosing a local updates solution, there are several questions to consider.

- Does the organization use multiple operating systems?

  Windows, Linux, Mac OS X
- Does the organization need to manage updates for many applications or just a few?
- Do the benefits justify the cost of the solution?
- What is the end goal of implementing update management, and will the solution serve that goal?



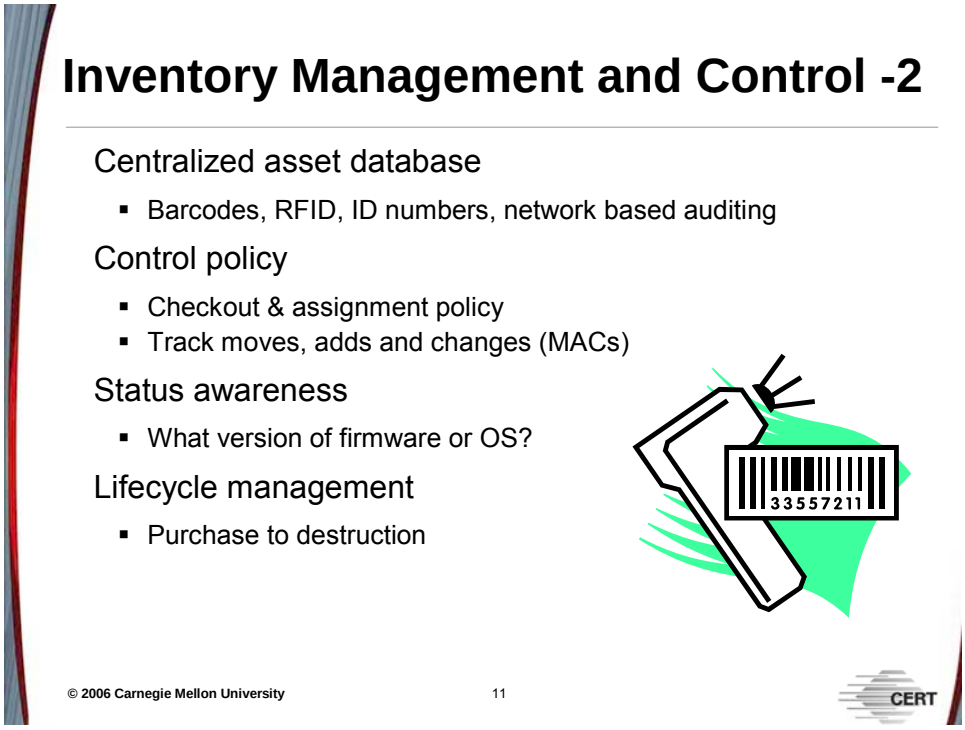
8.2.5 Inventory Management and Control

Inventory management is a critical component of information assurance and configuration management. After all, you can only control assets of which you are aware. There are multiple facets of the asset management process; these include the acquisition and tracking of assets, as well as seeing them through their individual changes and evolutions. Relevant assets include both physical (hardware) and logical (software or information) assets, such as computers, networking equipment, cell phones, PDAs, and software.

8.2.5.1 Why Is It Important?

Having necessary resources is fundamental to conducting business. It becomes a significant challenge when you aren't aware of all available resources and there is no accountability of assets. What if you are called upon to account for the status of organizational assets? It is not only financially inefficient to lack proper understanding of available computing resources, but also an administrative failure.

Thorough inventory control will enable you to more fully utilize your computing resources and better secure your organization's assets.



8.2.6 Important Components

8.2.6.1 Centralized Database

A centralized database for recording assets is fundamental to controlling the individual components of your organization's information infrastructure. Understanding this infrastructure depends on knowing what devices belong to the company. Implementing a centralized asset database is therefore the first step toward efficient asset management, accountability, and, ultimately, network configuration management.

Various methods of collecting asset information abound, but some are easier than others. Solutions such as barcodes, ID tags, and RFID (radio frequency identification) involve manually collecting data and placing tags on each asset. In some organizations, this may be necessary to supplement other means of tracking inventory.

Another means of collecting inventory data is network-based auditing, using a centralized application that actively scans a network and then inventories the network's physical and logical assets. This can be an effective and efficient method for collecting large quantities of inventory data from remote locations or for collecting a baseline inventory on which to build.

8.2.6.2 Control Policy

Once an organization has recorded its assets, it must keep track of them. An organization cannot implement Defense-in-Depth if it cannot locate the assets it needs to defend.

Therefore, every organization should have a policy defining specific procedures for assignment and use of equipment and software. These procedures can vary from organization to organization, but the end result should be the same: At any given time, an organization must be able to account for who has what. Following standard procedure should make this a straightforward process.

As part of such a policy, there should be documentation corresponding to the assignment of assets. This should be a standardized document that states what equipment is being assigned or checked out and by whom. Additionally, the document should state when the assets are to be returned and if they are being assigned as part of a project or for general use. This documentation should be acknowledged and signed by the individual responsible for it, who can be either the employee using the equipment or a project manager who will accept responsibility for apportioning assets among a group. Much of this information should be kept electronically, and reports should be generated and reviewed on a regular basis to follow up on assets that should have been returned. This form of accountability should also be integrated into organizational procedures for hiring and terminating of employees. Before employees are released of their obligation to the company, all assets should be returned according to the organizational policy.

This type of control policy should maintain accountability of assets throughout their life cycles, including the inevitable moves, additions, and changes that will affect those assets. This helps ensure that no asset is being wasted; an asset no longer suitable for use in one part of the organization might be useful in other areas that have different requirements.

8.2.6.3 Status Awareness

It is not enough to know your organization's assets and their location. It is also critical to know their status. This means knowing what firmware or code is on each specific device. It also means knowing if each device's configuration is adequate to meet the security policy of the organization. In the end, status awareness is about achieving accountability and compliance with security policy.

Not all assets can be monitored by performing software inventory and utilizing typical update management applications. Many devices require considerable monitoring and management—more than can be provided through a single management application. Devices that may require special attention include routers, switches, and wireless access points, as well as handheld or mobile computing devices.

Many of these devices can be monitored and managed through vendor-provided solutions or packages; often, these will augment the solutions that suit your other inventory and management needs.

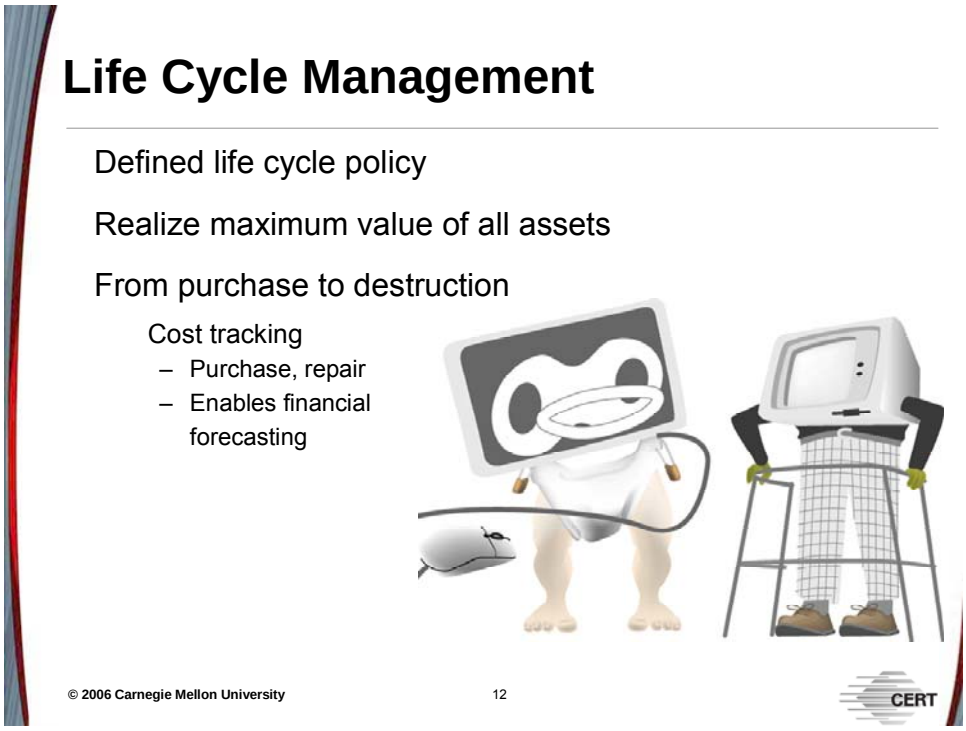
Status awareness can be achieved in multiple ways, but you may find that several solutions are necessary for managing mobile devices and networking components. In fact, there is no

solution to date that addresses all aspects of configuration management or inventory management across all devices.

8.2.6.4 Infrastructure Solutions

Below is a list of common enterprise network management software applications. These are just a sampling of the many tools that help administrators and IT managers perform configuration management effectively.

- Afaria by iAnywhere
- Cisco Works
- Computer Associates Unicenter
- HP OpenView
- IBM Tivoli



8.3 Life-Cycle Management Policy

Inventory management must address every step of an asset's life cycle, so life-cycle management is really about policy. From purchase to destruction, there must be a specific procedure for managing assets. This includes how your organization records newly acquired products, how products are assigned and tracked, and how they are disposed of when no longer usable. This also involves consideration of device-specific contracts for leased products.

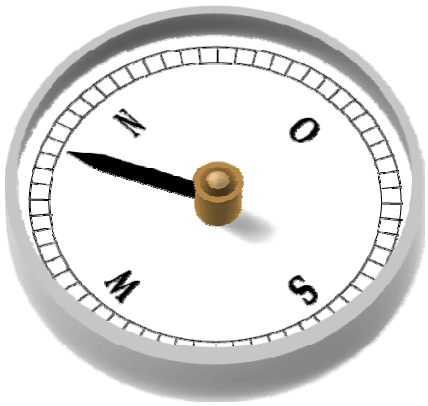
Life-cycle management is essential to realizing the full value of assets. If an organization cannot account for assets, those assets may very well be lost or stolen. Regardless of the status of a missing asset, it is being underutilized and is a wasted resource.

Life-cycle management also is not solely an IT issue. Although each organization should have procedures for determining how long an asset is to be used, this often will be determined as much by the accounting department as by the technical staff. Just because a device *appears* to have outlived its usefulness doesn't mean it can't be used. It is important to weigh the needs of the organization against its current resources, not just against what is available on the market.

Change Management -1

Providing accountability for configuration and environment changes

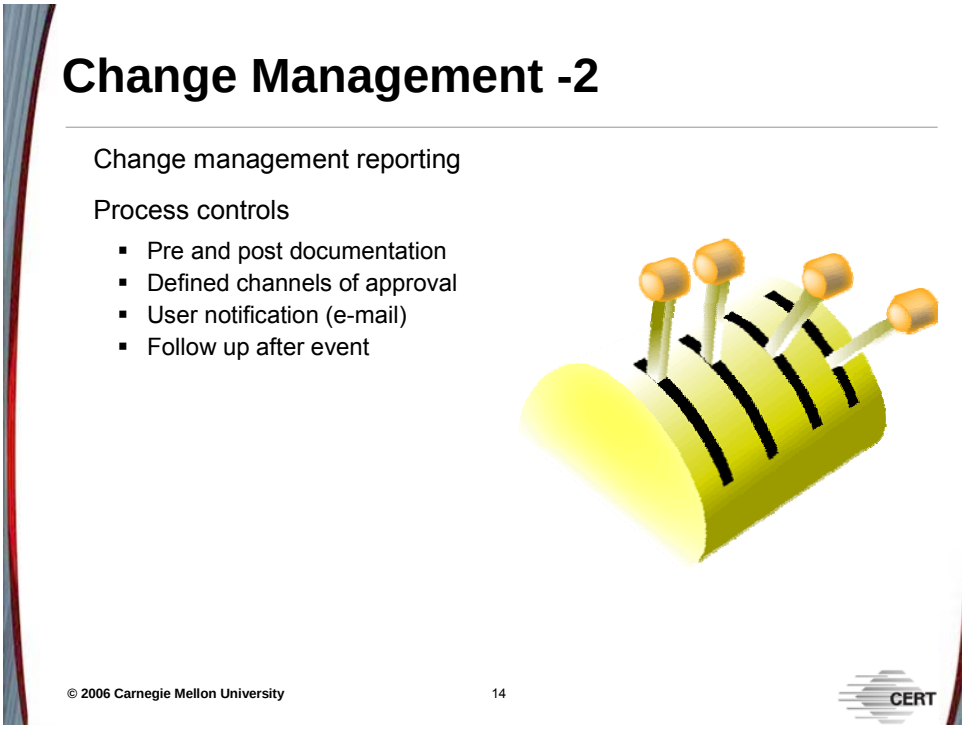
- Hardware
- Software
- Infrastructure
- Environment



13
CERT

8.3.1 Change Management

Change management is a process for managing change in the information architecture. It can include basic changes such as patching systems, upgrading essential servers, or updating the firmware of routers and switches. Each of these activities should involve adequate consideration, planning, testing, and notification before changes are made to essential devices.



In essence, change management is another step toward fulfilling the goals of configuration management. It helps ensure that network infrastructure is constructed and managed in a way that supports usability and the security of the organization. Each organization should have a process that includes specific steps for implementing changes to the network infrastructure. This process should take into consideration organizational politics, work environments, and timing issues. For example, would you make a major change to your network just before rolling out a new service that requires significant bandwidth? Change management is about managing necessary updates and changes to the network, as identified in the other areas of this module, *with a minimum of disruption*. This includes the updating of software, as identified in the update management section, as well as the management of firmware or configuration changes, as identified in the inventory management section.

8.3.1.1 Pre- and Post-Documentation

Very little documentation should need to be done during the planning stage of change management—because it already should have been done in the process of carrying out other configuration management tasks. Any additional documentation done as part of change management should be ancillary to the work already completed. This documentation should cover the changes to occur and should include a detailed explanation of why those changes are needed. Testing performed to confirm that the changes will not adversely affect the network's functionality should be completed as part of other processes (such as during the upgrade management process, as described below).

For example, a system administrator might think that the Windows operating system on the organization's workstations needs upgrading to Service Pack 2. After doing the evaluation and testing of the update management process, the administrator determines that the organization would benefit from the update and that the update would be compatible with the organization's integral business applications. He has documented all of his testing and results to process through the channels of approval.

8.3.1.2 Defined Channels of Approval

As part of the change management process, there should be clearly defined and understood channels for obtaining approval before changes are made. Many times, change will be initiated by management, while at other times it will be a part of an administrator's regular duties.

Defined approval channels can vary based on the level or type of change. Some organizations have minor changes approved by an upper-level systems administrator, while major changes affecting a large number of workstations or involving changes to the infrastructure require approval by more senior management. Channels of approval should include interactions with other systems administration staff, since responsibilities for systems may be divided among different people. By involving other administration staff, feedback also can be provided about possible problems that could result from proposed changes. Business service owners, incident response staff, and security staff also should not be left out of this feedback loop.

Approval should only be granted once the appropriate analysis and documentation have been performed. This will ensure not only that the approving manager has a full understanding of the changes to be made and their potential effects, but also that the staff carrying out the changes has thought through the entire process, including potential consequences, as well.

Once approval is granted, the rollout and testing plan should be followed as expected, with follow-up with the approving authority required if plans change significantly.

8.3.1.3 User Notification

Users cannot be left out of the equation when it is time to make changes to systems. While you may be responsible for the maintenance and reliability of the infrastructure, users are the ones who are most affected by network change.

Notification may vary depending on the effect of the change on users. Even seemingly minor changes to the network may have unexpected effects, so it's important to notify users of any change being made and possible side effects. If a change will require users' action, such as interacting with an update procedure on their host, an announcement must be made well in advance specifying the required user actions. If the change will affect user procedures or user interaction, users will require at least a minimal amount of training and/or documentation. Experience always shows that users don't like to be surprised.

Users must also be aware of any anticipated downtime for systems. It is best to have a standardized procedure for downtime notification, rather than basing notification decisions on the perceived effect or transparency to users. Mechanisms of notification can range from sending emails to posting on an intranet forum within the organization.

8.3.1.4 Monitor

Monitoring is the process of ensuring that systems are functioning as defined in the original and follow-up configuration stages. It is both an ending and a beginning to the change management process; after the original implementation of the baseline configuration, it is monitoring that will indicate when it is time for additional changes and updates to the infrastructure. In this sense, monitoring is also the beginning of the assessment phase of configuration management.

Change Management -3

Automated record keeping and controls

- What has changed?
- Was it approved?
- Does user have authority?

Products

- TripWire
- Ecora

CERT

Change management not only encompasses the implementation of change. It also involves the status of those implementations and the monitoring of the network to ensure that no unauthorized changes take place. Information assurance can only be managed if you can ensure that only authorized personnel are making changes.

Monitoring change within an infrastructure is also a full-time job that may involve many technical staff. As already noted in this module and in the Compliance Management module, their job can be made far easier by development and enforcement of a security policy.

There are also applications available for monitoring change on a network, many of which provide both intrusion detection and availability monitoring capabilities, as well as solutions for remediation of unauthorized use. We have discussed some of these solutions in the context of configuration management as well as in other modules, but a couple of applications specifically designed for automated recordkeeping and network auditing are

- Tripwire at tripwire.com
- Ecora at ecora.com

Change Management -4

Positives

- Identify possible consequences of change
- Pinpoint reasons for failure
- Accountability for actions

Negatives

- Rigid process can delay change
- Staff may not adapt easily or quickly

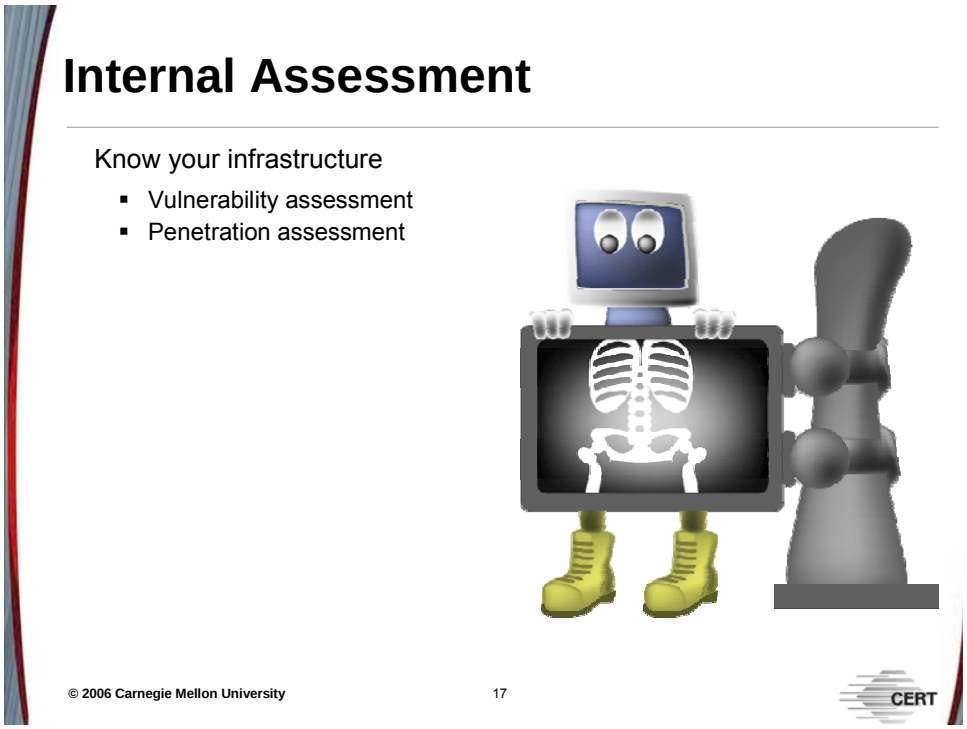
CERT

Overall, the benefits of a change management process include

- the ability to pinpoint reasons for failure more easily than when changes are made with little or no evaluation and testing. This additional network transparency is a result of following defined processes and procedures.
- increased ability to identify possible consequences of changes before they are implemented, stemming from defined channels for approval and increased interaction with organization members responsible for various elements of the network infrastructure.
- increased accountability for updates and changes to the network, not only among individuals responsible for the updates, but among all management and technical staff who participate in the change management process.

Despite these benefits, change management at times may seem like an impediment to progress rather than a helpful business process. Indeed, when change management is made a new priority at an organization, technical staff often find it difficult to adapt. There may also be concern that particular tasks should be accomplished in a shorter time frame than allowed by the change management process. To address these concerns, organizations should build in some flexibility to their change management process, with emergency procedures available to enable quick changes when required.

The key thing to remember is that an organization's infrastructure is intended to enable the mission of the organization. All changes should be evaluated and managed with this purpose in mind. The wishes of technical staff should be considered to the extent that they promote this primary purpose.

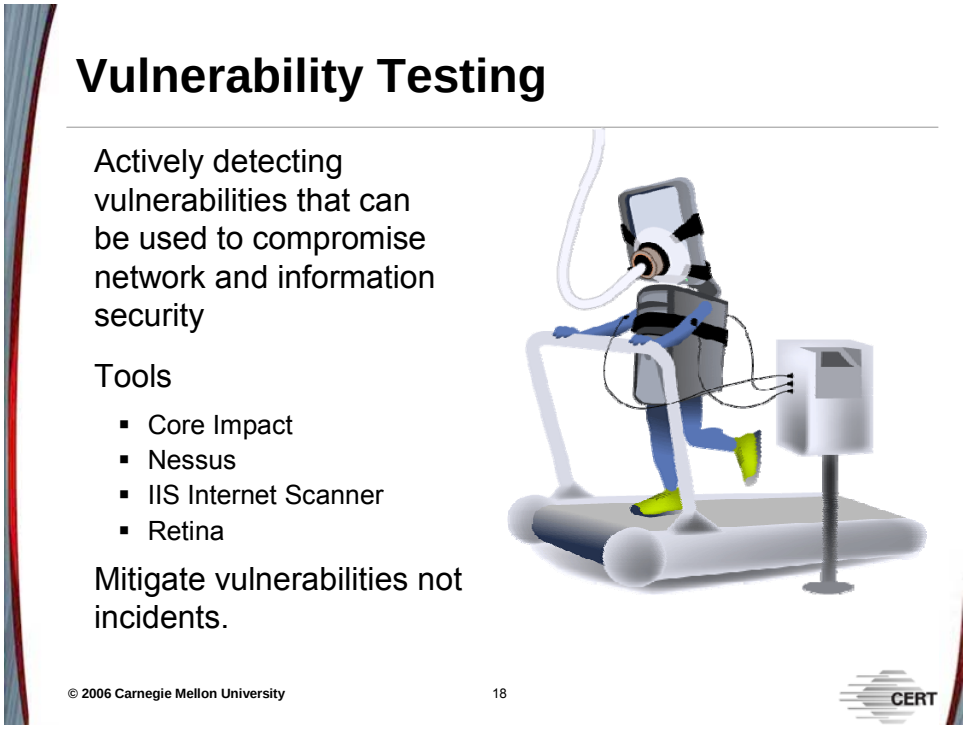


8.4 Internal Assessment

Internal assessment is another invaluable tool for organizations. Achieving absolute assurance that a network is secure is unlikely, especially without tested controls. Risk always exists. An important thing to remember when deciding if your organization will perform penetration and vulnerability testing is that if you don't, attackers will. When determining what amount of resources to apply to this effort, there should be several considerations:

- Is the infrastructure worth protecting?
- Are the assets being protected sensitive?
- Which assets are the most important?
- Is the cost of the assessment less than the cost of a security incident?

Performing internal assessment can provide significant information about vulnerabilities and weaknesses in a network. These can range from insecure devices or services to the structure of the enterprise architecture. It may also lead to the discovery of insecurities in areas previously thought to be either secure or not under threat.



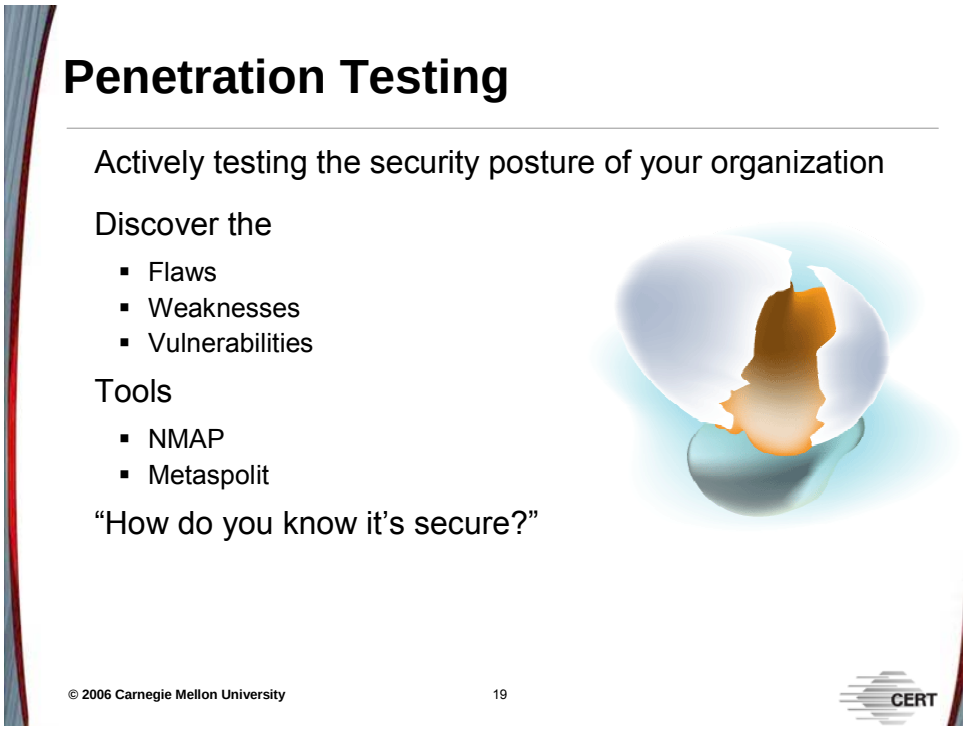
8.4.1 Vulnerability Testing

Vulnerability testing is the examination of an environment to determine its weaknesses. This can involve testing in a lab environment or scanning of a production network to detect known vulnerabilities. There are many tools available for vulnerability scanning. Many of them can enable an intruder with very little technical knowledge to exploit vulnerabilities. Tools for scanning for vulnerabilities include

- Core Impact
- IIS Internet Scanner
- eEye's Retina
- Nessus
- Foundstone's FoundScan

8.4.1.1 Why Vulnerability Testing?

To preempt attackers using vulnerability scanners, organizations may find it useful to engage in vulnerability testing. By doing so, an organization can determine possible points of weakness in its network infrastructure before an attacker finds and exploits them to gain unauthorized access to network resources.



8.4.2 Penetration Testing

If your infrastructure were a house, think of penetration testing as checking all the doors. By putting your security through a systematic series of tests, you can increase your awareness of what works, where the weaknesses are, and what must be addressed. Penetration testing involves the exploitation of vulnerabilities found in vulnerability testing, as well as an attempt to find all possible ways into a network. Think of it as verification of the vulnerability assessment and then some.

8.4.2.1 Why Penetration Testing?

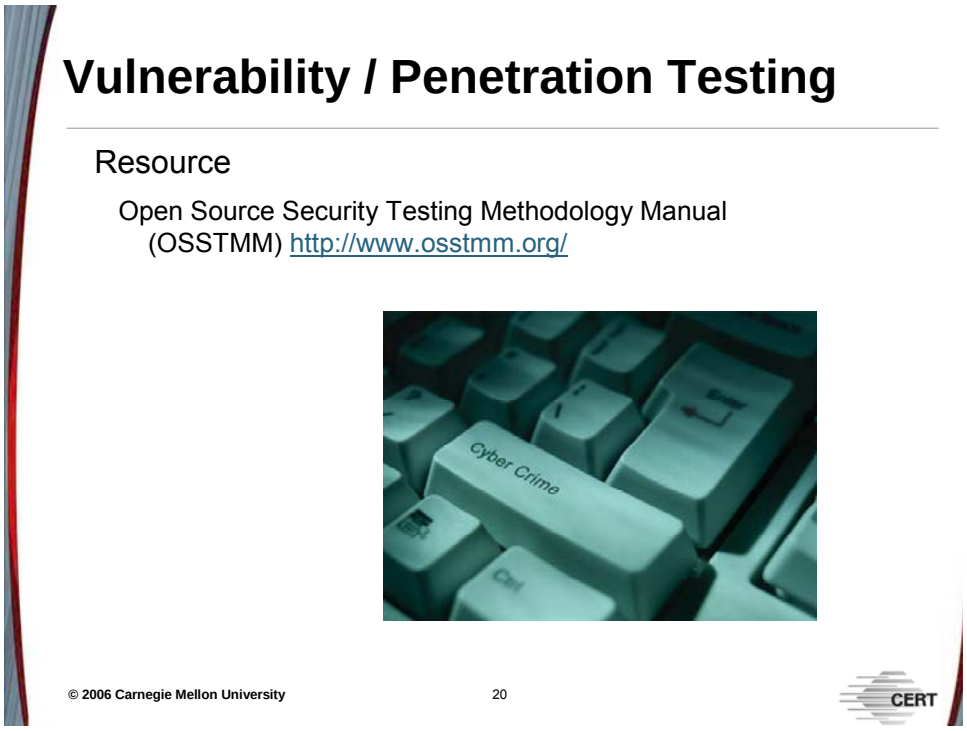
Penetration testing can be a valuable tool to aid in enumerating risks and providing concrete evidence of the threat to system resources. A penetration test can be a very convincing tool, both in highlighting the general need for increased attention to information security and in helping to identify specific goals and targets for information security efforts.

8.4.2.2 Timing for Penetration Testing

Timing is directly correlated with the usefulness of penetration testing. If you wait until an attacker or intruder has put your security through its paces, you turn your opportunity to test your own network into an incident response scenario. Once a baseline configuration is compromised, integrity is uncertain. It will take considerable effort to determine what data and systems were compromised and to test and ensure the integrity of those assets once you recover from the intrusion.

8.4.2.3 Considerations

Internal assessment is becoming increasingly easy with the development of automated tools. Many commercial and open-source tools are available for both vulnerability and penetration testing. It is important to understand that such testing should be carried out only in a systematic and documented process. Penetration testing can affect the availability, stability, and integrity of part or all of your network resources.



8.4.2.4 Outsourcing

Depending on the available resources and expertise, your organization can perform penetration and vulnerability tests using existing staff or can outsource these tasks to a managed security service provider (MSSP). Both options have benefits and drawbacks, and due care should be taken when determining the right solution for your organization.

CERT has published "Outsourcing Managed Security Services," available at http://www.sei.cmu.edu/publications/documents/sims/sim012.html, which is a valuable resource for choosing and managing an outsourced MSSP [Allen 03].

Some reasons for outsourcing vulnerability and penetration testing include

- lack of specific vulnerability assessment (VA) technical knowledge and expertise
- insufficient staff time and resources
- benefit of an outsider's objectivity and experience gained by working with a wide range of clients
- a requirement for customized vulnerability reporting and corrective action
- a requirement for ongoing, regularly scheduled VA activities

- benefit of an external “intruder’s eye view” of the organization’s security posture
- a requirement for independent affirmation of the client’s security posture to build customer and partner confidence[33]

[33] Allen, J., Gabbard, D., & May, C. (2003). *Outsourcing Managed Security Service.* http://www.sei.cmu.edu/publications/documents/sims/sim012.html.

Review Questions

1. What are the benefits of properly implemented configuration management?
2. What are several reasons for implementing localized software updates?
3. Explain the purpose of inventory control.
4. Explain the importance of change management.
5. Explain two mechanisms for performing internal assessment and their benefits.

Module 9: Incident Management

This module discusses various types of incidents, the benefits of incident management, the incident response process, and developing an incident response program.

Instructional Objectives

Define an incident.

Outline the benefits of incident management.

Describe the incident response process and relate it to incident management.

Develop a long-term incident response process.

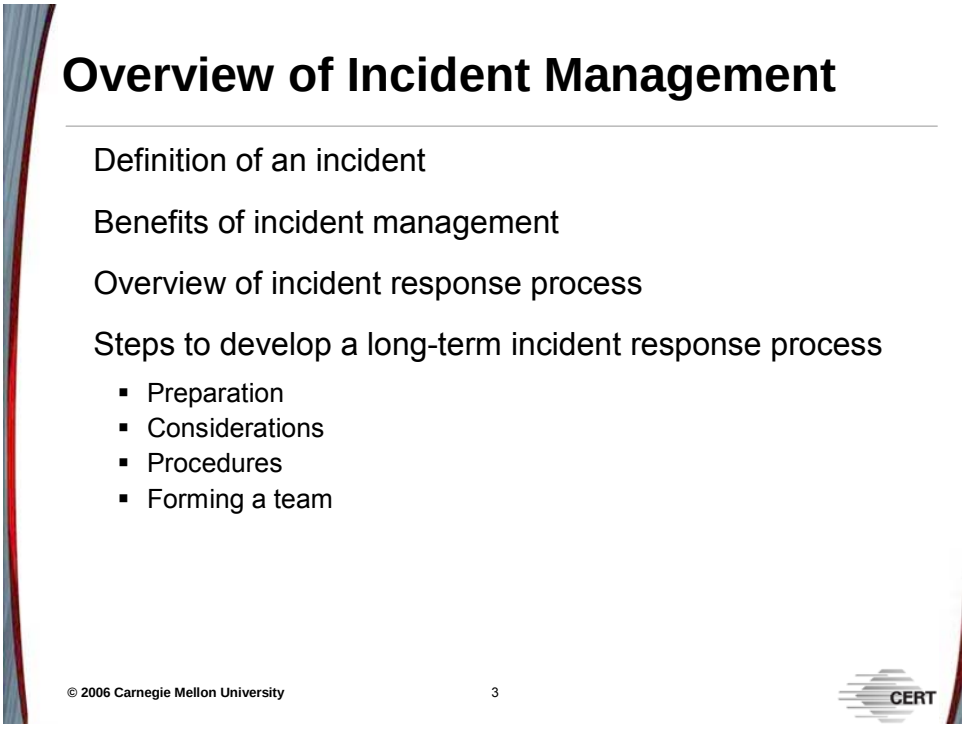
- Preparation
- Considerations
- Procedures
- Forming a Team

2

CERT

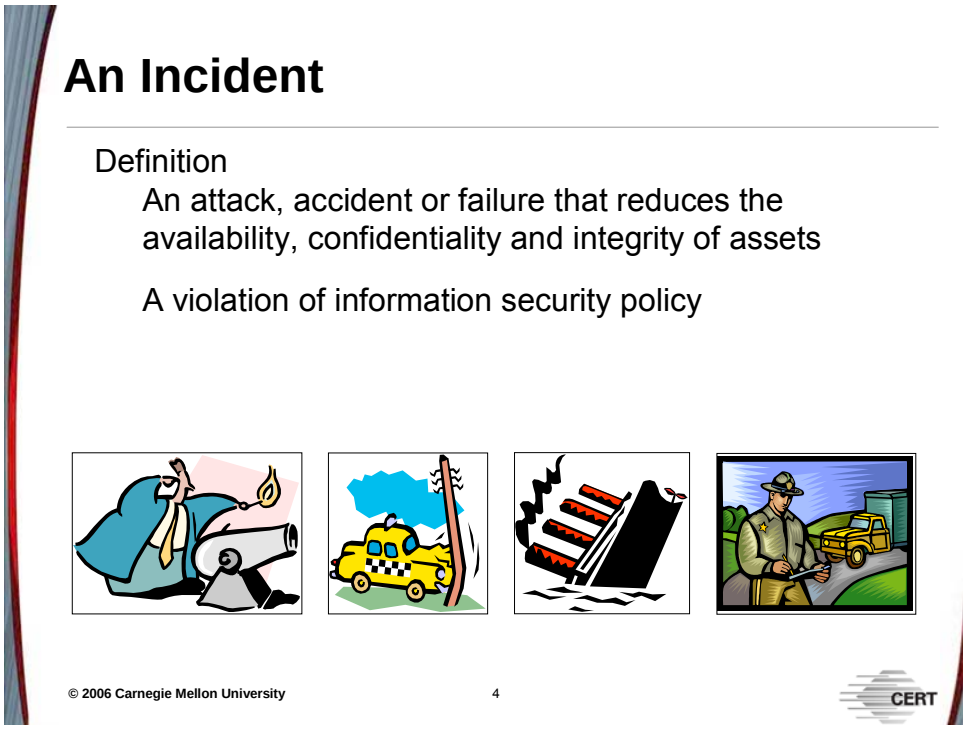
This instructional module will enable students to complete all of the above learning objectives.

9 Overview of Incident Management



The goal of this module is to educate managers and decision makers about incident management (IM) concepts so that they can develop an effective organizational incident response capability. A solid incident management plan must take into consideration many issues that involve both internal and external organizations and factors. This module attempts to cover many of these relevant issues on a level broad enough to enable individuals to make sound incident management decisions.

The first objective of this module is to present a useful and comprehensive definition of the term *incident*. The second objective is to define incident management and outline the benefits it brings to an organization. This is important; if you are not able to convey the importance of incident management to upper-level management and executives, the quality of incident response is likely to suffer due to a lack of support and resources. The third objective is to provide an overview of the incident response process within the overarching framework of incident management. The last objective is to outline a process for developing a long-term incident response process. It is organized into four main sections: preparation, considerations, procedures, and forming a Computer Security Incident Response Team (CSIRT).



9.1 Definition of an Incident

Before you can react and respond to an event, you must first understand what constitutes an event. Therefore, effective incident management capabilities and plans cannot be developed until an "incident" has been well defined. For the purposes of this module, we define an incident as[34]

1. an attack, accident, or failure that reduces the availability, confidentiality, or integrity of assets
2. a violation of information security policy

The first definition is fairly straightforward. However, it is important to note that we will confine our discussion of incidents mainly to computer security events, although other events such as power outages may occur in the course of your work and indirectly have security-related consequences.

Even within the realm of computer security events, one should avoid jumping to conclusions. Incidents initially should be viewed as events that are independent of intent. Only upon further investigation can an incident be classified as malicious (an attack, such as a denial of service) or benign (a failure, such as an inadvertent router misconfiguration).

The purpose the second definition of "incident" is to include events that do not reduce the availability, confidentiality, or integrity of a particular asset but nonetheless pose a security

[34] This definition was presented in the lecture *Security Architecture & Analysis* at Carnegie Mellon University.

risk to the organization. For example, suppose an organization has a policy that forbids employees from storing files containing confidential company information on their laptops. The end goal—the mission of the policy—is to reduce the risk of unauthorized access to those files. Violation of this policy would therefore be considered an incident. Even though the storage of sensitive files on a laptop computer does not immediately compromise confidentiality, it greatly increases the risk of doing so. Therefore, it is important to treat this type of violation as an incident that requires an appropriate response.

Now that we have defined an incident, we can define incident management. In brief, incident management is the ability to provide end-to-end management of computer security events and incidents across the enterprise or organization.

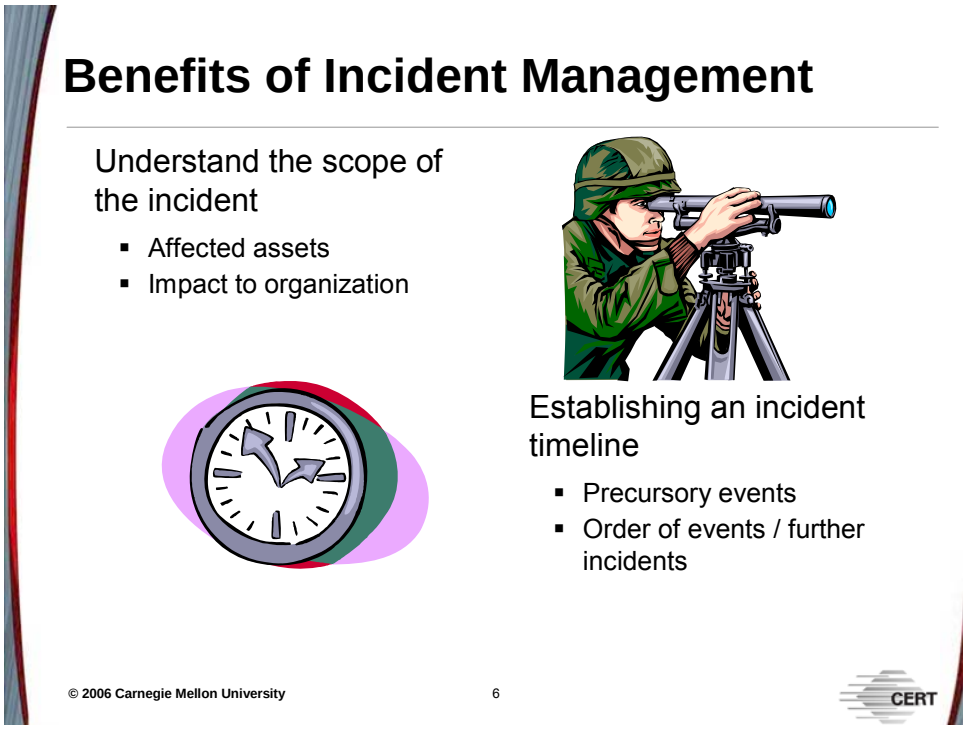
For computer security incident management to occur in an effective and successful way, all the tasks and processes being performed must be viewed from an enterprise perspective. This means identifying how

- tasks and processes relate
- information is exchanged
- actions are coordinated

Looking only at the response part of the process misses key actions that, if not done in a timely, consistent, and quality-driven manner, will impact the overall response, possibly delaying actions due to the confusion of roles and responsibilities, ownership of data and systems, and authority.

Response can also be delayed or ineffective because of communication problems (not knowing whom to contact) or poor quality information about the event or incident. Any impact on response timeliness and quality can cause further damage to critical assets and data during an incident. Indeed, identifying and defining the roles and responsibilities of various participants across the enterprise is a key part of setting up any incident management capability.

Incident management, then, is an abstract, enterprise-wide capability, potentially involving every business unit within the organization. As such, it is a subset of other security management activities and functions that can be applied to achieve Defense-in-Depth, and therefore often crosses into and includes some general security tasks and practices.



9.2 The Benefits of Incident Management

Understanding the benefits of incident management is important not only for developing an incident response plan; it also enables individuals to communicate the significance of IM to upper-level management and executives. This is critical for garnering top-level support, without which no security effort can succeed.

We'll discuss the benefits of incident management first, and then look at how it can be implemented as a defined, repeatable, sustainable process.

9.2.1 Understanding the Scope of an Incident

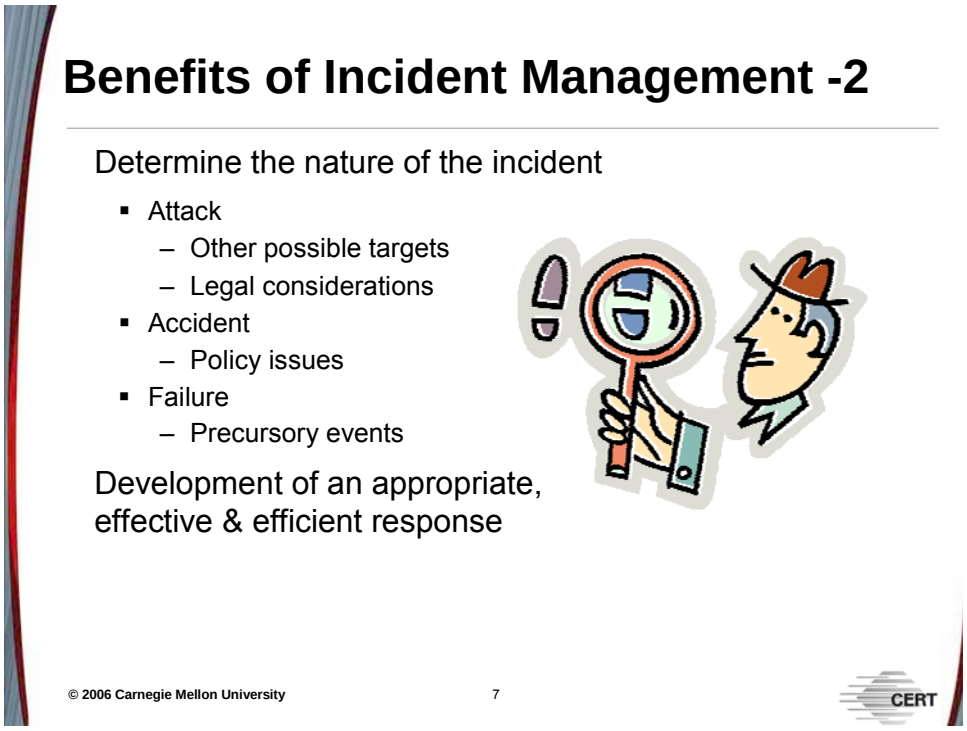
One of the main benefits of implementing an IM plan is that it enables an organization to put repeatable, consistent processes in place to detect, assess the scope of, and analyze incidents, with the end goal of providing an effective response. Specifically, implementing IM processes helps an organization determine which assets have been affected by an incident as well as the organization-wide impact. For example, if a company is hit by a worm, an organization would follow its procedures for identifying the computer that has been infected, the effect the virus will have on the infected computer, and the effect this will have on day-to-day business operations over the long term.

9.2.2 Establishing a Timeline

Another component of formulating a complete response is establishing an incident timeline—a chronological map of events occurring prior to, during, and after an incident. This helps the

organization better understand the incident because it illustrates the incident's life cycle and explains how it unfolded.

For example, a company discovers that its Web server has been attacked. Investigation into the incident reveals that port scanning was detected on the network just prior to the attack. Furthermore, examination of the Web server reveals that the attacker was able to gain root access and compromise the company's internal network. Through IM, the company is able to establish an incident timeline, identify the events leading up to the attack, and determine the order of events occurring during the full attack. It is thus able to demonstrate that the Web server attack led to another incident.



9.2.3 Determination of Intent

Processes for IM also provide a clear path to determine the intent or nature of an incident. Remember that an incident can be an attack, accident, or failure. The nature of the incident affects how the organization will respond to it. All responses to an incident will involve some form of mitigation; however, additional considerations sometimes must be taken into account. For example, if a malicious attack occurs, an organization will have to consider legal ramifications, public relations issues, and other possible attack targets. For an accident such as an employee's inadvertent deletion of a critical file share, a company's response will be narrower but may involve a look at current policies and employee training programs. Lastly, in the event of a failure, such as a Web server crash, the response may be more heavily focused on establishing an incident timeline to determine the events that led to the failure.

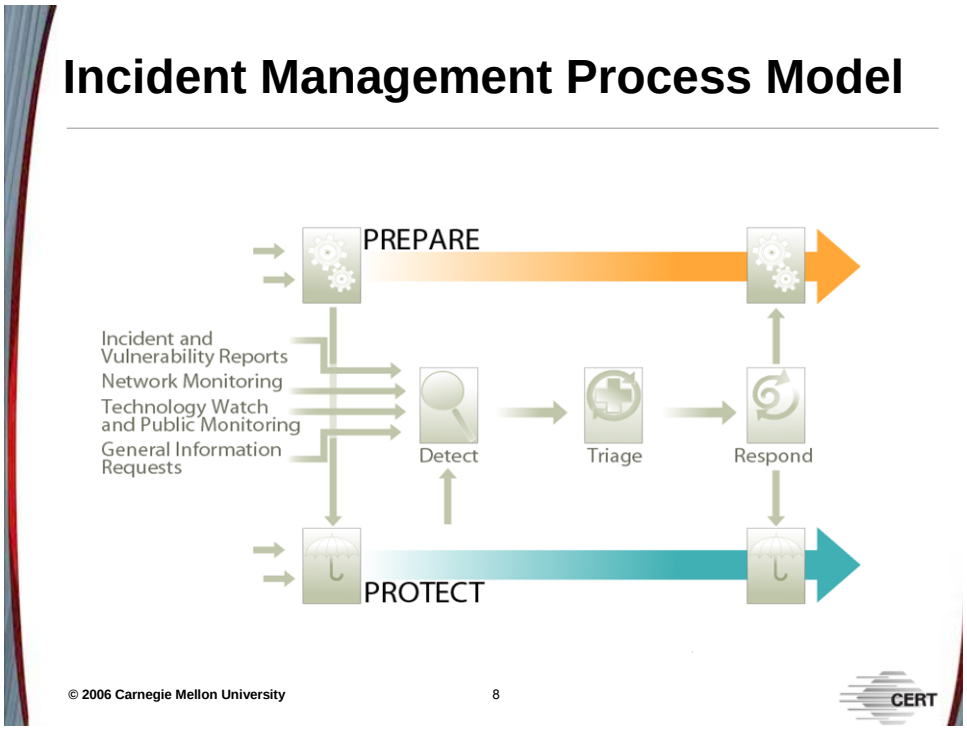
9.2.4 Responding Appropriately, Effectively, and Efficiently

One primary reason for implementing IM is the improved capability for effective, efficient response that minimizes downtime and maximizes system resiliency. Responders in a defined process know their roles and responsibilities and are able to take quick action when needed. The end result is a reduction in the cost and time required to recover from an incident. In other words, effective incident management is Defense-in-Depth in action.

9.2.5 Proactive Prevention

Incident management is not just about response. It also includes proactive activities that help prevent incidents, for example by identifying vulnerabilities in software that can be addressed before they are exploited. Another proactive action is the training of end users to understand the importance of computer security in their daily operations and to define what constitutes abnormal or malicious behavior, so that end users can identify and report this behavior when they see it. Some of these tasks may be done by persons outside of the security department.

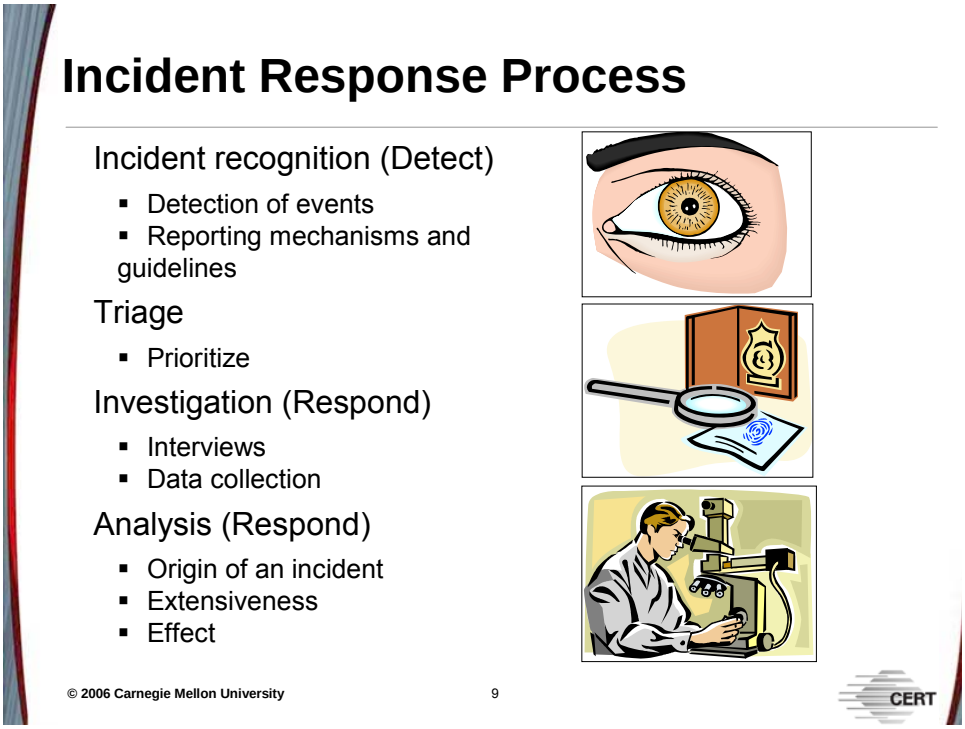
We'll discuss some of these proactive measures in more detail later in this module, when we talk about implementing a CSIRT to ensure high-quality incident management over the long term.



In order to unlock the benefits of incident management described above, an organization will need to implement a repeatable, sustainable process. One possible model, developed by the CSIRT Development Team in the CERT Program, is depicted here and consists of five steps:

1. Prepare
2. Protect
3. Detect
4. Triage
5. Respond

This model is described in more detail in SEI Technical Report CMU/SEI-2004-TR-015, *Defining Incident Management Processes: A Work in Progress* [Alberts 04]. This report is available at http://www.sei.cmu.edu/publications/documents/04.reports/04tr015.html.



9.3 Incident Response Process

The previous modules of this course have dealt in depth with various aspects of the Prepare and Protect steps of the incident management process: compliance management, risk management, identity management, authorization management, accountability management, availability management, and configuration management. We now will examine the remaining three steps—Detect, Triage, and Respond—which become important once an incident actually occurs. These three steps, taken together, can be viewed as an incident response process (IRP), which can be further broken down into the steps outlined on this slide.

This section is intended to serve as a broad overview of the incident response process for those who are unfamiliar with the concept. A fictitious company called Ambisoft will be used to illustrate each of the incident response process phases.

9.3.1 Incident Recognition (Detect)

The incident response process is usually initiated by the detection of anomalous activity, either through an alert from an intrusion detection system or in some other manner. For example, suppose an employee at Ambisoft, Chuck, has been noticing lately that his user account is temporarily locked out when he arrives at work in the morning. Chuck notifies the IT department of his problem, which is immediately flagged as suspicious since Ambisoft has a policy in place to lock out users for an hour after five failed login attempts. The IT staff,

suspecting a potential attack, hands the issue over to the Ambisoft incident response team. It is important to note that an organization with a well-formed overall incident management process will have a defined incident reporting mechanism in place. The section on *Incident Management Procedures* will cover incident reporting mechanisms in more detail.

9.3.2 Triage

At this point, the Ambisoft incident response team will triage the incident. Depending on what other events have been reported to the team, Chuck's report may be assigned a high, medium, or low priority. For example, if the incident response team is fighting to contain a new worm that threatens to take down the entire network, Chuck's report may not be investigated immediately. On the other hand, if the team has sufficient resources to look into Chuck's report, a best practice is to handle the report as quickly as possible.

No matter the priority of a report, it should always be investigated as soon as possible, and the incident response team should follow up with the person or group who made the initial report.

9.3.3 Investigation (Respond)

Incident investigation and analysis correspond to the Respond step of the process model presented earlier in this module. Generally, the investigation phase consists of gaining a better understanding of the situation and gathering as much pertinent information as possible. The first part of an investigation may involve interviewing the parties who first discovered the incident. Next, the incident response team will begin to collect data from computers and equipment involved in the incident. If the incident seems to call for forensic analysis capabilities, this process can include volatile data collection as well as forensic duplication of persistent data. *Volatile data* is information about a computer's current state and can include time/date stamps, system time, currently running processes, and a list of open network connections [Mandia 03]. At any point in time, this data could change. On the other hand, forensic duplication is bit-for-bit duplication of an object, ranging from a single file to an entire hard drive. This data is termed "persistent" because it is unlikely to change as rapidly as volatile data and, under normal circumstances, will not be lost if a machine is rebooted. Each collection method gathers data that the other method cannot. There are other methods for collecting data; however, they are outside of the scope of this module.

So, back at Ambisoft, the incident response team interviews Chuck to gather more information about his incident. They learn that the lockouts have been occurring for the past week, but, curiously, nothing happened this morning. The first thing the incident response team does is to collect volatile data from Chuck's computer. This action creates a "snapshot" of the current state of Chuck's computer. Next, the team performs a forensic duplication of Chuck's hard drive. With this data in hand, the team is now ready to begin its analysis. (For this simplified example, we will not consider chain of custody, but it is important to consult forensic specialists about this issue in real-world investigations.)

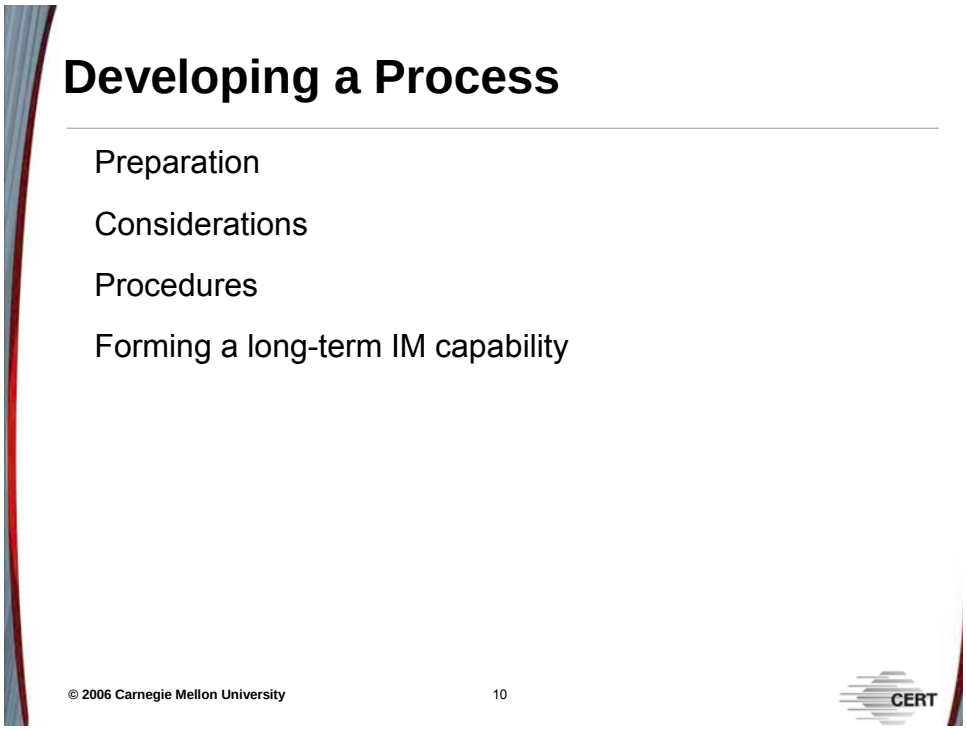
9.3.4 Analysis (Respond)

The analysis phase consists of examining the data gathered during the investigation to determine the root cause of the incident. Analysis can involve sifting through log files and volatile data in search of anomalous activity. There are also sophisticated tool sets such as Encase[35] that can perform highly specialized, complex forensic analysis if needed [Guidance 05]. The ultimate goal of the analysis phase is to determine an incident's origin, extensiveness, and effect on information and hardware assets, as well as possible mitigation or response strategies.

The Ambisoft incident response team began to analyze data collected from Chuck's computer and noticed that the volatile data revealed a mysterious network connection at the time of the collection. Using the forensic duplication of the hard drive, the team was able to examine the event logs from Chuck's computer, which showed an unnaturally high number of failed login attempts between 10 p.m. and 4 a.m. The event log also showed that Chuck's account had been successfully accessed the previous night at 11:30 p.m. This led the team to another entry in the event log shortly thereafter, indicating that the privileges to a guest account had been elevated to Administrator level. As it turned out, the mysterious connection found on Chuck's computer was initiated by the compromised guest account, and Chuck's computer was being used to host an illegal file share.

This example is a simplification of the incident response process. However, it should provide enough detail for you to understand the essence of incident response. Also, keep in mind that incident response is just one component of incident management. The purpose of this module is to impart a broad understanding of the key issues pertinent to all incident management.

[35] http://www.guidancesoftware.com.

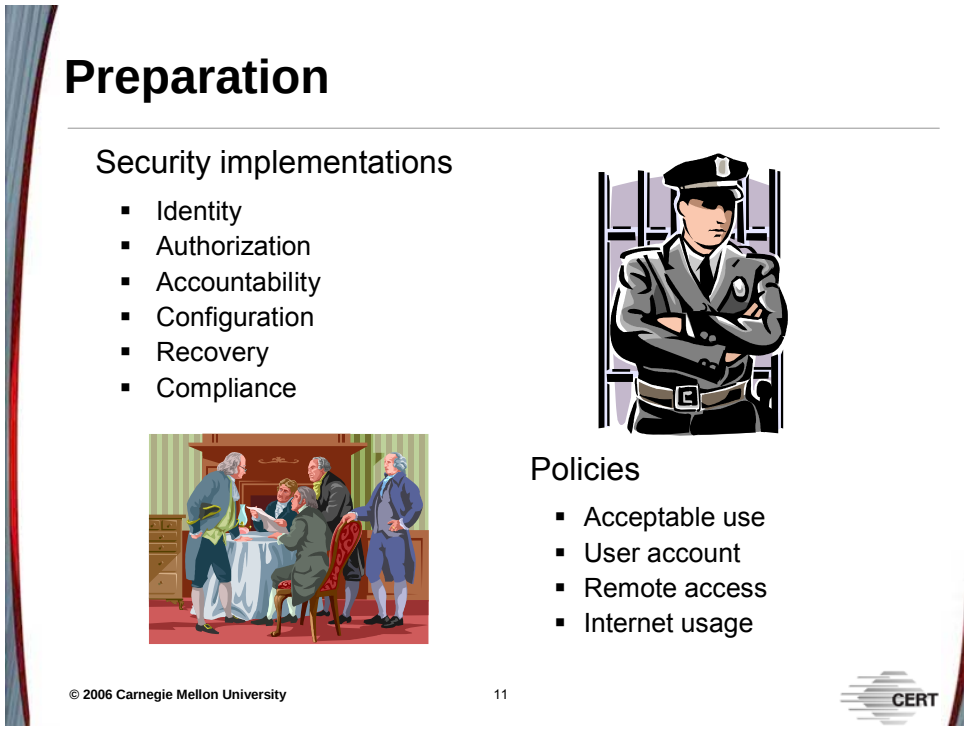


9.4 Developing an Incident Response Process

It seems clear that an incident response process is vital to the overall incident management endeavor—but where do you start? It is management's duty to ensure that a sound process is put in place and continuously revised to keep up with changing trends and technologies.

The first step in implementing any successful program is to lay a strong foundation. In the case of developing an incident response process, this involves making sure the appropriate supporting components exist within the organization. Second, several considerations must be taken into account during development of an incident response process: legal issues, effect on business, and handling of critical assets. Third, there are processes and procedures that the organization must implement for the overall effort to be successful. One example of such a process is a mechanism for incident reporting and categorization.

Finally, an organization must consider whether to implement some type of CSIRT or other incident response team. Members of this team will be the staff members who actually handle and respond to incidents within the organization. They may be distributed across several geographically distant branches of the organization or centralized at its headquarters or another location. They may work on incident response full-time or perform IR as needed in addition to their regular job roles. The point is, there are many ways of implementing an incident response team. We will discuss one, forming a CSIRT, in more detail later in this module, but you need not feel bound to the CSIRT model.

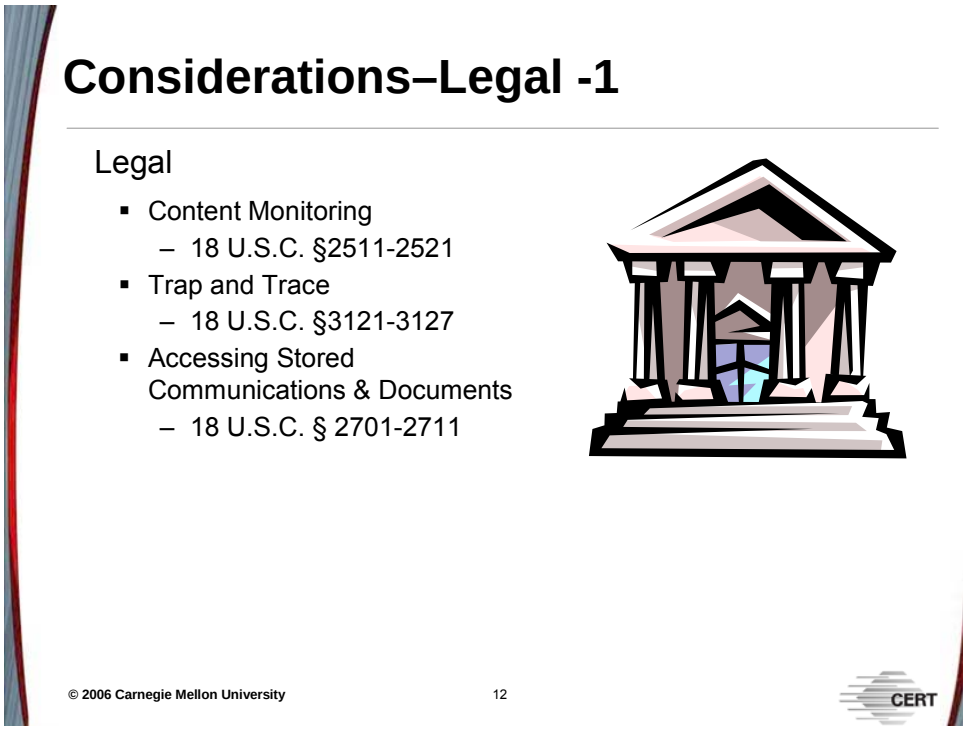


9.4.1 Preparations

As defined earlier, incident response (Detect, Triage, Respond) occurs in reaction to an attack, accident, failure, or security policy violation. However, it is important not to overlook the other two steps of incident management: Prepare and Protect. The Defense-in-Depth approach to incident management aims to keep incidents—and, therefore, the need for incident response—to a minimum. Therefore, a core goal of sound incident management is to Prepare and Protect to ensure that an organization's proactive security components are sufficient.

We have discussed many of these components in previous modules, including accountability management, authorization management, availability management, identity management, configuration management, risk management, and compliance management. Regarding compliance management, it is especially important to ensure that sufficient security policies are in place, including user account, remote access , and acceptable use policies. The next section, which will cover legal considerations related to computer security incident management, will specifically stress the importance of having policies in place that require employees' consent to have their network traffic monitored.

All of these security components are a part of this curriculum; you should be familiar with them by now. For more information about any security component mentioned here, please refer to its corresponding module in the course.



9.4.2 Considerations

9.4.2.1 Legal Considerations

An organization must take into account several legal issues as part of developing a computer security incident response process. These issues pertain to 1) the legality of an organization's actions in response to an incident and 2) incidents involving law enforcement as well as illegal activities. Often, an organization will monitor packet header information and traffic content and search through data devices in response to an incident. There are a number of statutes relevant to these actions. It is important to grasp them, at least at a high level, so that you understand the effect these laws will have on your organization's incident response.

> *It is important to understand that the following sections are intended to provide individuals with a general understanding of legal matters that may be relevant to developing an incident response process. Individuals also should consult with their organizations' legal counsel.*

9.4.2.2 Full Content Monitoring: 18 U.S.C. §2511-2521

Full content monitoring can be an extremely useful tool during the incident response process because it enables observation of all the traffic on a network. This type of monitoring captures packet headers along with payload information. For example, if an intruder were to open an unauthorized telnet session on an internal server, full content monitoring would make

it possible to view the entire telnet session and see exactly what the intruder did. Despite its usefulness, however, full content monitoring is subject to restrictions.

United States Code Title 18 Section 2511 makes it illegal for anyone to intercept wire, oral, or electronic communications while they are being transmitted [Mandia 03]. This statute is commonly known as the federal wiretap law. However, exceptions exist that make it possible for organizations to capture (intercept) traffic on their network without violating this statute. Specifically, 18 U.S.C. §2511 (2)(d) states the following:

18 U.S.C. §2511 (2)(d)
It shall not be unlawful under this chapter for a person not acting under color of law to intercept a wire, oral, or electronic communication where such person is a party to the communication or where one of the parties to the communication has given prior consent to such interception unless such communication is intercepted for the purpose of committing any criminal or tortious act in violation of the Constitution or laws of the United States or of any state.

Therefore, an organization can perform full content monitoring so long as it obtains consent from its employees beforehand. One of the best ways to obtain employee consent is through creation of a policy that employees must sign to acknowledge that they have read the policy, understand it, and agree to it. Banners can also be set up on login screens to inform individuals that by using the network they consent to monitoring of their traffic.

It is important to point out that this exception to the federal wiretap statute requires the consent of only *one* of the two parties involved. The user is one party, and the system administrator of the machine receiving the communication is usually considered the second party [Mandia 03]. This means that garnering prior consent from the system administrator would also enable an organization to perform full content monitoring. However, some states, such as Pennsylvania, have more restrictive laws that require the consent of all parties to the communication. It is generally best to obtain the consent of employees to ensure a sound legal footing.

Another exception that may be applicable is 18 U.S.C. §2511 (2)(a)(i), quoted below.

18 U.S.C. §2511 (2)(a)(i)
It shall not be unlawful under this chapter for an operator of a switchboard, or an officer, employee, or agent of a provider of wire or electronic communication service, whose facilities are used in the transmission of a wire or electronic communication, to intercept, disclose, or use that communication in the normal course of his employment while engaged in any activity which is a necessary incident to the rendition of his service or to the protection of the rights or property of the provider of that service, except that a provider of wire communication service to the public shall not utilize service observing or random monitoring except for mechanical or service quality control checks.

This exception states that monitoring can be conducted by the provider (the business) of the electronic communication to protect its rights or property. However, this exception does not

permit an organization to conduct unlimited monitoring; rather, any monitoring done on the basis of this exception must remain within the scope of the situation. For example, if an employee is suspected of revealing trade secrets, the organization cannot monitor resources that are unrelated to this suspicion.[36]

Overall, developers of an incident response process should consult legal counsel to determine the applicability of these exceptions to their organizations. As noted in our description of the exception for consent, laws can vary from state to state.

Pen Registers/Trap and Traces: 18 U.S.C. §3121-3127

A pen register and trap and trace are similar to full content monitoring except that they do not capture the content of the communication. You can think of full content monitoring as a wiretap that records the conversation between two parties, whereas a pen register or trap and trace records only the phone numbers being dialed. Specifically, a pen register records outgoing information, while a trap and trace records incoming information.[37] In the cyber world, executing a trap and trace or pen register is the equivalent of capturing a packet's header information, such as the source or destination address and the source or destination port.

18 U.S.C. §3121, which is entitled "General prohibition on pen register and trap, and trace device use; exception" states the following:

18 U.S.C. §3121 (a): In General
Except as provided in this section, no person may install or use a pen register or a trap and trace device without first obtaining a court order under section 3123 of this title…

Exceptions to this statute are as follows:

18 U.S.C. §3121 (b): Exception
The prohibition of subsection (a) does not apply with respect to the use of a pen register or a trap and trace device by a provider of electronic or wire communication service--
(1) relating to the operation, maintenance, and testing of a wire or electronic communication service or to the protection of the rights or property of such provider, or to the protection of users of that service from abuse of service or unlawful use of service; or
(2) to record the fact that a wire or electronic communication was initiated or completed in order to protect such provider, another provider furnishing service toward the completion of the wire communication, or a user of that service, from fraudulent, unlawful or abusive use of service; or **(3)** where the consent of the user of that service has been obtained.

The first exception, (b)(1), deals with protection of the provider's rights and property, similar to the exception with regard to full content monitoring. The second exception, (b)(2), allows

[36] Schwartz, Joel. "Cyber Security - the Laws that Govern Incident Response." Security Professionals Conference, 2004.

[37] Ibid.

providers to perform pen registers and trap and traces to ensure that electronic communications are working properly. An example of this would be a network administrator collecting TCP header information to determine whether network components are communicating properly. The third exception, (b)(3), requires the consent of the user.

As with full content monitoring, be sure to check with a lawyer to determine the applicability of these statutes and exceptions to your own organization.

Accessing Stored Communications and Documents: 18 U.S.C. § 2701-2711

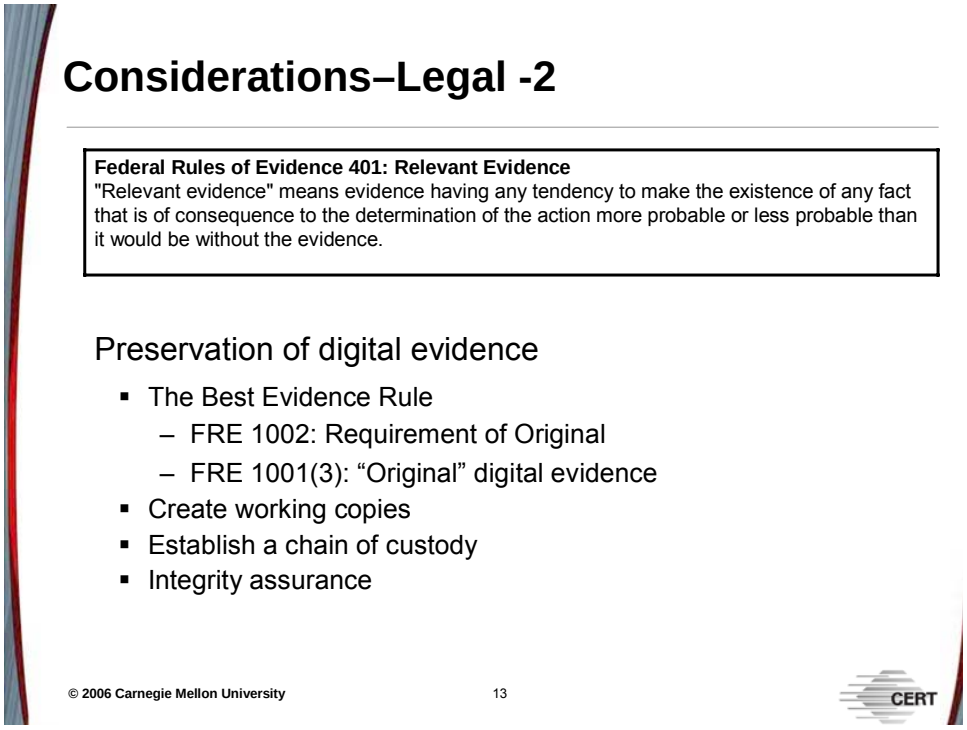
Regulations governing stored communications differ from the wiretap, pen register, and trap and trace laws governing communications in transit. In all cases, the content of the communication can be the same; however, the statutes are less strict for stored data. For example, intercepting an email using a network sniffer would be considered a wiretap and would be covered under 18 U.S.C. §2511. On the other hand, if the email were accessed from storage on the company mail server, it would fall under 18 U.S.C. §2701, which defines unlawful access to stored communications[Mandia 03]. The following excerpt is taken from a paper written by Fraser A. McAlpine and Michael Droke of the Littler Mendelson law firm regarding an organization's ability to access stored communications and documents: [38]

> The Electronic Communications Privacy Act, 18 U.S.C. ' 2701, *et seq.*, clearly gives an employer the right to access an employee's e-mail and voice-mail messages if the messages are maintained on a system *provided by the employer*. However, employers may not access messages if the system is provided by an outside entity without the authorization of the employee who communicated the message or the intended receiver of the message.

Other Laws

In this section, we have provided an overview of the most common laws facing computer security incident response personnel. However, depending on the scope of a team's incident response efforts, it also may need to take into account laws concerning personnel privacy and First Amendment issues, as well as any company-specific policies that may restrict the scope of incident response.

[38] McAlpine, Fraser A. & Droke, Michael. "Electronic Privacy In Employment." January 1998. http://library.findlaw.com/1998/Jan/1/126935.html.



9.4.2.3 Preservation of Digital Evidence

If an organization anticipates that an incident will result in a legal proceeding, it should ensure the preservation of all relevant digital evidence. Rule 401 of the Federal Rules of Evidence (FRE) provides the following definition:

FRE 401: Relevant Evidence
"Relevant evidence" means evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence.

This section will review a few important concepts for gaining a better understanding of digital evidence and its preservation. It is not intended to serve as legal advice. For detailed information, please consult a lawyer.

The Best Evidence Rule

Before evidence preservation concepts can be reviewed, it is important to understand the best evidence rule. The idea behind this rule is that for a piece of evidence to be admissible in court, it must best represent the original item for greatest accuracy. In most cases, this means it must be the original item itself. This rule arose in the eighteenth century, when copies were handwritten by clerks and it was assumed that unless the original was produced in court,

there was a significant change of the copies containing errors or fraud.[39] The best evidence rule is outlined in FRE 1002:

FRE 1002: Requirement of Original
To prove the content of a writing, recording, or photograph, the original writing, recording, or photograph is required, except as otherwise provided in these rules or by Act of Congress.

Since the purpose of the best evidence rule is to verify the accurate representation of a piece of writing, recording, or photograph, electronic copies of files are considered as originals. This rule is codified in FRE 1001(3):

FRE 1001(3): Original
An "original" of a writing or recording is the writing or recording itself or any counterpart intended to have the same effect by a person executing or issuing it. An "original" of a photograph includes the negative or any print therefrom. If data are stored in a computer or similar device, any printout or other output readable by sight, shown to reflect the data accurately, is an "original."

This is important because it is not always feasible to preserve and store the original media that contains the evidence. For example, suppose a small company detects an incident on its Web servers. It would not be practical for the company to take down its Web site due to the hard drives' being stored as evidence. Because of FRE 1001(3), forensic duplications of the hard drives would suffice as original evidence.

To summarize, in most cases the Federal Rules of Evidence dictate that only original writings, recordings, or photographs are admissible in a court of law. For the purposes of digital evidence, however, an electronic copy can serve as an original.

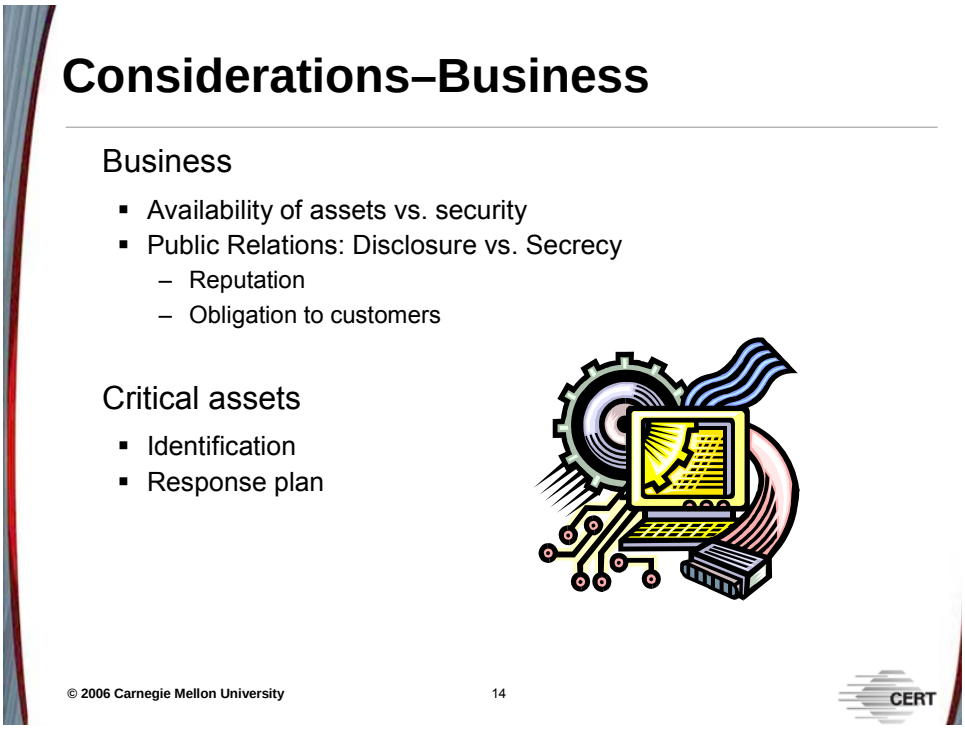
Digital Evidence Preservation Procedures

Procedures such as creating working copies, establishing a chain of custody, and assuring integrity are important for preserving digital evidence. Working copies are important for performing forensic analysis on digital data without tampering with the original evidence. A common practice for creating working copies is forensic duplication. As noted earlier, forensic duplication involves creating a bit-by-bit exact copy of persistent data (ranging from a file to an entire hard drive).

Additionally, digital evidence must be stored in a protected area, and proof of its integrity must be provided. Therefore, it is important that a chain of custody be established, starting the moment that a piece of digital evidence is collected. The chain of custody will provide a trail for the piece of evidence from its collection point to the time it is presented in court. The purpose of the chain of custody is to ensure the evidence has not been accessed by unauthorized individuals and has been handled in a tamper-proof manner [Mandia 03].

[39] For more information see "Best Evidence Rule." Wikipedia. 15 Nov. 2005. http://en.wikipedia.org/wiki/Best_evidence_rule.

It is also important to ensure the integrity of any evidence collected. It should be provable in court that the data being presented exactly matches the data that was collected. A common practice for ensuring evidence integrity is to generate cryptographic hash values of the data when it is being collected. Later, if the integrity of the data comes into question, one can refer to the hash values generated at the time of collection, compute the hash values of the data being presented, and concretely demonstrate that the data collected and the data presented in court are one and the same.



9.4.2.4 Business Considerations

The impact of an incident can range from harmless to catastrophic for an organization and its customers. As part of developing an incident response process, therefore, the organization will need to consider *in advance* various issues that may arise. One such issue is the tradeoff between availability and security.

Suppose a critical server within an organization requires 24/7 uptime. If an incident is detected on that server, the organization will need to decide how the response process will affect availability. This determination will depend on multiple factors. First, the organization must evaluate the importance of the asset in question. If the asset provides a critical function, response options may be limited to those that can ensure the asset's availability. Second, the security risk that the incident poses to the organization must be assessed. The potential for damage from a high-risk incident may outweigh the benefits of keeping a particular asset functioning. For example, if an unauthorized root account is detected on a critical server that is exposed to the outside world, the costs of temporarily taking the server offline may pale in comparison to the consequences of the entire network being compromised. Conversely, certain incident response processes, such as monitoring a hacker's activities, require that the asset be available. Again, in this type of situation the benefit of the response must be weighed against the potential damage that could be incurred by allowing the attacker to continue malicious activity. These issues demonstrate that there is no clear-cut correct decision and that any incident response team must consider the tradeoffs between security and availability. This requirement to consider tradeoffs is related to risk management and the identification of critical assets, which we discussed in detail in an earlier module of this

course. If this activity is performed before an incident ever occurs, the organization will be able to respond promptly and confidently when the inevitable happens.

Another issue that can greatly affect an organization is disclosure of an incident to customers and to the public. Historically, organizations have shied away from reporting incidents out of fear that they will lose their customers' trust. Indeed, attackers often extort money from businesses by taking advantage of this fear. It is not uncommon for an attacker to compromise a company's network, steal customer records (e.g., credit card numbers), and then threaten to publicize the breach unless he is paid a certain amount of money. Cyber extortion is outside the scope of this module; however, it attests to organizations' strong motivation to avoid incident disclosure due to the severe damage it can cause to their business and reputation.

However, it is important for organizations to prepare for incidents that may require some form of disclosure to their customers or the public. For example, if sensitive customer information, such as dates of birth or Social Security numbers, is stolen by a malicious individual, the organization likely will need to reveal the incident to its affected customers, especially in light of recently passed state laws that were discussed in the "Compliance Management" module. Preparing for the worst ahead of time can help an organization determine an appropriate balance between disclosure and secrecy to protect both its business interests and its customers.

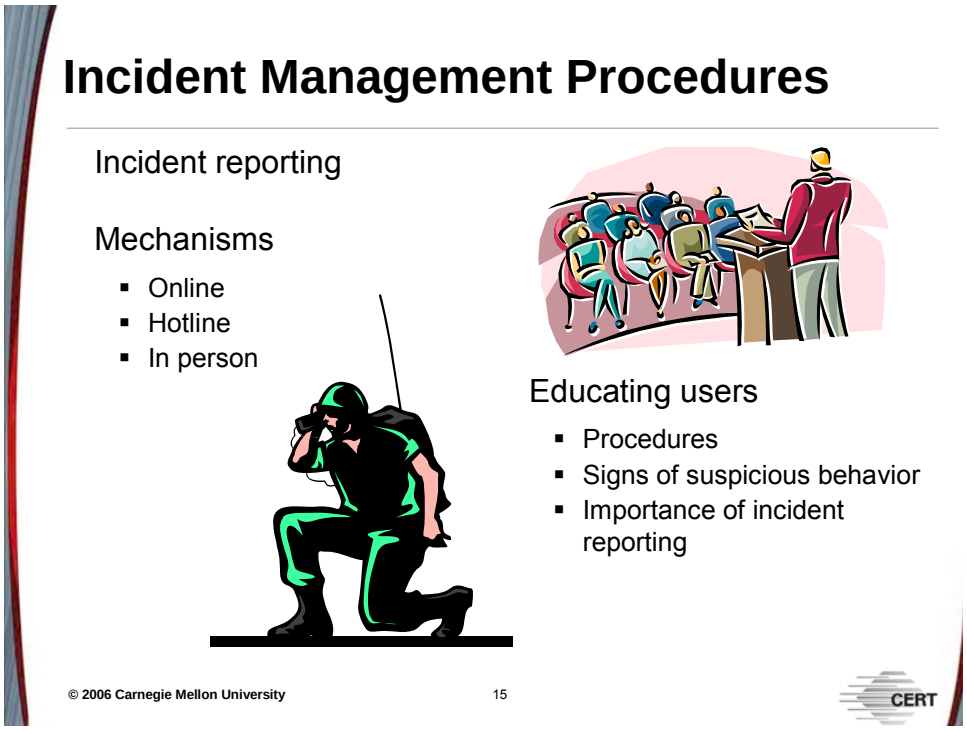
Disclosure of an incident can also be made to a third-party organization such as CERT/CC, which can help investigate the incident in a discreet way, or to law enforcement for a criminal investigation. For U.S. federal government agencies, this is important because incident reporting outside of the organization is a part of FIPS 200 compliance. FIPS 200 is a set of mandatory minimum standards for security, including incident response, with which all agencies must comply in the wake of FISMA's passage. These standards are very general; NIST Special Publication 800-53 outlines specific security controls to implement the minimum requirements summarized in FIPS 200. For more information about FIPS 200 and NIST SP 800-53, see http://csrc.nist.gov.

9.4.2.5 Critical Asset Considerations

The previous section, "Business Considerations," explained the tradeoff between availability and security during an incident. For certain key assets, this tradeoff is delicate and should not be made in the heat of the moment. Identifying these critical assets enables customized response plans to be formulated before an incident occurs. This includes deciding ahead of time on an acceptable level of availability for specific assets should they become involved in an incident.

For example, suppose a fictitious company called WeeStuff Inc. sells Weebles online. The WeeStuff Web servers can be identified as a critical asset, since the Web is WeeStuff's main source of revenue. After some forethought, WeeStuff determines that the availability of its

Web servers cannot be reduced, even in the event of an incident. However, the IT staff at WeeStuff realizes that keeping compromised Web servers online is a security risk. For example, transaction information may be tampered with, or customer information may be stolen. Therefore, WeeStuff determines that it will keep a full set of backup servers that can be immediately swapped with the current Web servers. This example illustrates how identification of critical assets enables forethought about how they will be handled during an incident.



9.4.3 Incident Management Procedures

9.4.3.1 Incident Reporting

When a computer security incident occurs, it will most likely be detected first by someone outside of the incident response team. Therefore, it is important to set up an incident reporting mechanism for everyone in the organization. Incident reporting can be handled online, in person, or over the phone. The actual means of incident reporting is not as important as how the reporting mechanism is designed. Since not all users will be technically savvy, it is critical to educate users on the signs of suspicious behavior as well as the procedures for reporting an incident. For example, incident reporting education could be as simple as developing posters to inform employees about the process and why it is important. Break areas, lunch rooms, and even bathrooms are excellent locations in which to hang posters.

Reporting mechanisms also help track incidents over time and in an organized manner. This is important for government agencies and is a part of FIPS 200 compliance.

What is a CSIRT?

An organization or team that provides services and support, to a defined constituency, for ***preventing***, ***handling*** and ***responding*** to computer security incidents

CERT

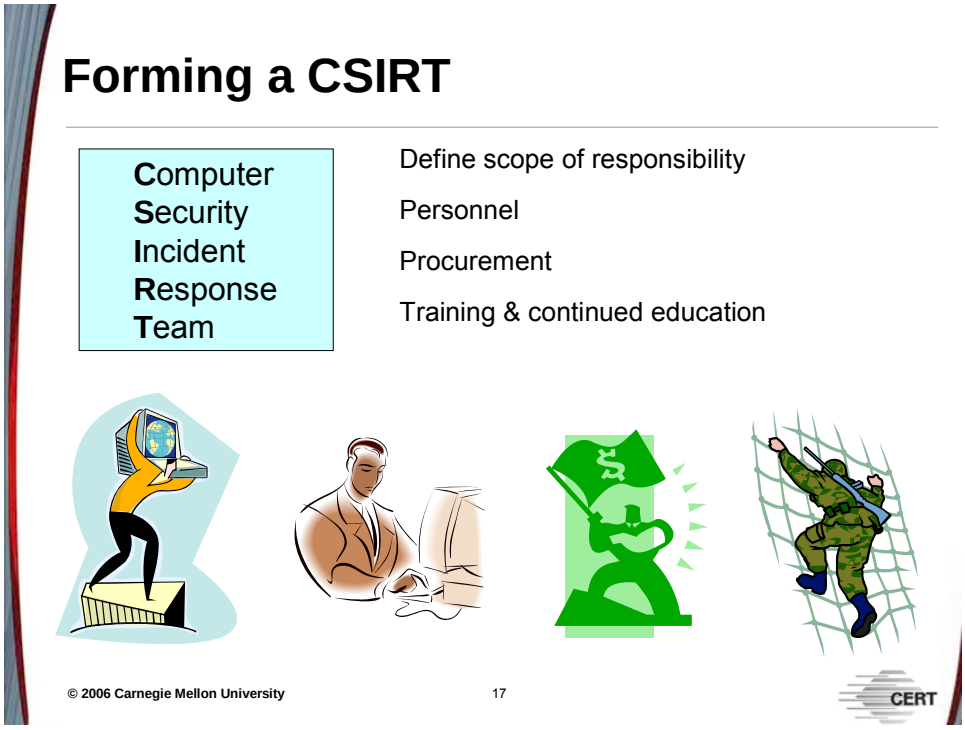
9.4.4 Forming a CSIRT

Forming and running a Computer Security Incident Response Team (CSIRT) is one way to approach computer security incident management. There are other ways to form an incident response team that may be equally valid; we provide information about a CSIRT in this module as an in-depth example.

In essence, a CSIRT is a concrete organizational entity (i.e., one or more staff) that is assigned the responsibility of providing part of the incident management capability for a particular organization. When a CSIRT exists in an organization, it is generally the focal point for coordinating and supporting incident response. By definition a CSIRT must perform at least incident handling activities.

CSIRT work is very similar to emergency response work in other sectors. Not only do you need to have the necessary tools and plans in place to respond effectively, but you also must perform other proactive functions to prevent disasters from happening, where possible.

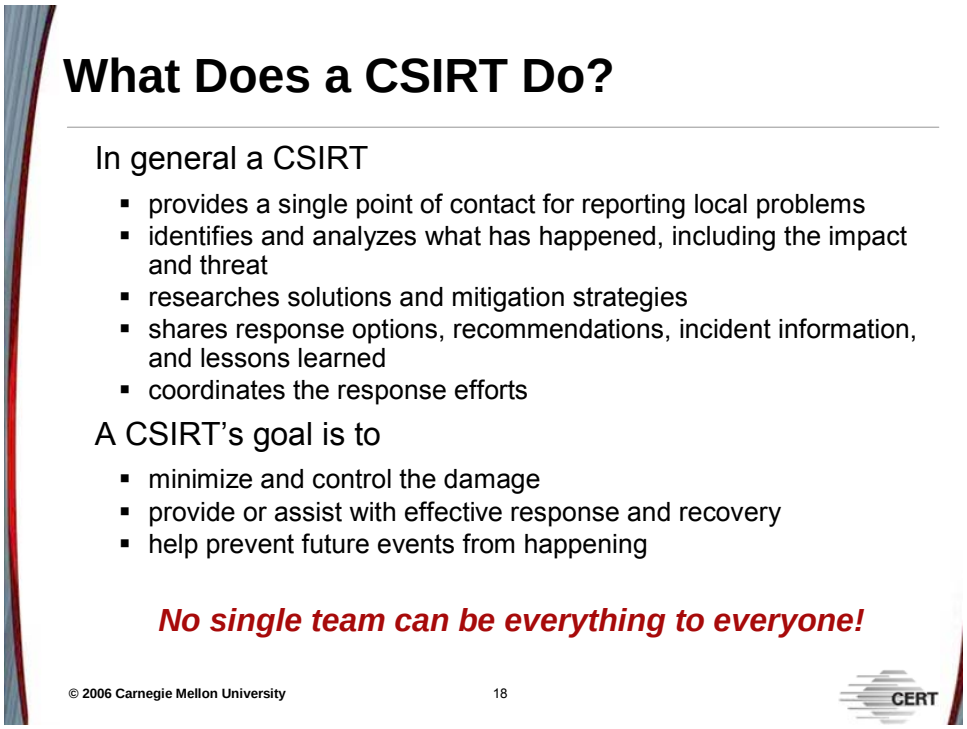
We'll discuss all of these roles of a CSIRT in the following slides.



When forming a CSIRT, an organization must follow four basic steps:

1. Decide what types of responsibilities the CSIRT will shoulder. (Define its mission.)
2. Choose personnel carefully to ensure their technical and "soft" skills match the requirements of the job.
3. Procure assets, as the incident response process often requires use of specialized software and dedicated equipment.
4. Ensure CSIRT staff stays up to date through continued education and training, since the field of incident response is constantly changing with the emergence of new technologies and security threats.

We'll examine each of these steps in detail next.



9.4.4.1 Defining the Scope of Responsibility

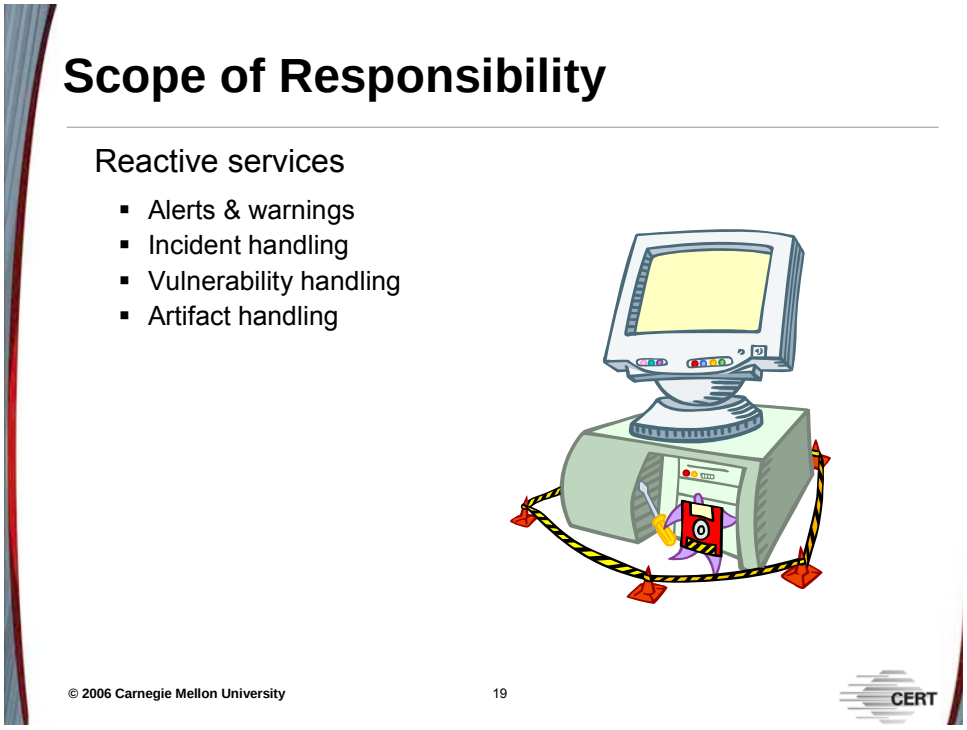
At a basic level, the goal of a CSIRT is to minimize and control damage stemming from computer security incidents, provide effective response and recovery, and work to prevent future events from happening. Often, the CSIRT is the group that coordinates incident response analysis and implementation.

Many organizations also assign the CSIRT additional responsibilities, such as protecting infrastructure, detecting security events, and conducting triage. It is up to each organization to determine which of the incident management processes (prepare, protect, detect, triage, respond) its CSIRT will perform.

The goals of a CSIRT must be based on the business goals of the constituent or parent organization and may depend on that organization's size, mission, and available resources. Protecting critical assets is key to the success of both an organization and its CSIRT.

Each CSIRT's unique mission, as defined by its parent organization, will determine the types of services it offers and its specific modes of interaction with others within the organization. Typically, however, CSIRT services can be broken down into three categories: reactive services, proactive services, and security quality management services [CERT 02].[40]

[40] See "CSIRT Services," 2002. http://www.cert.org/csirts/services.html.



Reactive Services

Reactive services encompass the procedures and responsibilities generally associated with incident response. These services involve responding to events and requests. An example of an event could be an alert or warning from an intrusion detection system. Other services that are considered reactive are the incident response process, vulnerability handling, and artifact handling. Vulnerability handling is an aspect of incident management that involves dealing with software or hardware flaws that could be exploited by attackers to cause computer security events. Artifact handling is an aspect of incident management that involves analyzing files or objects that might be involved in probing or attacking networked systems or that are being used to defeat security measures.[41]

These three services are similar in that they all involve the analysis, response, and mitigation of a particular security threat through reaction to something that is already there.

[41] See "CSIRT Services," 2002. http://www.cert.org/csirts/services.html.

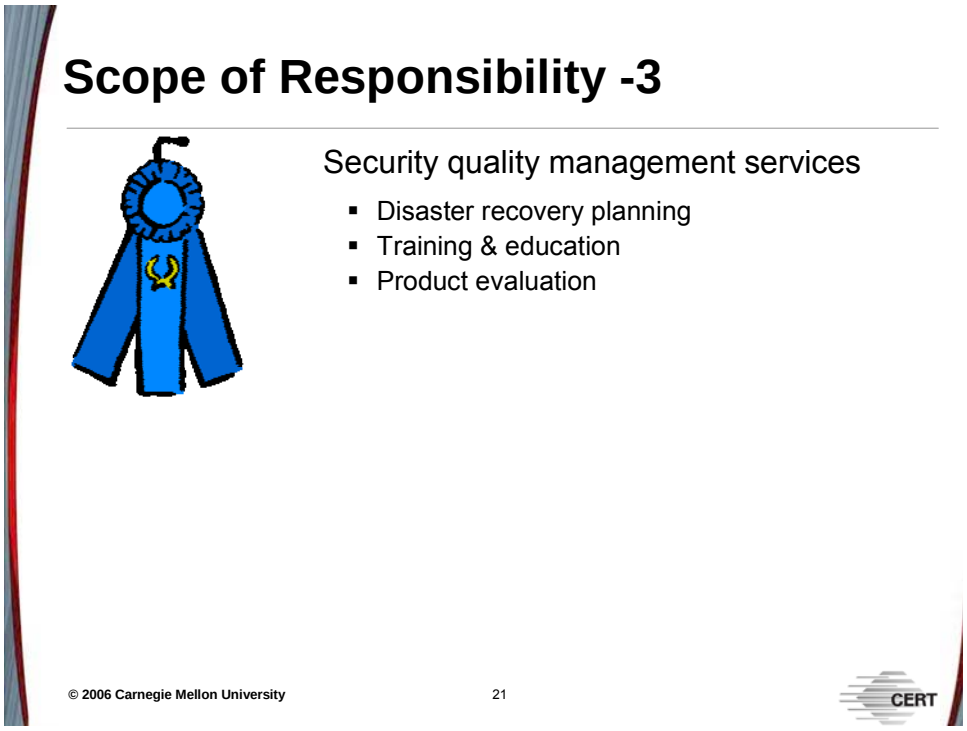
Proactive Services

The main goal of proactive services is to improve security within an organization to reduce the number of incidents that occur. Proactive services can be thought of as procedures and operations associated with information assurance, including but not limited to alerts and announcements, security audits, implementation of security tools and measures, and dissemination of security information.

Alerts and announcements can inform employees about immediate or near-term security threats and the release of new security advisories. Audits can ensure that security measures meet a minimum acceptable standard and may also include a review of best practices and policies. The implementation of security measures involves the installation, configuration, and maintenance of such tools as firewalls, authentication mechanisms, and intrusion detection systems. Dissemination of security information is intended to promote greater security awareness in the long term and to make available policies, security guidelines, and best practices to constituents.[42]

Again, it is important to keep in mind that the purpose of all of these services is to prevent incidents from occurring in the first place.

[42] See "CSIRT Services," 2002. http://www.cert.org/csirts/services.html.



Security Quality Management Services

The purpose of security quality management services is to improve the overall security posture of an organization at a high level and on a long-term basis. Such services can include disaster recovery planning, training and education, and product evaluation.[43] Disaster recovery is fairly self-explanatory and has been discussed in the Availability Management module. Training and education can be conducted through seminars, courses, or tutorials and are intended to help improve employees' understanding of security policies, procedures, and practices. This is especially vital for federal government agencies because providing training to employees is part of FIPS 200 compliance. Finally, product evaluation is a means of assessing applications and equipment to ensure they meet an acceptable level of security, which can help reduce the number of vulnerabilities to which an organization is exposed.

Proactive and security quality management services are not typically associated with incident management. However, these categories illustrate the potential cross-functional nature of a CSIRT. For some organizations, it may be more efficient to streamline all three categories of services into one group, since reactive services may not occupy all of the CSIRT's time. It is important to understand that the scope of responsibility of a CSIRT will vary for each organization and should be customized to suit the organization's specific needs.

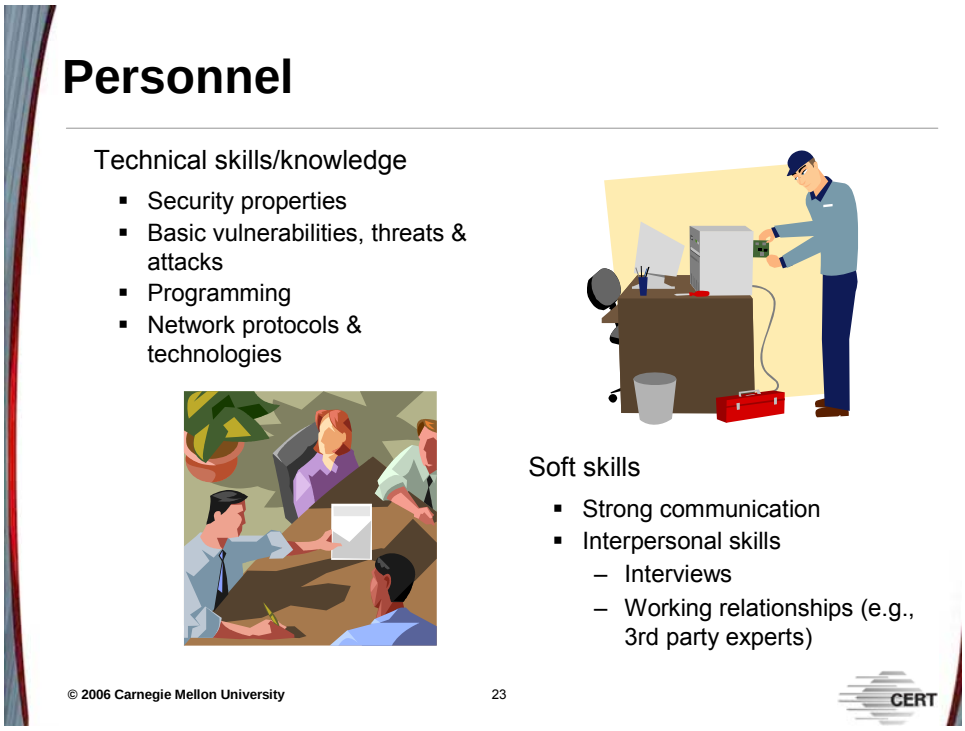
[43] See "CSIRT Services," 2002. http://www.cert.org/csirts/services.html.

Range of CSIRT Services

CSIRTs provide one or more of the following services:

CERT

This slide shows a more comprehensive list of the categories and services associated with each category of service.



9.4.4.2 Personnel

Once the scope of responsibility for the CSIRT has been defined, the next step is to find the right people to staff it. A CSIRT will obviously need personnel with the appropriate technical expertise, but it is also important for CSIRT members to possess strong "soft" skills such as oral communication, written communication, and sociability.

Technical Skills

The CSIRT as a whole should possess a wide range of technical skills and expertise. The skill sets sought will largely depend on the CSIRT's scope of responsibility. At a minimum, CSIRT members should have an understanding of basic vulnerabilities, threats, and attacks, including viruses, worms, denial-of-service, buffer overflows, spoofing, and man-in-the-middle attacks. Since so much of the digital world is connected through networks, team members also should have a broad understanding of various network technologies, protocols, and applications. Indeed, team members' knowledge of security issues should encompass both hosts and networks.

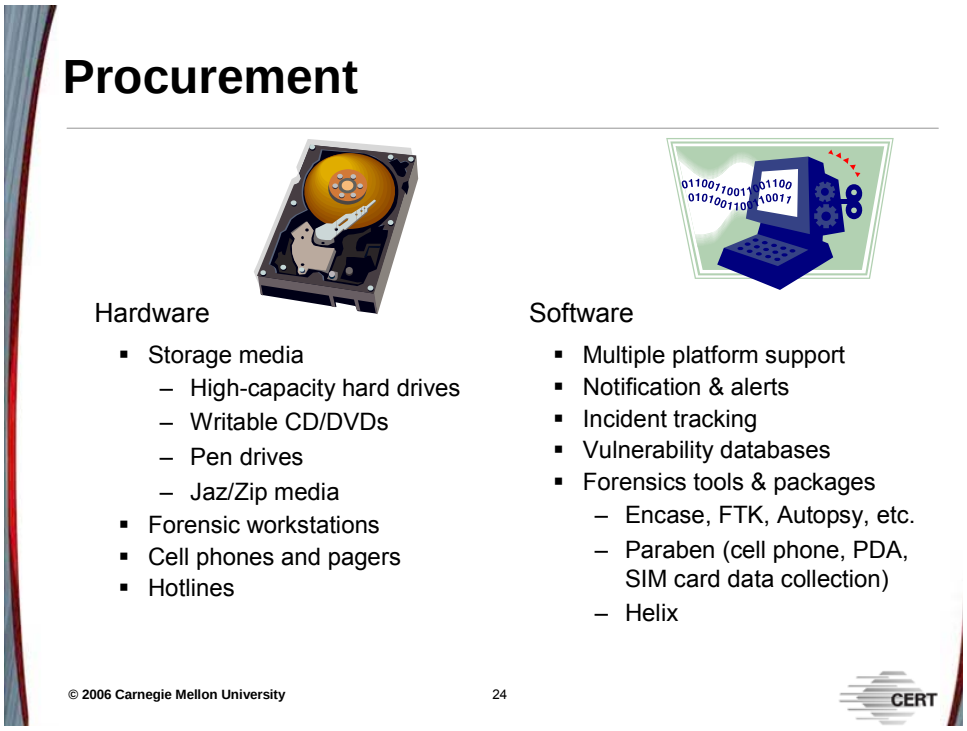
Some CSIRT team members should also possess programming skills so that they can understand source code when reviewing artifacts and exploits. It is vital to understand how operating system, network, and application software works in order to know how to defend it. Programming skills also can be useful for such tasks as creating tools or scripts. For example, many incident response processes can be automated and batched together through a script.

Lastly, all CSIRT members should be intimately familiar with the properties of confidentiality, integrity, availability, access control, authentication, and nonrepudiation.[44] No matter what type of services the CSIRT provides, these properties are universally applicable.

Soft Skills

Good oral communication, written communication, and interpersonal skills are just as important to a CSIRT as strong technical skills. This is because a large portion of the work that a CSIRT performs involves interacting with many different parties. Someone who is able to communicate effectively and promote a cohesive work environment will ultimately help the CSIRT provide high-quality, efficient services. For example, when an incident occurs, one of the tasks of the CSIRT is to gather more information by interviewing the involved parties. The ability to extract the maximum amount of relevant information from an interviewee can greatly accelerate the incident response process and lead to a quick resolution. Similarly, in the course of their work, CSIRT members likely will need to develop relationships with parties outside of the organization, such as third-party experts or upstream providers. CSIRT members who are able to establish and maintain good working relationships can add significant value to the team.

[44] See "Staffing Your Computer Security Incident Response Team – What Basic Skills Are Needed?" 2003. http://www.cert.org/csirts/csirt-staffing.html.



9.4.4.3 Procurement

For a CSIRT to be successful, it needs the appropriate resources to provide its services, including both hardware and software components. These may include incident tracking systems, notification and alerting systems, vulnerability tracking databases, test labs, cell phones, hotlines, and pagers, to name just a few. Procuring all of these resources is a daunting but necessary task.

In this section, we'll take a closer look at hardware and software components required for forensic analysis—just one service that may be provided by a CSIRT—but the important takeaway message is that this level of consideration should be given to components required for *all* incident management processes to be conducted by the CSIRT. *All* necessary equipment should be identified and procured in advance. The last thing you want to experience is a failed response due to a preventable lack of resources.

That said, let's delve into an analysis of required forensic hardware and software components. These components compose the CSIRT's trusted resource kit. Such a kit is important because when the CSIRT collects data from an incident, it must assume it is dealing with a hostile environment and can trust only its own hardware and software tools.

Hardware Components

The type of hardware that should be most abundantly available to a CSIRT forensics operation is storage media. This is because at least two data copies will need to be made during the collection process: an "original" copy and a working copy. Also, it is not unusual

for more than one working copy to be made so that analysis can be performed by more than one individual. With that said, storage media can include high-capacity hard drives, writable CDs and DVDs, USB drives, and other media types. It is also important that peripherals for these types of media, including ribbon cables and labels for clearly marking used media, be available as well. Computer systems should be allocated or purchased for the exclusive purpose of forensic duplication, data collection, and analysis. There are actually some companies, such as Digital Intelligence, Inc.[45], that create workstations and toolkits specifically for forensic use.

Software Components

A CSIRT should possess the software to support multiple environments and operating systems. Some noteworthy forensic toolkits include Encase by Guidance Software, FTK (Forensic Toolkit) by AccessData, SMART by ASR Data, and Autopsy. With the exception of Autopsy, all of these forensic packages are commercial software. Also, Paraben offers an array of forensic tools that are useful for collecting data from cell phones, PDAs, and SIM cards [Paraben 06].[46] Lastly, Helix is a bootable live CD that is freely available for download and contains many applications dedicated to incident response. It has been designed not to modify the host computer in any way, thereby ensuring that it remains forensically sound.[47]

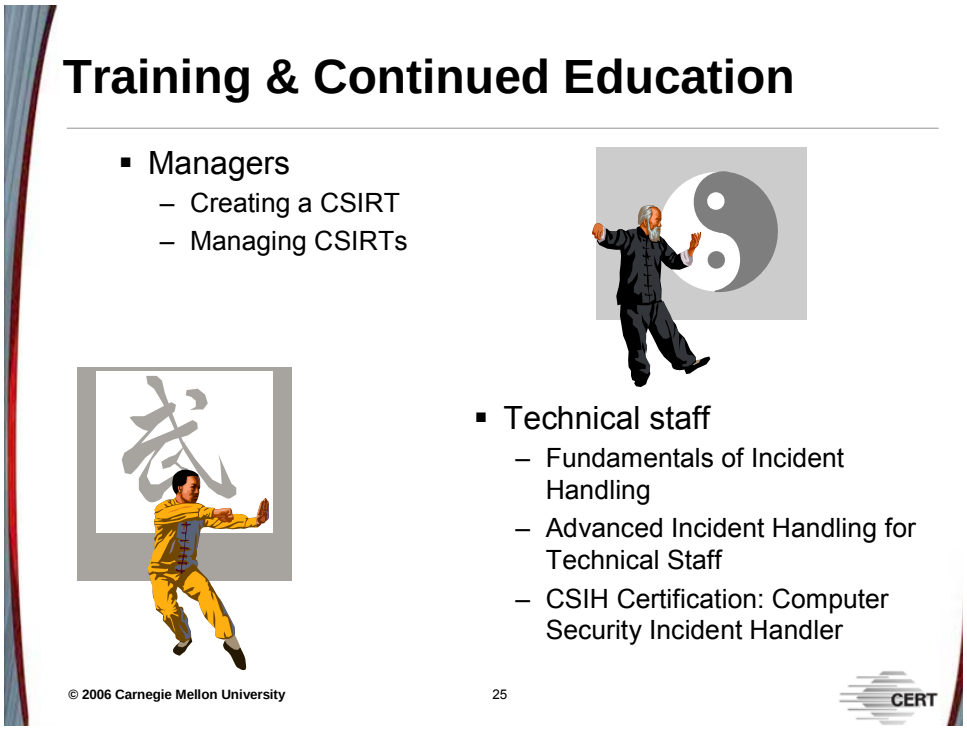
This section is intended to give you some general background on the types of toolkits available. The packages that best suit your organization will largely depend on its needs.

---

45 See http://www.digitalintelligence.com/.

46 See "Handheld Digital Forensics." Paraben Corporation. 17 Nov. 2005. http://www.paraben-forensics.com/handheld_forensics.html.

47 See "The Helix Live CD Page." e-fense, Inc. 17 Nov. 2005. http://www.e-fense.com/helix/.



9.4.4.4 Training and Continued Education

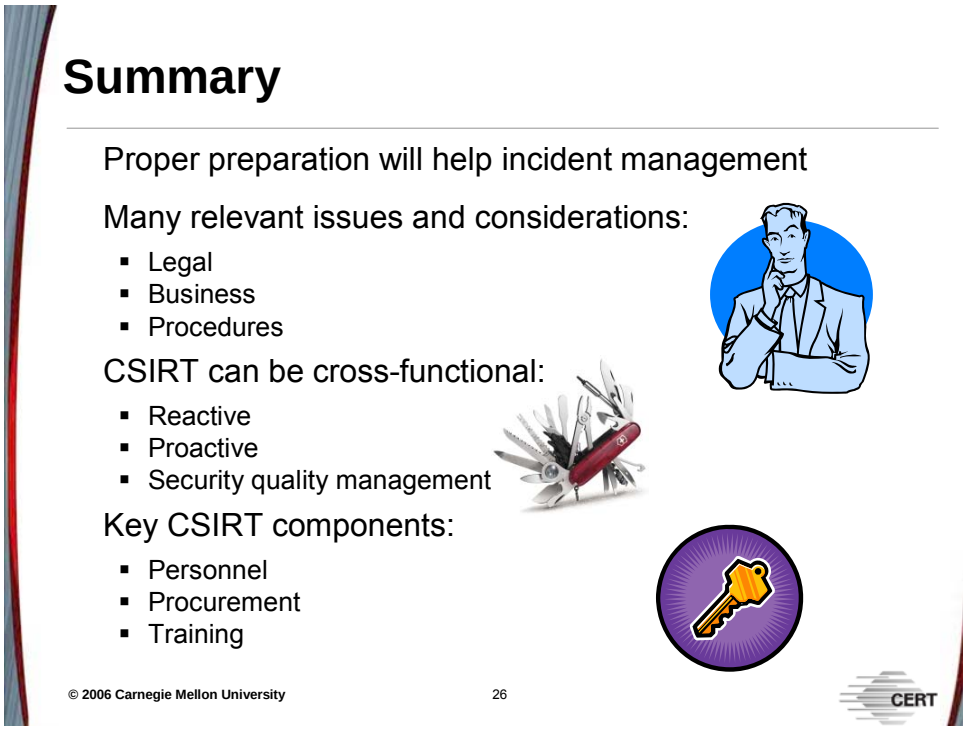
Even highly skilled CSIRT members do not permanently possess the skills to provide adequate services. Trends, techniques, methodologies, attacks, and threat environments change over time. Therefore, it is important for both technical and managerial staff of any CSIRT to improve their knowledge and skills through continuing education. This practice will ensure that technical staff members remain at the forefront of their field and that managers stay up to date with the latest trends, practices, and issues, so they can continue to steer the CSIRT in the right direction.

While this module is intended to provide a backbone of knowledge for incident management, there are a number of other courses offered by CERT that also can serve as continuing education for managers and technical staff members.

For managers, CERT offers two courses: *Creating a Computer Security Incident Response Team* and *Managing Computer Security Incident Response Teams*. Likewise, CERT offers the following courses geared toward technical staff members: *Fundamentals of Incident Handling* and *Advanced Incident Handling for Technical Staff*. For more information regarding these classes, refer to http://www.cert.org/nav/training.html. In addition, CERT offers a Computer Security Incident Handler (CSIH) certification program, which encompasses some of the training courses previously mentioned for technical staff members. More information about the CSIH program can be found at http://www.cert.org/certification/. A benefit of receiving training from CERT is that it strives to remain at the forefront of the information security field and pass that expertise on to others.

In addition to continuing education, CSIRT members can find real benefit in conducting internal training exercises within their organization. This type of training can be especially useful for the complex processes and procedures of incident response, by keeping staff members' skills sharp and enabling them to be better prepared when they need to put those skills to use. Some CSIRTs go as far as creating mock incidents and stepping through the entire incident response process from start to finish. Such training exercises can help the CSIRT work out kinks and refine its processes to maximize effectiveness when a real incident occurs.

Another training approach that may be helpful is a mentoring program for new CSIRT staff in which new staff members are assigned to an experienced staff member who can help guide them through what they need to know: the media policy, the information disclosure policy, standards for secure information handling, and so on.



Summary

This module has discussed the benefits of incident management and laid a basic knowledge foundation for implementing an incident response process as a component of Defense-in-Depth.

We have seen that a big-picture view of incident management enables you to understand the effects an incident may have on your organization and formulate an appropriate response to mitigate this risk and recover quickly if an incident does occur.

The main portion of the module focused on the development of an incident response process as briefly summarized below.

1. Before developing an incident response process, it is important to ensure the existence of certain security implementations, policies, and procedures. These preparations can reduce the number of incidents that occur and contribute to the efficiency of the incident response process.

2. Legal and business considerations must be taken into account to form highly tailored response programs for individual organizations.

3. Despite this customization, there are still a number of procedures such as reporting mechanisms and evidence preservation that must be followed by every incident response team.

4. A CSIRT is one way or organizing an incident response team to carry out the incident response process. CSIRT services are not limited to traditional, reactive incident response procedures. A CSIRT can also provide proactive and security quality management services. However, in order for a CSIRT to provide high-quality services, it is important that the right personnel are hired, the appropriate assets are acquired, and adequate training is made available to team members.

Review Questions -1

1. An incident is an _____, _____, or _____ that reduces the availability, confidentiality, and integrity of assets.
2. Name two benefits of incident management.
3. In what part of the incident response process does data collection occur?
4. What are three legal considerations related to developing an incident response process?
5. What is the potential tradeoff between an asset's availability and security during an incident?

Review Questions -2

6. What is the “Best Evidence Rule”?
7. What are the three types of services that a CSIRT can offer and what is their purpose?
8. Name two reasons why “soft skills” are just as important as technical skills for CSIRT personnel.
9. Why is it important for a CSIRT to have resources abundantly available?
10. What does a CSIRT achieve by conducting internal training exercises?

References/Bibliography

All urls valid at the publication date of this document.

[Alberts 01] Alberts, C. & Dorofee, A. “An Introduction to the OCTAVE Method.” Pittsburgh, PA: Software Engineering Institute, Carnegie Mellon University. http://www.cert.org/octave /methodintro.html (2001).

[Alberts 04] Alberts, C.; Dorofee, A.; Killcrece, G.; Ruefle. R. & Zajicek, M. *Defining Incident Management Processes for CSIRTs: A Work in Progress* (CMU/SEI-2004-TR-015). Pittsburgh, PA: Software Engineering Institute, Carnegie Mellon University, 2004. http://www.sei.cmu.edu/publications/documents /04.reports/04tr015.html.

[Allen 01] Allen, J. “CERT System and Network Security Practices” *news@sei,* 2nd quarter, 2001. http://www.sei.cmu.edu/news-at-sei/columns/security_matters/2001/2q01/security-2q01.htm.

[Allen 03] Allen, J., Gabbard, D., & May, C. *Outsourcing Managed Security Service* (CMU/SEI-SIM-012). Pittsburgh, PA: Software Engineering Institute, Carnegie Mellon University, 2003. http://www.sei.cmu.edu/publications/documents/sims /sim012.html.

[Allen 01] Allen, Stewart. *Importance of Understanding Logs from an Information Security Standpoint.* Bethesda, MD: SANS Institute, 2001. http://www.sans.org/reading_room /whitepapers /logging/200.php?portal=92d3b0ef42396d1bbf5c7fa6c4c0f17a.

[ATIS 00] ATIS. “ATIS Telecomm Glossary.” 2000. http://www.atis.org/tg2k/.

[BASE 04] The BASE (Basic Analysis and Security Engine) Project Team. “About Basic Analysis and Security Engine (BASE).” http://base.secureideas.net/about.php (2004).

[Bastille 06] Bastille Linux. “Bastille Hardening Program: increased security for your OS.” 2006. http://www.bastille-linux.org/.

[Caffrey 04] Caffrey, Andrew. “Bank of America Fined Record $10m.” *The Boston Globe*, March 11, 2004. Available through http://www.boston.com/globe/search/.

[CASB 02] California Senate Bill (CASB) 1386, Chapter 915. *California Security Breach Information Act of 2002*. http://www.securitymanagement.com/library /SB1386_ca0203.pdf.

[CERT 02] CERT. “CSIRT Services.” Pittsburgh, PA: Software Engineering Institute, Carnegie Mellon University, 2002. http://www.cert.org/csirts/services.html.

[CERT 03] CERT.“Staffing Your Computer Security Incident Response Team – What Basic Skills Are Needed?” Pittsburgh, PA: Software Engineering Institute, Carnegie Mellon University, 2003. http://www.cert.org/csirts/csirt-staffing.html.

[CISCO 06] Cisco. “Cisco Catalyst 5500 Switch” (product no longer available) 2006. http://www.cisco.com/en/US/products/hw /switches/ps686/ps687/index.html.

[DI 06] Digital Intelligence. http://www.digitalintelligence.com /forensichardware.php (2006).

[Eaton 02] Eaton, Ian. “The Ins and Outs of System Logging Using Syslog.” Bethesda, MD: SANS Institute, 2002. http://www.sans.org/reading_room/whitepapers/logging /1168.php?portal=a280db7784ccb20e0c23a53465c4df85.

[e-fense 05] e-fense Corporation. “The Helix Live CD Page.” 2005. http://www.e-fense.com/helix/.

[FPIC 05] Federal Partnership for Interoperable Communications (FPIC). “Response to Notice and Request for Comments on Draft implementation Guidance for Homeland Security Presidential Directive 12.” http://whitehouse.fed.us/omb/inforeg/hspd12/13.pdf (2005).

[FTC 99] Federal Trade Commission. *Gramm-Leach-Bliley Act: Disclosure of Nonpublic Personal Information.* 15 USC, Subchapter I, Sec. 6801-6809 (1999).

[Glorioso 05] Glorioso, R. *Assured Availability for the Digital Nervous System.* All Computer Solutions, courtesy of Marathon Technologies, 2005. http://www.disastertolerance.com/aawhitepaper.htm.

[GFi 06] GFi Software, Ltd. “GFi LANguard System Integrity Monitor.” 2006. http://www.gfi.hk/lansim/.

[Guidance 05] Guidance Software. “EnCase® Enterprise Detailed Product Description.” http://www.guidancesoftware.com/corporate /downloads /whitepapers/EEDetailedProductDescription04-25-2005.pdf (2005).

[HRG 06] Harvard Research Group. http://www.hrgresearch.com/ (2006).

[IBM 06] International Business Machines. “(Tavoli) Availability.” 2006. http://www-306.ibm.com/software/tivoli/solutions /availability/.

[IIA 06] Institute of Internal Auditors. “IT Audit.” http://www.theiia.org /itaudit/index.cfm?fuseaction=print&fid=5569 (2006).

[IEEE 90] Institute of Electrical and Electronics Engineers. *IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries.* New York, NY: IEEE, 1990.

[InterSect 06] InterSect Alliance. *The SNARE Toolset – A White Paper*. http://www.intersectalliance.com/resources /Documentation /Snare_Toolset_White_Paper-2.4.pdf (2006).

[ISECOM] Institute for Security and Open Methodologies. *Open Source Security Testing Methodology Manual* (OSSTMM). http://www.isecom.org/osstmm/ (2006).

[ISO 05] International Organization for Standardization. *ISO/IEC 17799:2005 Information technology - Security techniques - Code of practice for information security management.* 2005. http://www.iso.org/iso/en/prods-services/popstds /informationsecurity.html.

[Kenning 01] Kenning, M. J. “Security Management Standard — ISO 17799/BS 7799.” *BT Technology Journal 19*, 3 (July 2001): 132-136. http://www.springerlink.com/content /q7351214t803384x/fulltext.pdf.

[Kohl 93] Kohl, J. & Neuman, C. *The Kerberos Network Authentication Service (V5*) (RFC 1510). IETF, The Internet Society, February 1993. http://ietfreport.isoc.org/idref/rfc1510/. Obsoleted by RFC 4120, *The Kerberos Network Authentication Service (V5).* 2005. http://ietfreport.isoc.org/idref /rfc4120/.

[Krutz 01] Krutz, Ronald L. & Vines, Russell D. *The CISSP Prep Guide*. New York, NY: Wiley Computer Publishing, 2001.

[LAP 05] Liberty Alliance Project. “Specifications.” http://www.projectliberty.org/resources/specifications.php (2005).

[Linux 03] Linux. “Linux Advanced Routing and Traffic Control.” 2004. http://lartc.org/.

[LogLogic 05] LogLogic. *Best Practices for Log Management – Industry Organizations and Regulators Define How to Manage Log Data*. San Jose, CA: LogLogic, 2005. http://loglogic.com/documents/white-papers /Log%20Management%20Defined%20Jan05Final.pdf.

[Lonvick 01] Lonvick, C. *The BSD Syslog Protocol* (RFC 3164). IETF, The Internet Society, February 2001. http://ietfreport.isoc.org/idref/rfc3164/.

[MACE 06] Middleware Architecture Committee for Education (MACE). “About Shibboleth.” http://shibboleth.internet2.edu /about.html (2006).

[Mandia 03] Mandia, Kevin; Prosise, Chris; & Pepe, Matt. *Incident Response*, 2nd ed. Emeryville, CA: McGraw-Hill Osborne Media, 2003. (ISBN: 0-072-22696-X).

[Marathon 06] Marathon Technologie Corporation. 2006. http://www.marathontechnologies.com/.

[Markle 06] The Markle Foundation. *The Legal Realities of Logs.* New York, NY: The Markle Foundation, 2006. http://www.markle.org/downloadable_assets /nstf_IAL_020906.pdf.

[McAlpine 98] McAlpine, Fraser A. & Droke, Michael. “Electronic Privacy In Employment,” 1998. http://library.findlaw.com/1998/Jan/1/126935.html.

[Microsoft 05] Microsoft Corporation. “Microsoft’s Vision for an Identity Metasystem,” 2005. http://msdn.microsoft.com/library/default.asp?url=/library /en-us/dnwebsrv/html/identitymetasystem.asp.

[Microsoft 06] Microsoft Corporation. “Windows Server Update Services.” 2006. http://www.microsoft.com/windowsserversystem /updateservices/default.mspx.

[Nagios 06] Nagios. “About Nagios.”1999-2006. http://www.nagios.org/about.

[NSC 06] The Naval Safety Center. “Operational Risk Management.” http://www.safetycenter.navy.mil/orm/default.htm (2006).

[Nawyn 03] Nawyn, Kenneth E. *A Security Analysis of System Event Logging with Syslog*. Bethesda, MD: SANS Institute, 2003. http://www.sans.org/reading_room/whitepapers/logging/1101.php?portal=1deeb14e268cd2a8fb75e0e5019ce34d.

[Newton 06] Newton, Harry. *Newton's Telecom Dictionary*, 22nd ed. San Francisco, CA: CMP Books, 2006. (ISBN 1-578-20319-8).

[NIST 02] National Institute of Standards and Technology (NIST). *Federal Information Security Management Act of 2002*. Gaithersburg, MD: Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology, 2002. http://csrc.nist.gov/policies/FISMA-final.pdf.

[NIST 05] National Institute of Standards and Technology (NIST). *Recommended Security Controls for Federal Information Systems* (SP 800-53). Gaithersburg, MD: Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology, 2005. http://csrc.nist.gov/publications/nistpubs/800-53/SP800-53.pdf.

[NIST 06] National Institute of Standards and Technology (NIST). *Personal Identity Verification (PIV) of Federal Employees and Contractors (FIPS 201-1).* Gaithersburg, MD: Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology, 2006. http://csrc.nist.gov/publications/fips/fips201-1/FIPS-201-1-chng1.pdf.

[ntop 06] ntop.org. http://www.ntop.org/ntop.html (1998-2006).

[Oetiker 06] Oetiker, T. “Tobi Oetiker’s MRTG – The Multi-Router Traffic Grapher.” 2006. http://oss.oetiker.ch/mrtg/.

[Packeteer 06] Packeteer. http://www.packeteer.com/ (2006).

[Paraben -06] Paraben Corporation “Handheld Digital Forensics.” 2006. http://www.paraben-forensics.com/handheld_forensics.html.

[RAID 04] RAID. “Data Recovery Clinic.” 2004. http://www.datarecoveryclinic.com/raid_data_recovery.htm.

[Red Hat 03] Red Hat, Inc. “Red Hat Linus 9: Red Hat Linus Reference Guide.” 2003. https://www.redhat.com/docs/manuals/linux /RHL-9-Manual /ref-guide/ch-tripwire.html.

[Relex 01] Relex Software “MTBF and MTTF Calculation.” 2001. http://www.i-mtbf.com/.

[Russell 91] Russell, D. & Gangemi, G.T. *Computer Security Basics.* Sebastopol, CA: O'Reilly Media, 1991. (ISBN 0-937-17571-4).

[SANS 06] The SANS Institute. “Intrusion Detection: What is Host-Based Intrusion Detection?” http://www.sans.org/resources/idfaq /host_based.php (2000-2006).

[Schmoo 05] The Shmoo Group “Osiris User Handbook.” http://osiris.shmoo.com/handbook.html (2005).

[Sourcefire 06] Sourcefire, Inc. “snort.org.” http://www.snort.org/ (2006).

[Squid 06] Squid. “Team Squid.” http://www.squid-cache.org/ (2006).

[Summers 97] Summers, R. *Secure Computing: Threats and Safeguards*. New York, NY: McGraw-Hill, 1997. (ISBN: 0-070-69419-2).

[USPL 96] U.S. Public Law (USPL). *Health Insurance Portability and Accountability Act (HIPAA) of 1996.* http://aspe.hhs.gov/admnsimp/pl104191.htm (1996).

[USPL 02] U.S. Public Law (USPL). *Sarbanes-Oxley Act of 2002.* http://www.law.uc.edu/CCL/SOact/soact.pdf (2002).

[USC 99] United States Congress. *Gramm-Leach-Bliley Act of 1999.* http://www.ftc.gov/privacy/glbact/glbsub1.htm (1999).

[Venema 92] Venema, W. “TCP Wrapper: Network Monitoring, Access Control, and Booby Traps,” 85-92. *Proceedings of the 3rd Unix Security Symposium*. Baltimore, MD, Sept. 14-17, 1992. Berkeley, CA: USENIX, 1992. Available through http://www.usenix.org/publications/library/proceedings/.

[Visscher 06] Visscher, Bamm. “Sguil -The Analyst Console for Network Security Monitoring.” http://sguil.sourceforge.net/ (2006).

[Webopedia 06] Webopedia. “Mesh.” http://www.webopedia.com/TERM/M /mesh.html (2006).

[Wiki 06] Wikipedia, the Free Encyclopedia (Wiki). “MD5 (Message-Digest Algorithm 5).” http://en.wikipedia.org/wiki/MD5 (2006).

[Wiki 06] Wikipedia, the Free Encyclopedia (Wiki). “Best Evidence Rule.” http://en.wikipedia.org/wiki/Best_evidence_rule (2006).

[Wiki 06] Wikipedia, the Free Encyclopedia (Wiki). “Password.” http://en.wikipedia.org/wiki/Password (2006).

[Wiki 06] Wikipedia, the Free Encyclopedia (Wiki). “Sarbanes-Oxley Act.” http://en.wikipedia.org/wiki/Sarbanes-Oxley_Act (2006).

[Wiki 06] Wikipedia, the Free Encyclopedia (Wiki). “Sarbanes-Oxley and its Impact on IT Controls.” http://en.wikipedia.org /wiki/Information_Technology_Controls #Sarbanes-Oxley_and_its_impact_on_IT_Controls (2006).

[Wiki 06] Wikipedia, the Free Encyclopedia (Wiki). “SHA Hash Functions.” http://en.wikipedia.org/wiki/SHA-1 (2006).

[Wiki 06] Wikipedia, the Free Encyclopedia (Wiki). “Windows CardSpace.” http://en.wikipedia.org/wiki/Infocard (2006).

[Wireshark 06] Wireshark (formerly Ethereal) 2006. http://www.wireshark.org/.

[Wotring 04] Wotring, Brian. “Host Integrity Monitoring: Best Practices for Deployment.” http://www.securityfocus.com/infocus/1771 (2004).

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
(Leave Blank)	September 2006	Final

4. TITLE AND SUBTITLE	5. FUNDING NUMBERS
Defense in Depth: Foundation for Secure and Resilient IT Enterprises	FA8721-05-C-0003

6. AUTHOR(S)

Christopher J. May, Josh Hammerstein, Jeff Mattson, Kristopher Rush

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Software Engineering Institute Carnegie Mellon University Pittsburgh, PA 15213	CMU/SEI-2006-HB-003

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSORING/MONITORING AGENCY REPORT NUMBER
HQ ESC/XPK 5 Eglin Street Hanscom AFB, MA 01731-2116	

11. SUPPLEMENTARY NOTES

12A DISTRIBUTION/AVAILABILITY STATEMENT	12B DISTRIBUTION CODE
Unclassified/Unlimited, DTIC, NTIS	

13. ABSTRACT (MAXIMUM 200 WORDS)

The Defense-in-Depth Foundational Curriculum is designed for students ranging from system administrators to CIOs who have some technical understanding of information systems and want to delve into how technical assurance issues affect their entire organizations. The course material takes a big-picture view while also reinforcing concepts presented with some details about implementation. Therefore, this course can be a useful pursuit for system administrators and IT security personnel who would like to step up to the management level. It also can provide a refresher for IT managers and executives who want to stay up to date on the latest technological threats facing their enterprises.

The curriculum consists of eight main modules: (1) Compliance Management, (2) Risk Management, (3) Identity Management, (4) Authorization Management, (5) Accountability Management, (6) Availability Management, (7) Configuration Management, and (8) Incident Management. The document also contains an introduction, “Foundations of Information Assurance,” which focuses on how the overarching concepts of confidentiality, integrity, and availability can lead to a comprehensive security strategy.

14. SUBJECT TERMS	15. NUMBER OF PAGES
defense-in-depth, compliance management, identity management, authorization management, accountability management, availability management, configuration management, incident management	368

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL