capability, contacts will have been established with counterparts outside the organization. This allows for *early warning* of threats and vulnerabilities that the organization may have not yet experienced. Early preventative measures (generally more cost-effective than repairing damage) can then be taken to reduce future losses. Data is also shared outside the organization to allow others to learn from the organization's experiences.

12.1.3 Side Benefits

Finally, establishing an incident handling capability helps an organization in perhaps unanticipated ways. Three are discussed here.

Uses of Threat and Vulnerability Data: Incident handling can greatly enhance the risk assessment process. An incident handling capability will allow organizations to collect threat data that may be useful in their risk assessment and safeguard selection processes (e.g., in designing new systems). Incidents can be logged and analyzed to determine whether there is a recurring problem (or if other patterns are present, as are sometimes seen in hacker attacks), which would not be noticed if each incident were only viewed in isolation. Statistics on the numbers and types of incidents in the organization can be used in the risk assessment process as an indication of vulnerabilities and threats.[94]

Enhancing Internal Communications and Organization Preparedness. Organizations often find that an incident handling capability enhances internal communications and the readiness of the organization to respond to any type of incident, not just computer security incidents. Internal communications will be improved; management will be better organized to receive communications; and contacts within public affairs, legal staff, law enforcement, and other groups will have been preestablished. The structure set up for reporting incidents can also be used for other purposes.

Enhancing the Training and Awareness Program. The organization's training process can also benefit from incident handling experiences. Based on incidents reported, training personnel will have a better understanding of users' knowledge of security issues. Trainers can use actual incidents to vividly illustrate the importance of computer security. Training that is based on current threats and controls recommended by incident handling staff provides users with information more specifically directed to their current needs – thereby reducing the risks to the organization from incidents.

[94] It is important, however, *not* to assume that since only *n* reports were made, that *n* is the total number of incidents; it is not likely that all incidents will be reported.

12.2 Characteristics of a Successful Incident Handling Capability

A successful incident handling capability has several core characteristics:

- an understanding of the constituency it will serve;
- an educated constituency;
- a means for centralized communications;
- expertise in the requisite technologies; and
- links to other groups to assist in incident handling (as needed).

12.2.1 Defining the Constituency to Be Served

The constituency includes computer users and program managers. Like any other customer-vendor relationship, the constituency will tend to take advantage of the capability if the services rendered are valuable.

The constituency is not always the entire organization. For example, an organization may use several types of computers and networks but may decide that its incident handling capability is cost-justified only for its personal computer users. In doing so, the organization may have determined that computer viruses pose a much larger risk than other malicious technical threats on other platforms. Or, a large organization composed of several sites may decide that current computer security efforts at some sites do not require an incident handling capability, whereas other sites do (perhaps because of the criticality of processing).

> The focus of a computer security incident handling capability may be external as well as internal. An incident that affects an organization may also affect its trading partners, contractors, or clients. In addition, an organization's computer security incident handling capability may be able to help other organizations and, therefore, help protect the community as a whole.

12.2.2 Educated Constituency

Users need to know about, accept, and trust the incident handling capability or it will not be used. Through training and awareness programs, users can become knowledgeable about the existence of the capability and how to recognize and report incidents. Users trust in the value of the service will build with

> Managers need to know details about incidents, including who discovered them and how, so that they can prevent similar incidents in the future. However users will not be forthcoming if they fear reprisal or that they will become scapegoats. Organizations may need to offer incentives to employees for reporting incidents and offer guarantees against reprisal or other adverse actions. It may also be useful to consider anonymous reporting.

reliable performance.

12.2.3 Centralized Reporting and Communications

Successful incident handling requires that users be able to report incidents to the incident handling team in a convenient, straightforward fashion; this is referred to as *centralized reporting*. A successful incident handling capability depends on timely reporting. If it is difficult or time consuming to report incidents, the incident handling capability may not be fully used. Usually, some form of a hotline, backed up by pagers, works well.

Centralized communications is very useful for accessing or distributing information relevant to the incident handling effort. For example, if users are linked together via a network, the incident handling capability can then use the network to send out timely announcements and other information. Users can take advantage of the network to retrieve security information stored on servers and communicate with the incident response team via e-mail.

12.2.4 Technical Platform and Communications Expertise

The technical staff members who comprise the incident handling capability need specific knowledge, skills, and abilities. Desirable qualifications for technical staff members may include the ability to:

- work expertly with some or all of the constituency's core technology;
- work in a group environment;
- communicate effectively with different types of users, who will range from system administrators to unskilled users to management to law-enforcement officials;
- be on-call 24 hours as needed; and
- travel on short notice (of course, this depends upon the physical location of the constituency to be served).

12.2.5 Liaison With Other Organizations

Due to increasing computer connectivity, intruder activity on networks can affect many organizations, sometimes including those in foreign countries. Therefore, an organization's incident handling team may need to work with other teams or security groups to effectively handle incidents that range beyond its constituency. Additionally, the team may need to pool its knowledge with other teams at various times. Thus, it is vital to the success of an incident handling capability that it establish ties and contacts with other related counterparts and supporting organizations.

Especially important to incident handling are contacts with investigative agencies, such as federal (e.g., the FBI), state, and local law enforcement. Laws that affect computer crime vary among localities and states, and some actions may be state (but not federal) crimes. It is important for teams to be familiar with current laws and to have established contacts within law enforcement and investigative agencies.

Incidents can also garner much media attention and can reflect quite negatively on an organization's image. An incident handling capability may need to work closely with the organization's public affairs office, which is trained in dealing with the news media. In presenting information to the press, it is important that (1) attackers are not given information that would place the organization at greater risk and (2) potential legal evidence is properly protected.

The Forum of
Incident Response and Security Teams

The 1988 Internet worm incident highlighted the need for better methods for responding to and sharing information about incidents. It was also clear that any single team or "hot line" would simply be overwhelmed. Out of this was born the concept of a coalition of response teams – each with its own constituency, but working together to share information, provide alerts, and support each other in the response to incidents. The Forum of Incident Response and Security Teams (FIRST) includes teams from government, industry, computer manufacturers, and academia. NIST serves as the secretariat of FIRST.

12.3 Technical Support for Incident Handling

Incident handling will be greatly enhanced by technical mechanisms that enable the dissemination of information quickly and conveniently.

12.3.1 Communications for Centralized Reporting of Incidents

The technical ability to report incidents is of primary importance, since without knowledge of an incident, response is precluded. Fortunately, such technical mechanisms are already in place in many organizations.

For rapid response to constituency problems, a simple telephone "hotline" is practical and convenient. Some agencies may already have a number used for emergencies or for obtaining help with other problems; it may be practical (and cost-effective) to also use this number for incident handling. It may be necessary to provide 24-hour coverage for the hotline. This can be done by staffing the answering center, by providing an answering service for nonoffice hours, or by using a combination of an answering machine and personal pagers.

If additional mechanisms for contacting the incident handling team can be provided, it may increase access and thus benefit incident handling efforts. A centralized e-mail address that forwards mail to staff members would permit the constituency to conveniently exchange information with the team. Providing a fax number to users may also be helpful.

One way to establish a centralized reporting and incident response capability, while minimizing expenditures, is to use an existing Help Desk. Many agencies already have central Help Desks for fielding calls about commonly used applications, troubleshooting system problems, and providing help in detecting and eradicating computer viruses. By expanding the capabilities of the Help Desk and publicizing its telephone number (or e-mail address), an agency may be able to significantly improve its ability to handle many different types of incidents at minimal cost.

12.3.2 Rapid Communications Facilities

Some form of rapid communications is essential for quickly communicating with the constituency as well as with management officials and outside organizations. The team may need to send out security advisories or collect information quickly, thus some convenient form of communications, such as electronic mail, is generally highly desirable. With electronic mail, the team can easily direct information to various subgroups within the constituency, such as system managers or network managers, and broadcast general alerts to the entire constituency as needed. When connectivity already exists, e-mail has low overhead and is easy to use. (However, it is possible for the e-mail system itself to be attacked, as was the case with the 1988 Internet worm.)

Although there are substitutes for e-mail, they tend to increase response time. An electronic bulletin board system (BBS) can work well for distributing information, especially if it provides a convenient user interface that encourages its use. A BBS connected to a network is more convenient to access than one requiring a terminal and modem; however, the latter may be the only alternative for organizations without sufficient network connectivity. In addition, telephones, physical bulletin boards, and flyers can be used.

12.3.3 Secure Communications Facilities

Incidents can range from the trivial to those involving national security. Often when exchanging information about incidents, using encrypted communications may be advisable. This will help prevent the unintended distribution of incident-related information. Encryption technology is available for voice, fax, and e-mail communications.

12.4 Interdependencies

An incident handling capability generally depends upon other safeguards presented in this handbook. The most obvious is the strong link to other components of the contingency plan. The following paragraphs detail the most important of these interdependencies.

Contingency Planning. As discussed in the introduction to this chapter, an incident handling capability can be viewed as the component of contingency planning that deals with responding to technical threats, such as viruses or hackers. Close coordination is necessary with other contingency planning efforts, particularly when planning for contingency processing in the event of a serious unavailability of system resources.

Support and Operations. Incident handling is also closely linked to support and operations, especially user support and backups. For example, for purposes of efficiency and cost savings, the incident handling capability is often co-operated with a user "help desk." Also, backups of system resources may need to be used when recovering from an incident.

Training and Awareness. The training and awareness program can benefit from lessons learned during incident handling. Incident handling staff will be able to help assess the level of user awareness about current threats and vulnerabilities. Staff members may be able to help train system administrators, system operators, and other users and systems personnel. Knowledge of security precautions (resulting from such training) helps reduce future incidents. It is also important that users are trained what to report and how to report it.

Risk Management. The risk analysis process will benefit from statistics and logs showing the numbers and types of incidents that have occurred and the types of controls that are effective in preventing incidents. This information can be used to help select appropriate security controls and practices.

12.5 Cost Considerations

There are a number of start-up costs and funding issues to consider when planning an incident handling capability. Because the success of an incident handling capability relies so heavily on users' perceptions of its worth and whether they use it, it is very important that the capability be able to meet users' requirements. Two important funding issues are:

Personnel. An incident handling capability plan might call for at least one manager and one or more technical staff members (or their equivalent) to accomplish program objectives. Depending on the scope of the effort, however, full-time staff members may not be required. In some situations, some staff may be needed part-time or on an on-call basis. Staff may be performing incident handling duties as an adjunct responsibility to their normal assignments.

Education and Training. Incident handling staff will need to keep current with computer system and security developments. Budget allowances need to be made, therefore, for attending conferences, security seminars, and other continuing-education events. If an organization is located in more than one geographic areas, funds will probably be needed for travel to other sites for handling incidents.

References

Brand, Russell L. *Coping With the Threat of Computer Security Incidents: A Primer from Prevention Through Recovery.* July 1989.

Fedeli, Alan. "Organizing a Corporate Anti-Virus Effort." *Proceedings of the Third Annual Computer VIRUS Clinic*, Nationwide Computer Corp. March 1990.

Holbrook, P., and J. Reynolds, eds. *Site Security Handbook.* RFC 1244 prepared for the Internet Engineering Task Force, 1991. FTP from csrc.nist.gov:/put/secplcy/rfc1244.txt.

National Institute of Standards and Technology. "Establishing a Computer Security Incident Response Capability." Computer Systems Laboratory Bulletin. Gaithersburg, MD. February 1992.

Padgett, K. *Establishing and Operating an Incident Response Team.* Los Alamos, NM: Los Alamos National Laboratory, 1992.

Pethia, Rich, and Kenneth van Wyk. *Computer Emergency Response - An International Problem.* 1990.

Quarterman, John. *The Matrix - Computer Networks and Conferencing Systems Worldwide.* Digital Press, 1990.

Scherlis, William, S. Squires, and R. Pethia. *Computer Emergency Response.* 1989.

Schultz, E., D. Brown, and T. Longstaff. *Responding to Computer Security Incidents: Guidelines for Incident Handling.* University of California Technical Report UCRL-104689, 1990.

Proceedings of the Third Invitational Workshop on Computer Security Incident Response. August 1991.

Wack, John. *Establishing an Incident Response Capability.* Special Publication 800-3. Gaithersburg, MD: National Institute of Standards and Technology. November 1991.

Chapter 13

AWARENESS, TRAINING, AND EDUCATION

People, who are all fallible, are usually recognized as one of the weakest links in securing systems. The purpose of computer security awareness, training, and education is to enhance security by:

- improving awareness of the need to protect system resources;

- developing skills and knowledge so computer users can perform their jobs more securely; and

- building in-depth knowledge, as needed, to design, implement, or operate security programs for organizations and systems.

Making computer system users aware of their security responsibilities and teaching them correct practices helps users change their behavior.[95] It also supports *individual accountability*, which is one of the most important ways to improve computer security. Without knowing the necessary security measures (and to how to use them), users cannot be truly accountable for their actions. The importance of this training is emphasized in the Computer Security Act, which requires training for those involved with the management, use, and operation of federal computer systems.

This chapter first discusses the two overriding benefits of awareness, training, and education, namely: (1) improving employee behavior and (2) increasing the ability to hold employees accountable for their actions. Next, awareness, training, and education are discussed separately, with techniques used for each. Finally, the chapter presents one approach for developing a computer security awareness and training program.[96]

13.1 Behavior

People are a crucial factor in ensuring the security of computer systems and valuable information resources. Human actions account for a far greater degree of computer-related loss than all other sources combined. Of such losses, the actions of an organization's insiders normally cause far more harm than the actions of outsiders. (Chapter 4 discusses the major sources of computer-related loss.)

[95] One often-cited goal of training is changing people's attitudes. This chapter views changing attitudes as just one step toward changing behavior.

[96] This chapter does not discuss the specific contents of training programs. See the references for details of suggested course contents.

The major causes of loss due to an organization's own employees are: errors and omissions, fraud, and actions by disgruntled employees. One principal purpose of security awareness, training, and education is to reduce errors and omissions. However, it can also reduce fraud and unauthorized activity by disgruntled employees by increasing employees' knowledge of their accountability and the penalties associated with such actions.

Management sets the example for behavior within an organization. If employees know that management does not care about security, no training class teaching the importance of security and imparting valuable skills can be truly effective. This "tone from the top" has myriad effects an organization's security program.

13.2 Accountability

One of the keys to a successful computer security program is security awareness and training. If employees are not informed of applicable organizational policies and procedures, they cannot be expected to act effectively to secure computer resources.

Both the *dissemination* and the *enforcement* of policy are critical issues that are implemented and strengthened through training programs. Employees cannot be expected to follow policies and procedures of which they are unaware. In addition, enforcing penalties may be difficult if users can claim ignorance when caught doing something wrong.

Training employees may also be necessary to show that a standard of *due care* has been taken in protecting information. Simply issuing policy, with no follow-up to implement that policy, may not suffice.

Many organizations use *acknowledgment statements* which state that employees have read and understand computer security requirements. (An example is provided in Chapter 10.)

13.3 Awareness

Security *awareness* programs: (1) set the stage for *training* by changing organizational attitudes to realize the importance of security and the adverse consequences of its failure; and (2) remind users of the procedures to be followed.

Awareness stimulates and motivates those being trained to care about security and to remind them of important security practices. Explaining what happens to an organization, its mission, customers, and employees if security fails motivates people to take security seriously.

Awareness can take on different forms for particular audiences. Appropriate awareness for management officials might stress management's pivotal role in establishing organizational attitudes toward security. Appropriate awareness for other groups, such as system programmers or information analysts, should address the need for security as it relates to their job. In today's systems environment, almost everyone in an organization may have access to system resources – and therefore may have the potential to cause harm.

Comparative Framework

	AWARENESS	**TRAINING**	**EDUCATION**
Attribute:	"What"	"How"	"Why"
Level:	Information	Knowledge	Insight
Objective:	Recognition	Skill	Understanding
Teaching Method:	Media - Videos -Newsletters -Posters, etc.	Practical Instruction - Lecture - Case study workshop - Hands-on practice	Theoretical Instruction - Discussion Seminar - Background reading
Test Measure:	True/False Multiple Choice (identify learning)	Problem Solving (apply learning)	Eassay (interpret learning)
Impact Timeframe:	Short-term	Intermediate	Long-term

Figure 13.1 compares some of the differences in awareness, training, and education.

Awareness is used to reinforce the fact that security supports the mission of the organization by protecting valuable resources. If employees view security as just bothersome rules and procedures, they are more likely to ignore them. In addition, they may not make needed suggestions about improving security nor recognize and report security threats and vulnerabilities.

Awareness also is used to remind people of basic security practices, such as logging off a computer system or locking doors.

Techniques. A security awareness program can use many teaching methods, including video tapes, newsletters, posters, bulletin boards, flyers, demonstrations, briefings, short reminder notices at log-on, talks, or lectures. Awareness is often incorporated into basic security training and can use any method that can change employees' attitudes.

Effective security awareness programs need to be designed with the recognition that people tend to practice a *tuning out* process (also known as *acclimation*). For example, after a while, a security poster, no matter how well designed, will be ignored; it will, in effect, simply blend into the environment. For this reason, awareness techniques should be creative and frequently changed.

> Employees often regard computer security as an obstacle to productivity. A common feeling is that they are paid to produce, not to protect. To help motivate employees, awareness should emphasize how security, from a broader perspective, contributes to productivity. The consequences of poor security should be explained, while avoiding the fear and intimidation that employees often associate with security.

13.4 Training

The purpose of training is to teach people the skills that will enable them to perform their jobs more securely. This includes teaching people *what* they should do and *how* they should (or can) do it. Training can address many levels, from basic security practices to more advanced or specialized skills. It can be specific to one computer system or generic enough to address all systems.

Training is most effective when targeted to a specific audience. This enables the training to focus on security-related job skills and knowledge that people need performing their duties. Two types of audiences are general users and those who require specialized or advanced skills.

General Users. Most users need to understand good computer security practices, such as:

- protecting the physical area and equipment (e.g., locking doors, caring for floppy diskettes);
- protecting passwords (if used) or other authentication data or tokens (e.g., never divulge PINs); and
- reporting security violations or incidents (e.g., whom to call if a virus is suspected).

In addition, general users should be taught the organization's policies for protecting information and computer systems and the roles and responsibilities of various organizational units with which they may have to interact.

In teaching general users, care should be taken not to overburden them with unneeded details. These people are the target of multiple training programs, such as those addressing safety, sexual harassment, and AIDS in the workplace. The training should be made useful by addressing security issues that *directly* affect the users. The goal is to improve basic security practices, *not* to make everyone literate in all the jargon or philosophy of security.

Specialized or Advanced Training. Many groups need more advanced or more specialized training than just basic security practices. For example, managers may need to understand security consequences and costs so they can factor security into their decisions, or system administrators may need to know how to implement and use specific access control products.

There are many different ways to identify individuals or groups who need specialized or advanced training. One method is to look at job categories, such as executives, functional managers, or technology providers. Another method is to look at job functions, such as system design, system operation, or system use. A third method is to look at the specific technology and products used, especially for advanced training for user groups and training for a new system. This is further discussed in the section 13.6 of this chapter.

One group that has been targeted for specialized training is executives and functional managers. The training for management personnel is specialized (rather than advanced) because managers do *not* (as a general rule) need to understand the technical details of security. However, they do need to understand how to organize, direct, and evaluate security measures and programs. They also need to understand risk acceptance.

Techniques. A security training program normally includes training classes, either strictly devoted to security or as added special sections or modules within existing training classes. Training may be computer- or lecture-based (or both), and may include hands-on practice and case studies. Training, like awareness, also happens on the job.

13.5 Education

Security education is more in-depth than security training and is targeted for security professionals and those whose jobs require *expertise* in security.

Techniques. Security education is normally outside the scope of most organization awareness and training programs. It is more appropriately a part of *employee career development.* Security education is obtained through college or graduate classes or through specialized training programs. Because of this, most computer security programs focus primarily on awareness and training, as does the remainder of this chapter.[97]

13.6 Implementation[98]

An effective computer security awareness and training (CSAT) program requires proper planning, implementation, maintenance, and periodic evaluation. The following seven steps constitute *one approach* for developing a CSAT program.[99]

Step 1: Identify Program Scope, Goals, and Objectives.

Step 2: Identify Training Staff.

Step 3: Identify Target Audiences.

Step 4: Motivate Management and Employees.

Step 5: Administer the Program.

Step 6: Maintain the Program.

Step 7: Evaluate the Program.

13.6.1 Identify Program Scope, Goals, and Objectives

The first step in developing a CSAT program is to determine the program's scope, goals, and objectives. The scope of the CSAT program should provide training to all types of people who interact with computer systems. The scope of the program can be an entire organization or a subunit. Since users need training which relates directly to their use of

> The Computer Security Act of 1987 requires federal agencies to "provide for the mandatory periodic training in computer security awareness and accepted computer practices of all employees who are involved with the management, use, or operation of each federal computer system within or under the supervision of that agency." The scope and goals of federal computer security awareness and training programs must implement this broad mandate. (Other federal requirements for computer security training are contained in OMB Circular A-130, Appendix III, and OPM regulations.)

[97] Unfortunately, college and graduate security courses are not widely available. In addition, the courses may only address general security.

[98] This section is based on material prepared by the Department of Energy's Office of Information Management for its unclassified security program.

[99] This approach is presented to familiarize the reader with some of the important implementation issues. It is not the only approach to implementing an awareness and training program.

particular systems, a large organizationwide program may need to be supplemented by more specific programs. In addition, the organization should specifically address whether the program applies to employees only or also to other users of organizational systems.

Generally, the overall goal of a CSAT program is to sustain an appropriate level of protection for computer resources by increasing employee awareness of their computer security responsibilities and the ways to fulfill them. More specific goals may need to be established. Objectives should be defined to meet the organization's specific goals.

13.6.2 Identify Training Staff

There are many possible candidates for conducting the training including internal training departments, computer security staff, or contract services. Regardless of who is chosen, it is important that trainers have sufficient knowledge of computer security issues, principles, and techniques. It is also vital that they know how to communicate information and ideas effectively.

13.6.3 Identify Target Audiences

Not everyone needs the same degree or type of computer security information to do their jobs. A CSAT program that distinguishes between groups of people, presents only the information needed by the particular audience, and omits irrelevant information will have the best results. Segmenting audiences (e.g., by their function or familiarity with the system) can also improve the effectiveness of a CSAT program. For larger organizations, some individuals will fit into more than one group. For smaller organizations, segmenting may not be needed. The following methods are some examples of ways to do this.

Segment according to level of awareness. Individuals may be separated into groups according to their current level of awareness. This may require research to determine how well employees follow computer security procedures or understand how computer security fits into their jobs.

Segment according to general job task or function. Individuals may be grouped as data providers, data processors, or data users.

Segment according to specific job category. Many organizations assign individuals to job categories. Since each job category generally has different job responsibilities, training for each will be different. Examples of job categories could be general management, technology management, applications development, or security.

Segment according to level of computer knowledge. Computer experts may be expected to find a program containing highly technical information more valuable than one covering the management issues in computer security. Similarly, a computer novice would benefit more from a training program that presents introductory fundamentals.

Segment according to types of technology or systems used. Security techniques used for each off-the-shelf product or application system will usually vary. The users of major applications will normally require training specific to that application.

13.6.4 Motivate Management and Employees

To successfully implement an awareness and training program, it is important to gain the *support* of management and employees. Consideration should be given to using motivational techniques to show management and employees how their participation in the CSAT program will benefit the organization.

Management. Motivating management normally relies upon increasing awareness. Management needs to be aware of the losses that computer security can reduce and the role of training in computer security. Management commitment is necessary because of the resources used in developing and implementing the program and also because the program affects their staff.

Employees. Motivation of managers alone is not enough. Employees often need to be convinced of the merits of computer security and how it relates to their jobs. Without appropriate training, many employees will not fully comprehend the value of the system resources with which they work.

> Employees and managers should be solicited to provide input to the CSAT program. Individuals are more likely to support a program when they have actively participated in its development.

Some awareness techniques were discussed above. Regardless of the techniques that are used, employees should feel that their cooperation will have a beneficial impact on the organization's future (and, consequently, their own).

13.6.5 Administer the Program

There are several important considerations for administering the CSAT program.

Visibility. The visibility of a CSAT program plays a key role in its success. Efforts to achieve high visibility should begin during the early stages of CSAT program development. However, care should be give not to promise what cannot be delivered.

> The Federal Information Systems Security Educators' Association and NIST Computer Security Program Managers' Forum provide two means for federal government computer security program managers and training officers to share training ideas and materials.

Training Methods. The methods used in the CSAT program should be consistent with the material presented and tailored to the audience's needs. Some training and awareness methods and techniques are listed above (in the *Techniques* sections). Computer security awareness and training can be added to existing courses and presentations or taught separately. On-the-job training should also be considered.

Training Topics. There are more topics in computer security than can be taught in any one course. Topics should be selected based on the audience's requirements.

Training Materials. In general, higher-quality training materials are more favorably received and are more expensive. Costs, however, can be minimized since training materials can often be obtained from other organizations. The cost of modifying materials is normally less than developing training materials from scratch.

Training Presentation. Consideration should be given to the frequency of training (e.g., annually or as needed), the length of training presentations (e.g., 20 minutes for general presentations, one hour for updates or one week for an off-site class), and the style of training presentation (e.g., formal presentation, informal discussion, computer-based training, humorous).

13.6.6 Maintain the Program

Computer technology is an ever-changing field. Efforts should be made to keep abreast of changes in computer technology and security requirements. A training program that meets an organization's needs today may become ineffective when the organization starts to use a new application or changes its environment, such as by connecting to the Internet. Likewise, an awareness program can become obsolete if laws or organization policies change. For example, the awareness program should make employees aware of a new policy on e-mail usage. Employees may discount the CSAT program, and by association the importance of computer security, if the program does not provide current information.

13.6.7 Evaluate the Program

It is often difficult to measure the effectiveness of an awareness or training program. Nevertheless, an evaluation should attempt to ascertain how much information is retained, to what extent computer security procedures are being followed, and general attitudes toward computer security. The results of such an evaluation should help identify and correct problems. Some evaluation methods (which can be used in conjunction with one another) are:

- Use student evaluations.
- Observe how well employees follow recommended security procedures.
- Test employees on material covered.

- Monitor the number and kind of computer security incidents reported before and after the program is implemented.[100]

13.7 Interdependencies

Training can, and in most cases should, be used to support every control in the handbook. All controls are more effective if designers, implementers, and users are thoroughly trained.

Policy. Training is a critical means of informing employees of the contents of and reasons for the organization's policies.

Security Program Management. Federal agencies need to ensure that appropriate computer security awareness and training is provided, as required under the Computer Security Act of 1987. A security program should ensure that an organization is meeting all applicable laws and regulations.

Personnel/User Issues. Awareness, training, and education are often included with other personnel/user issues. Training is often required before access is granted to a computer system.

13.8 Cost Considerations

The major cost considerations in awareness, training, and education programs are:

- the cost of preparing and updating materials, including the time of the preparer;
- the cost of those providing the instruction;
- employee time attending courses and lectures or watching videos; and
- the cost of outside courses and consultants (both of which may including travel expenses), including course maintenance.

References

Alexander, M. ed. "Multimedia Means Greater Awareness." *Infosecurity News*. 4(6), 1993. pp. 90-94.

[100] The number of incidents will not necessarily go down. For example, virus-related losses may decrease when users know the proper procedures to avoid infection. On the other hand, reports of incidents may go up as users employ virus scanners and find more viruses. In addition, users will now know that virus incidents should be reported and to whom the reports should be sent.

Burns, G.M. "A Recipe for a Decentralized Security Awareness Program." *ISSA Access*. Vol. 3, Issue 2, 2nd Quarter 1990. pp. 12-54.

Code of Federal Regulations. 5 CFR 930. Computer Security Training Regulation.

Flanders, D. "Security Awareness - A 70% Solution." Fourth Workshop on Computer Security Incident Handling, August 1992.

Isaacson, G. "Security Awareness: Making It Work." *ISSA Access*. 3(4), 1990. pp. 22-24.

National Aeronautics and Space Administration. *Guidelines for Development of Computer Security Awareness and Training (CSAT) Programs*. Washington, DC. NASA Guide 2410.1. March 1990.

Maconachy, V. "Computer Security Education, Training, and Awareness: Turning a Philosophical Orientation Into Practical Reality." *Proceedings of the 12th National Computer Security Conference*. National Institute of Standards and Technology and National Computer Security Center. Washington, DC. October 1989.

Maconachy, V. "Panel: Federal Information Systems Security Educators' Association (FISSEA)." *Proceeding of the 15th National Computer Security Conference*. National Institute of Standards and Technology and National Computer Security Center. Baltimore, MD. October 1992.

Suchinsky, A. "Determining Your Training Needs." *Proceedings of the 13th National Computer Security Conference*. National Institute of Standards and Technology and National Computer Security Center. Washington, DC. October 1990.

Todd, M.A. and Guitian C. *"Computer Security Training Guidelines."* Special Publication 500-172. Gaithersburg, MD: National Institute of Standards and Technology. November 1989.

U.S. Department of Energy. *Computer Security Awareness and Training Guideline* (Vol. 1). Washington, DC. DOE/MA-0320. February 1988.

Wells, R.O. "Security Awareness for the Non-Believers." *ISSA Access*. Vol. 3, Issue 2, 2nd Quarter 1990. pp. 10-61.

Chapter 14

SECURITY CONSIDERATIONS
IN
COMPUTER SUPPORT AND OPERATIONS

Computer support and operations refers to everything done to run a computer system. This includes both system administration and tasks external to the system that support its operation (e.g., maintaining documentation). It does not include system planning or design. The support and operation of any computer system, from a three-person local area network to a worldwide application serving thousands of users, is critical to maintaining the security of a system. Support and operations are routine activities that enable computer systems to function correctly. These include fixing software or hardware problems, loading and maintaining software, and helping users resolve problems.

> System management and administration staff generally perform support and operations tasks although sometimes users do. Larger systems may have full-time operators, system programmers, and support staff performing these tasks. Smaller systems may have a part-time administrator.

The failure to consider security as part of the support and operations of computer systems is, for many organizations, their Achilles heel. Computer security system literature includes many examples of how organizations undermined their often expensive security measures because of poor documentation, old user accounts, conflicting software, or poor control of maintenance accounts. Also, an organization's policies and procedures often fail to address many of these important issues.

The important security considerations within some of the major categories of support and operations are:

- user support,
- software support,
- configuration management,
- backups,
- media controls,
- documentation, and
- maintenance.

> The primary goal of computer support and operations is the continued and correct operation of a computer system. One of the goals of computer security is the availability and integrity of systems. These goals are very closely linked.

Some special considerations are noted for larger or smaller systems.[101]

This chapter addresses the support and operations activities directly related to security. Every control discussed in this handbook relies, in one way or another, on computer system support and operations. This chapter, however, focuses on areas *not covered in other chapters*. For example, operations personnel normally create user accounts on the system. This topic is covered in the Identification and Authentication chapter, so it is not discussed here. Similarly, the input from support and operations staff to the security awareness and training program is covered in the Security Awareness, Training, and Education chapter.

14.1 User Support

In many organizations, user support takes place through a Help Desk. Help Desks can support an entire organization, a subunit, a specific system, or a combination of these. For smaller systems, the system administrator normally provides direct user support. Experienced users provide informal user support on most systems.

An important security consideration for user support personnel is being able to recognize which problems (brought to their attention by users) are security-related. For example, users' inability to log onto a computer system may result from the disabling of their accounts due to too many failed access attempts. This could indicate the presence of hackers trying to guess users' passwords.

User support should be closely linked to the organization's incident handling capability. In many cases, the same personnel perform these functions.

In general, system support and operations staff need to be able to identify security problems, respond appropriately, and inform appropriate individuals. A wide range of possible security problems exist. Some will be internal to custom applications, while others apply to off-the-shelf products. Additionally, problems can be software- or hardware-based.

The more responsive and knowledgeable system support and operation staff personnel are, the less user support will be provided informally. The support other users provide is important, but they may not be aware of the "whole picture."

Small systems are especially susceptible to viruses, while networks are particularly susceptible to hacker attacks, which can be targeted at multiple systems. System support personnel should be able to recognize attacks and know how to respond.

[101] In general, larger systems include mainframes, large minicomputers, and WANs. Smaller systems include PCs and LANs.

14.2 Software Support

Software is the heart of an organization's computer operations, whatever the size and complexity of the system. Therefore, it is essential that software function correctly and be protected from corruption. There are many elements of software support.

One is *controlling what software is used on a system.* If users or systems personnel can load and execute any software on a system, the system is more vulnerable to viruses, to unexpected software interactions, and to software that may subvert or bypass security controls. One method of controlling software is to inspect or test software before it is loaded (e.g., to determine compatibility with custom applications or identify other unforeseen interactions). This can apply to new software packages, to upgrades, to off-the-shelf products, or to custom software, as deemed appropriate. In addition to controlling the loading and execution of new software, organizations should also give care to the configuration and use of powerful system utilities. System utilities can compromise the integrity of operating systems and logical access controls.

A second element in software support can be to ensure that *software has not been modified without proper authorization.* This involves the protection of software and backup copies. This can be done with a combination of logical and physical access controls.

Many organizations also include a program to ensure that software is properly licensed, as required. For example, an organization may audit systems for illegal copies of copyrighted software. This problem is primarily associated with PCs and LANs, but can apply to any type of system.

Viruses take advantage of the weak software controls in personal computers. Also, there are powerful utilities available for PCs that can restore deleted files, find hidden files, and interface directly with PC hardware, bypassing the operating system. Some organizations use personal computers without floppy drives in order to have better control over the system.

There are several widely available utilities that look for security problems in both networks and the systems attached to them. Some utilities look for and try to exploit security vulnerabilities. (This type of software is further discussed in Chapter 9.)

14.3 Configuration Management

Closely related to software support is *configuration management* – the process of keeping track of changes to the system and, if needed, approving them.[102] Configuration management normally addresses hardware, software, networking, and other changes; it can be formal or informal. The primary security goal of configuration management is ensuring that changes to the system do not unintentionally or unknowingly diminish security. Some of the methods discussed under software

[102] This chapter only addresses configuration management during the operational phase. Configuration management can have extremely important security consequences during the development phase of a system.

support, such as inspecting and testing software changes, can be used. Chapter 9 discusses other methods.

Note that the security goal is to know what changes occur, not to prevent security from being changed. There may be circumstances when security will be reduced. However, the decrease in security should be the result of a decision based on all appropriate factors.

For networked systems, configuration management should include external connections. Is the computer system connected? To what other systems? In turn, to what systems are these systems and organizations connected?

A second security goal of configuration management is ensuring that changes to the system are reflected in other documentation, such as the contingency plan. If the change is major, it may be necessary to reanalyze some or all of the security of the system. This is discussed in Chapter 8.

14.4 Backups

Support and operations personnel and sometimes users back up software and data. This function is critical to contingency planning. Frequency of backups will depend upon how often data changes and how important those changes are. Program managers should be consulted to determine what backup schedule is appropriate. Also, as a safety measure, it is useful to test that backup copies are actually usable. Finally, backups should be stored securely, as appropriate (discussed below).

Users of smaller systems are often responsible for their own backups. However, in reality they do not always perform backups regularly. Some organizations, therefore, task support personnel with making backups periodically for smaller systems, either automatically (through server software) or manually (by visiting each machine).

14.5 Media Controls

Media controls include a variety of measures to provide physical and environmental protection and accountability for tapes, diskettes, printouts, and other media. From a security perspective, media controls should be designed to prevent the loss of confidentiality, integrity, or availability of information, including data or software, when stored outside the system. This can include storage of information before it is input to the system and after it is output.

The extent of media control depends upon many factors, including the type of data, the quantity of media, and the nature of the user environment. Physical and environmental protection is used to prevent unauthorized individuals from accessing the media. It also protects against such factors as heat, cold, or harmful magnetic fields. When necessary, logging the use of individual media (e.g., a tape cartridge) provides detailed accountability – to hold authorized people responsible for their actions.

14.5.1 Marking

Controlling media may require some form of physical labeling. The labels can be used to identify media with special handling instructions, to locate needed information, or to log media (e.g., with serial/control numbers or bar codes) to support accountability. Identification is often by colored labels on diskettes or tapes or banner pages on printouts.

If labeling is used for special handling instructions, it is critical that people be appropriately trained. The marking of PC input and output is generally the responsibility of the *user*, not the system support staff. Marking backup diskettes can help prevent them from being accidentally overwritten.

Typical markings for media could include: Privacy Act Information, Company Proprietary, or Joe's Backup Tape. In each case, the individuals handling the media must know the applicable handling instructions. For example, at the Acme Patent Research Firm, proprietary information may not leave the building except under the care of a security officer. Also, Joe's Backup Tape should be easy to find in case something happens to Joe's system.

14.5.2 Logging

The logging of media is used to support accountability. Logs can include control numbers (or other tracking data), the times and dates of transfers, names and signatures of individuals involved, and other relevant information. Periodic spot checks or audits may be conducted to determine that no controlled items have been lost and that all are in the custody of individuals named in control logs. Automated media tracking systems may be helpful for maintaining inventories of tape and disk libraries.

14.5.3 Integrity Verification

When electronically stored information is read into a computer system, it may be necessary to determine whether it has been read correctly or subject to any modification. The integrity of electronic information can be verified using error detection and correction or, if intentional modifications are a threat, cryptographic-based technologies. (See Chapter 19.)

14.5.4 Physical Access Protection

Media can be stolen, destroyed, replaced with a look-alike copy, or lost. Physical access controls, which can limit these problems, include locked doors, desks, file cabinets, or safes.

If the media requires protection at all times, it may be necessary to actually output data to the media in a secure location (e.g., printing to a printer in a locked room instead of to a general-purpose printer in a common area).

Physical protection of media should be extended to backup copies stored offsite. They generally should be accorded an equivalent level of protection to media containing the same information stored onsite. (Equivalent protection does not mean that the security measures need to be exactly the same. The controls at the off-site location are quite likely to be different from the controls at the regular site.) Physical access is discussed in Chapter 15.

14.5.5 Environmental Protection

Magnetic media, such as diskettes or magnetic tape, require environmental protection, since they are sensitive to temperature, liquids, magnetism, smoke, and dust. Other media (e.g., paper and optical storage) may have different sensitivities to environmental factors.

14.5.6 Transmittal

Media control may be transferred both within the organization and to outside elements. Possibilities for securing such transmittal include sealed and marked envelopes, authorized messenger or courier, or U.S. certified or registered mail.

14.5.7 Disposition

When media is disposed of, it may be important to ensure that information is not improperly disclosed. This applies both to media that is *external* to a computer system (such as a diskette) and to media *inside* a computer system, such as a hard disk. The process of removing information from media is called *sanitization.*

Many people throw away old diskettes, believing that erasing the files on the diskette has made the data unretrievable. In reality, however, erasing a file simply removes the pointer to that file. The pointer tells the computer where the file is physically stored. Without this pointer, the files will not appear on a directory listing. This does *not* mean that the file was removed. Commonly available utility programs can often retrieve information that is presumed deleted.

Three techniques are commonly used for media sanitization: overwriting, degaussing, and destruction. *Overwriting* is an effective method for clearing data from magnetic media. As the name implies, overwriting uses a program to write (1s, 0s, or a combination) onto the media. Common practice is to overwrite the media three times. Overwriting should not be confused with merely deleting the pointer to a file (which typically happens when a *delete* command is used). Overwriting requires that the media be in working order. *Degaussing* is a method to magnetically erase data from magnetic media. Two types of degausser exist: strong permanent magnets and electric degaussers. The final method of sanitization is *destruction* of the media by shredding or burning.

14.6 Documentation

Documentation of all aspects of computer support and operations is important to ensure continuity and consistency. Formalizing operational practices and procedures with sufficient detail helps to eliminate security lapses and oversights, gives new personnel sufficiently detailed instructions, and provides a quality assurance function to help ensure that operations will be performed correctly and efficiently.

The security of a system also needs to be documented. This includes many types of documentation, such as security plans, contingency plans, risk analyses, and security policies and procedures. Much of this information, particularly risk and threat analyses, has to be protected against unauthorized disclosure. Security documentation also needs to be both current and accessible. Accessibility should take special factors into account (such as the need to find the contingency plan during a disaster).

Security documentation should be designed to fulfill the needs of the different types of people who use it. For this reason, many organizations separate documentation into *policy* and *procedures*. A *security procedures manual* should be written to inform various system users how to do their jobs securely. A security procedures manual for systems operations and support staff may address a wide variety of technical and operational concerns in considerable detail.

14.7 Maintenance

System maintenance requires either physical or logical access to the system. Support and operations staff, hardware or software vendors, or third-party service providers may maintain a system. Maintenance may be performed on site, or it may be necessary to move equipment to a repair site. Maintenance may also be performed remotely via communications connections. If someone who does not normally have access to the system performs maintenance, then a security vulnerability is introduced.

In some circumstances, it may be necessary to take additional precautions, such as conducting background investigations of service personnel. Supervision of maintenance personnel may prevent some problems, such as "snooping around" the physical area. However, once someone has access to the system, it is very difficult for supervision to prevent damage done through the maintenance process.

Many computer systems provide *maintenance accounts*. These special log-in accounts are normally preconfigured at the factory with pre-set, widely known passwords. *It is critical to change these passwords or*

> One of the most common methods hackers use to break into systems is through maintenance accounts that still have factory-set or easily guessed passwords.

otherwise disable the accounts until they are needed. Procedures should be developed to ensure that only authorized maintenance personnel can use these accounts. If the account is to be used remotely, authentication of the maintenance provider can be performed using call-back confirmation. This helps ensure that remote diagnostic activities actually originate from an established phone number at the vendor's site. Other techniques can also help, including encryption and decryption of diagnostic communications; strong identification and authentication techniques, such as tokens; and remote disconnect verification.

Larger systems may have *diagnostic ports.* In addition, manufacturers of larger systems and third-party providers may offer more diagnostic and support services. It is critical to ensure that these ports are only used by authorized personnel and cannot be accessed by hackers.

14.8 Interdependencies

There are support and operations components in most of the controls discussed in this handbook.

Personnel. Most support and operations staff have special access to the system. Some organizations conduct background checks on individuals filling these positions to screen out possibly untrustworthy individuals.

Incident Handling. Support and operations may include an organization's incident handling staff. Even if they are separate organizations, they need to work together to recognize and respond to incidents.

Contingency Planning. Support and operations normally provides technical input to contingency planning and carries out the activities of making backups, updating documentation, and practicing responding to contingencies.

Security Awareness, Training, and Education. Support and operations staff should be trained in security procedures and should be aware of the importance of security. In addition, they provide technical expertise needed to teach users how to secure their systems.

Physical and Environmental. Support and operations staff often control the immediate physical area around the computer system.

Technical Controls. The technical controls are installed, maintained, and used by support and operations staff. They create the user accounts, add users to access control lists, review audit logs for unusual activity, control bulk encryption over telecommunications links, and perform the countless operational tasks needed to use technical controls effectively. In addition, support and operations staff provide needed input to the selection of controls based on their knowledge of system capabilities and operational constraints.

Assurance. Support and operations staff ensure that changes to a system do not introduce security vulnerabilities by using assurance methods to evaluate or test the changes and their effect on the system. Operational assurance is normally performed by support and operations staff.

14.9 Cost Considerations

The cost of ensuring adequate security in day-to-day support and operations is largely dependent upon the size and characteristics of the operating environment and the nature of the processing being performed. If sufficient support personnel are already available, it is important that they be trained in the security aspects of their assigned jobs; it is usually not necessary to hire additional support and operations security specialists. Training, both initial and ongoing, is a cost of successfully incorporating security measures into support and operations activities.

Another cost is that associated with creating and updating documentation to ensure that security concerns are appropriately reflected in support and operations policies, procedures, and duties.

References

Bicknell, Paul. "Data Security for Personal Computers." *Proceedings of the 15th National Computer Security Conference.* Vol. I. National Institute of Standards and Technology and National Computer Security Center. Baltimore, MD. October 1992.

Caelli, William, Dennis Longley, and Michael Shain. *Information Security Handbook.* New York, NY: Stockton Press, 1991.

Carnahan, Lisa J. "A Local Area Network Security Architecture." *Proceedings of the 15th National Computer Security Conference.* Vol. I. National Institute of Standards and Technology and National Computer Security Center. Baltimore, MD. 1992.

Carroll, J.M. *Managing Risk: A Computer-Aided Strategy.* Boston, MA: Butterworths, 1984.

Chapman, D. Brent. "Network (In)Security Through IP Packet Filtering." *Proceedings of the 3rd USENIX UNIX Security Symposium*, 1992.

Curry, David A. *UNIX System Security: A Guide for Users and System Administrators.* Reading, MA: Addison-Wesley Publishing Co., Inc., 1992.

Garfinkel, Simson, and Gene Spafford. *Practical UNIX Security.* Sebastopol, CA: O'Reilly & Associates, 1991.

Holbrook, Paul, and Joyce Reynolds, eds. *Site Security Handbook.* Available by anonymous ftp from nic.ddn.mil (in rfc directory).

Internet Security for System & Network Administrators. Computer Emergency Response Team Security Seminars, CERT Coordination Center, 1993.

Murray, W.H. "Security Considerations for Personal Computers." *Tutorial: Computer and Network Security.* Oakland, CA: IEEE Computer Society Press, 1986.

Parker, Donna B. *Managers Guide to Computer Security.* Reston, VA: Reston Publishing, Inc., 1981.

Pfleeger, Charles P. *Security in Computing*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1989.

Chapter 15

PHYSICAL AND ENVIRONMENTAL SECURITY

Physical and environmental security controls are implemented to protect the facility housing system resources, the system resources themselves, and the facilities used to support their operation.

The term *physical and environmental security*, as used in this chapter, refers to measures taken to protect systems, buildings, and related supporting infrastructure against threats associated with their physical environment.[103] Physical and environmental security controls include the following three broad areas:

1. The physical facility is usually the building, other structure, or vehicle housing the system and network components. Systems can be characterized, based upon their operating location, as static, mobile, or portable. Static systems are installed in structures at fixed locations. Mobile systems are installed in vehicles that perform the function of a structure, but not at a fixed location. Portable systems are not installed in fixed operating locations. They may be operated in wide variety of locations, including buildings or vehicles, or in the open. The physical characteristics of these structures and vehicles determine the level of such physical threats as fire, roof leaks, or unauthorized access.

2. The facility's general geographic operating location determines the characteristics of *natural threats*, which include earthquakes and flooding; *man-made threats* such as burglary, civil disorders, or interception of transmissions and emanations; and *damaging nearby activities*, including toxic chemical spills, explosions, fires, and electromagnetic interference from emitters, such as radars.

3. Supporting facilities are those services (both technical and human) that underpin the operation of the system. The system's operation usually depends on supporting facilities such as electric power, heating and air conditioning, and telecommunications. The failure or substandard performance of these facilities may interrupt operation of the system and may cause physical damage to system hardware or stored data.

This chapter first discusses the benefits of physical security measures, and then presents an overview of common physical and environmental security controls. Physical and environmental security measures result in many benefits, such as protecting employees. This chapter focuses on the protection of computer systems from the following:

[103] This chapter draws upon work by Robert V. Jacobson, International Security Technology, Inc., funded by the Tennessee Valley Authority.

Interruptions in Providing Computer Services. An external threat may interrupt the scheduled operation of a system. The magnitude of the losses depends on the duration and timing of the service interruption and the characteristics of the operations end users perform.

Physical Damage. If a system's hardware is damaged or destroyed, it usually has to be repaired or replaced. Data may be destroyed as an act of sabotage by a physical attack on data storage media (e.g., rendering the data unreadable or only partly readable). If data stored by a system for operational use is destroyed or corrupted, the data needs to be restored from back-up copies or from the original sources before the system can be used. The magnitude of loss from physical damage depends on the cost to repair or replace the damaged hardware *and* data, as well as costs arising from service interruptions.

Unauthorized Disclosure of Information. The physical characteristics of the facility housing a system may permit an intruder to gain access both to media external to system hardware (such as diskettes, tapes and printouts) and to media within system components (such as fixed disks), transmission lines or display screens. All may result in loss of disclosure-sensitive information.

Loss of Control over System Integrity. If an intruder gains access to the central processing unit, it is usually possible to reboot the system and *bypass* logical access controls. This can lead to information disclosure, fraud, replacement of system and application software, introduction of a Trojan horse, and more. Moreover, if such access is gained, it may be very difficult to determine what has been modified, lost, or corrupted.

Physical Theft. System hardware may be stolen. The magnitude of the loss is determined by the costs to replace the stolen hardware and restore data stored on stolen media. Theft may also result in service interruptions.

This chapter discusses seven major areas of physical and environmental security controls:

- physical access controls,
- fire safety,
- supporting utilities,
- structural collapse,
- plumbing leaks,
- interception of data, and
- mobile and portable systems.

15.1 Physical Access Controls

Physical access controls restrict the entry and exit of personnel (and often equipment and media) from an area, such as an office building, suite, data center, or room containing a LAN server.

The controls over physical access to the elements of a system can include controlled areas, barriers that isolate each area, entry points in the barriers, and screening measures at each of the entry points. In addition, staff members who work in a restricted area serve an important role in providing physical security, as they can be trained to challenge people they do not recognize.

Life Safety

It is important to understand that the objectives of physical access controls may be in conflict with those of *life safety*. Simply stated, life safety focuses on providing easy exit from a facility, particularly in an emergency, while physical security strives to control entry. In general, life safety must be given first consideration, but it is usually possible to achieve an effective balance between the two goals.

For example, it is often possible to equip emergency exit doors with a time delay. When one pushes on the panic bar, a loud alarm sounds, and the door is released after a brief delay. The expectation is that people will be deterred from using such exits improperly, but will not be significantly endangered during an emergency evacuation.

Physical access controls should address not only the area containing system hardware, but also locations of wiring used to connect elements of the system, the electric power service, the air conditioning and heating plant, telephone and data lines, backup media and source documents, and any other elements required system's operation. This means that all the areas in the building(s) that contain system elements must be identified.

There are many types of physical access controls, including badges, memory cards, guards, keys, true-floor-to-true-ceiling wall construction, fences, and locks.

It is also important to review the effectiveness of physical access controls in each area, both during normal business hours, and at other times – particularly when an area may be unoccupied. Effectiveness depends on both the characteristics of the control devices used (e.g., keycard-controlled doors) and the implementation and operation. Statements to the effect that "only authorized persons may enter this area" are not particularly effective. Organizations should determine whether intruders can easily defeat the controls, the extent to which strangers are challenged, and the effectiveness of other control procedures. Factors like these modify the effectiveness of physical controls.

The feasibility of surreptitious entry also needs to be considered. For example, it may be possible to go over the top of a partition that stops at the underside of a suspended ceiling or to cut a hole in a plasterboard partition in a location hidden by furniture. If a door is controlled by a combination lock, it may be possible to observe an authorized person entering the lock combination. If keycards are not carefully controlled, an intruder may be able to steal a card left on a desk or use a card passed back by an accomplice.

Corrective actions can address any of the factors listed above. Adding an additional barrier reduces the risk to the areas behind the barrier. Enhancing the screening at an entry point can reduce the number of penetrations. For example, a guard may provide a higher level of screening than a keycard-controlled door, or an anti-passback feature can be added. Reorganizing traffic patterns, work flow, and work areas may reduce the number of people who need access to a restricted area. Physical modifications to barriers can reduce the vulnerability to surreptitious entry. Intrusion detectors, such as closed-circuit television cameras, motion detectors, and other devices, can detect intruders in unoccupied spaces.

Types of Building Construction

There are four basic kinds of building construction: (a) light frame, (b) heavy timber, (c) incombustible, and (d) fire resistant. Note that the term *fireproof* is not used because no structure can resist a fire indefinitely. Most houses are light frame, and cannot survive more than about thirty minutes in a fire. Heavy timber means that the basic structural elements have a minimum thickness of four inches. When such structures burn, the char that forms tends to insulate the interior of the timber and the structure may survive for an hour or more depending on the details. Incombustible means that the structure members will not burn. This almost always means that the members are steel. Note, however, that steel loses it strength at high temperatures, at which point the structure collapses. Fire resistant means that the structural members are incombustible and are insulated. Typically, the insulation is either concrete that encases steel members, or is a mineral wool that is sprayed onto the members. Of course, the heavier the insulation, the longer the structure will resist a fire.

Note that a building constructed of reinforced concrete can still be destroyed in a fire if there is sufficient fuel present and fire fighting is ineffective. The prolonged heat of a fire can cause differential expansion of the concrete which causes *spalling*. Portions of the concrete split off, exposing the reinforcing, and the interior of the concrete is subject to additional spalling. Furthermore, as heated floor slabs expand outward, they deform supporting columns. Thus, a reinforced concrete parking garage with open exterior walls and a relatively low fire load has a low fire risk, but a similar archival record storage facility with closed exterior walls and a high fire load has a higher risk even though the basic building material is incombustible.

15.2 Fire Safety Factors

Building fires are a particularly important security threat because of the potential for complete destruction of both hardware and data, the risk to human life, and the pervasiveness of the damage. Smoke, corrosive gases, and high humidity from a localized fire can damage systems throughout an entire building. Consequently, it is important to evaluate the fire safety of buildings that house systems. Following are important factors in determining the risks from fire.

Ignition Sources. Fires begin because something supplies enough heat to cause other materials to burn. Typical ignition sources are failures of electric devices and wiring, carelessly discarded cigarettes, improper storage of materials subject to spontaneous combustion, improper operation of heating devices, and, of course, arson.

Fuel Sources. If a fire is to grow, it must have a supply of fuel, material that will burn to support its growth, and an adequate supply of oxygen. Once a fire becomes established, it depends on the combustible materials in the building (referred to as the fire load) to support its further growth. The more fuel per square meter, the more intense the fire will be.

Building Operation. If a building is well maintained and operated so as to minimize the accumulation of fuel (such as maintaining the integrity of fire barriers), the fire risk will be minimized.

Building Occupancy. Some occupancies are inherently more dangerous than others because of an above-average number of potential ignition sources. For example, a chemical warehouse may contain an above-average fuel load.

Fire Detection. The more quickly a fire is detected, all other things being equal, the more easily it can be extinguished, minimizing damage. It is also important to accurately pinpoint the location of the fire.

Fire Extinguishment. A fire will burn until it consumes all of the fuel in the building or until it is extinguished. Fire extinguishment may be automatic, as with an automatic sprinkler system or a HALON discharge system, or it may be performed by people using portable extinguishers, cooling the fire site with a stream of water, by limiting the supply of oxygen with a blanket of foam or powder, or by breaking the combustion chemical reaction chain.

When properly installed, maintained, and provided with an adequate supply of water, automatic sprinkler systems are highly effective in protecting buildings and their contents.[104] Nonetheless, one often hears uninformed persons speak of the *water damage* done by sprinkler systems as a disadvantage. *Fires that trigger sprinkler systems* cause the water damage.[105] In short, sprinkler systems reduce fire damage, protect the lives of building occupants, and limit the fire damage to the building itself. All these factors contribute to more rapid recovery of systems

Halons have been identified as harmful to the Earth's protective ozone layer. So, under an international agreement (known as the Montreal Protocol), production of halons ended January 1, 1994. In September 1992, the General Services Administration issued a moratorium on halon use by federal agencies.

[104] As discussed in this section, many variables affect fire safety and should be taken into account in selecting a fire extinguishment system. While automatic sprinklers can be very effective, selection of a fire extinguishment system for a particular building should take into account the particular fire risk factors. Other factors may include rate changes from either a fire insurance carrier or a business interruption insurance carrier. Professional advice is required.

[105] Occurrences of accidental discharge are extremely rare, and, in a fire, only the sprinkler heads in the immediate area of the fire open and discharge water.

following a fire.

Each of these factors is important when estimating the occurrence rate of fires and the amount of damage that will result. The objective of a fire-safety program is to optimize these factors to minimize the risk of fire.

15.3 Failure of Supporting Utilities

Systems and the people who operate them need to have a reasonably well-controlled operating environment. Consequently, failures of heating and air-conditioning systems will usually cause a service interruption and may damage hardware. These utilities are composed of many elements, each of which must function properly.

For example, the typical air-conditioning system consists of (1) air handlers that cool and humidify room air, (2) circulating pumps that send chilled water to the air handlers, (3) chillers that extract heat from the water, and (4) cooling towers that discharge the heat to the outside air. Each of these elements has a mean-time-between-failures (MTBF) and a mean-time-to-repair (MTTR). Using the MTBF and MTTR values for each of the elements of a system, one can estimate the occurrence rate of system failures and the range of resulting service interruptions.

This same line of reasoning applies to electric power distribution, heating plants, water, sewage, and other utilities required for system operation or staff comfort. By identifying the failure modes of each utility and estimating the MTBF and MTTR, necessary failure threat parameters can be developed to calculate the resulting risk. The risk of utility failure can be reduced by substituting units with lower MTBF values. MTTR can be reduced by stocking spare parts on site and training maintenance personnel. And the outages resulting from a given MTBF can be reduced by installing redundant units under the assumption that failures are distributed randomly in time. Each of these strategies can be evaluated by comparing the reduction in risk with the cost to achieve it.

15.4 Structural Collapse

A building may be subjected to a load greater than it can support. Most commonly this is a result of an earthquake, a snow load on the roof beyond design criteria, an explosion that displaces or cuts structural members, or a fire that weakens structural members. Even if the structure is not completely demolished, the authorities may decide to ban its further use, sometimes even banning entry to remove materials. This threat applies primarily to high-rise buildings and those with large interior spaces without supporting columns.

15.5 Plumbing Leaks

While plumbing leaks do not occur every day, they can be seriously disruptive. The building's plumbing drawings can help locate plumbing lines that might endanger system hardware. These lines include hot and cold water, chilled water supply and return lines, steam lines, automatic sprinkler lines, fire hose standpipes, and drains. If a building includes a laboratory or manufacturing spaces, there may be other lines that conduct water, corrosive or toxic chemicals, or gases.

As a rule, analysis often shows that the cost to relocate threatening lines is difficult to justify. However, the location of shutoff valves and procedures that should be followed in the event of a failure must be specified. Operating and security personnel should have this information immediately available for use in an emergency. In some cases, it may be possible to relocate system hardware, particularly distributed LAN hardware.

15.6 Interception of Data

Depending on the type of data a system processes, there may be a significant risk if the data is intercepted. There are three routes of data interception: direct observation, interception of data transmission, and electromagnetic interception.

Direct Observation. System terminal and workstation display screens may be observed by unauthorized persons. In most cases, it is relatively easy to relocate the display to eliminate the exposure.

Interception of Data Transmissions. If an interceptor can gain access to data transmission lines, it may be feasible to tap into the lines and read the data being transmitted. Network monitoring tools can be used to capture data packets. Of course, the interceptor cannot control what is transmitted, and so may not be able to immediately observe data of interest. However, over a period of time there may be a serious level of disclosure. Local area networks typically broadcast messages.[106] Consequently, all traffic, including passwords, could be retrieved. Interceptors could also transmit spurious data on tapped lines, either for purposes of disruption or for fraud.

Electromagnetic Interception. Systems routinely radiate electromagnetic energy that can be detected with special-purpose radio receivers. Successful interception will depend on the signal strength at the receiver location; the greater the separation between the system and the receiver, the lower the success rate. TEMPEST shielding, of either equipment or rooms, can be used to minimize the spread of electromagnetic signals. The signal-to-noise ratio at the receiver,

[106] An insider may be able to easily collect data by configuring their ethernet network interface to receive all network traffic, rather than just network traffic intended for this node. This is called the *promiscuous* mode.

determined in part by the number of competing emitters will also affect the success rate. The more workstations of the same type in the same location performing "random" activity, the more difficult it is to intercept a given workstation's radiation. On the other hand, the trend toward wireless (i.e., deliberate radiation) LAN connections may increase the likelihood of successful interception.

15.7 Mobile and Portable Systems

The analysis and management of risk usually has to be modified if a system is installed in a vehicle or is portable, such as a laptop computer. The system in a vehicle will share the risks of the vehicle, including accidents and theft, as well as regional and local risks.

Portable and mobile systems share an increased risk of theft and physical damage. In addition, portable systems can be "misplaced" or left unattended by careless users. Secure storage of laptop computers is often required when they are not in use.

> Encryption of data files on stored media may also be a cost-effective precaution against disclosure of confidential information if a laptop computer is lost or stolen.

If a mobile or portable system uses particularly valuable or important data, it may be appropriate to either store its data on a medium that can be removed from the system when it is unattended or to encrypt the data. In any case, the issue of how custody of mobile and portable computers are to be controlled should be addressed. Depending on the sensitivity of the system and its application, it may be appropriate to require briefings of users and signed briefing acknowledgments. (See Chapter 10 for an example.)

15.8 Approach to Implementation

Like other security measures, physical and environmental security controls are selected because they are cost-beneficial. This does not mean that a user must conduct a detailed cost-benefit analysis for the selection of every control. There are four general ways to justify the selection of controls:

1. *They are required by law or regulation.* Fire exit doors with panic bars and exit lights are examples of security measures required by law or regulation. Presumably, the regulatory authority has considered the costs and benefits and has determined that it is in the public interest to require the security measure. A lawfully conducted organization has no option but to implement all required security measures.

2. *The cost is insignificant, but the benefit is material.* A good example of this is a facility with a key-locked low-traffic door to a restricted access. The cost of keeping the door

locked is minimal, but there is a significant benefit. Once a significant benefit/minimal cost security measure has been identified, no further analysis is required to justify its implementation.

3. The security measure addresses a potentially "fatal" security exposure but has a reasonable cost. Backing up system software and data is an example of this justification . For most systems, the cost of making regular backup copies is modest (compared to the costs of operating the system), the organization would not be able to function if the stored data were lost, and the cost impact of the failure would be material. In such cases, it would not be necessary to develop any further cost justification for the backup of software and data. However, this justification depends on what constitutes a *modest* cost, and it does not identify the optimum backup schedule. Broadly speaking, a cost that does not require budgeting of additional funds would qualify.

4. The security measure is estimated to be cost-beneficial. If the cost of a potential security measure is significant, and it cannot be justified by any of the first three reasons listed above, then its cost (both implementation and ongoing operation) and its benefit (reduction in future expected losses) need to be analyzed to determine if it is cost-beneficial. In this context, *cost-beneficial* means that the reduction in expected loss is significantly greater than the cost of implementing the security measure.

Arriving at the fourth justification requires a detailed analysis. Simple rules of thumb do not apply. Consider, for example, the threat of electric power failure and the security measures that can protect against such an event. The threat parameters, rate of occurrence, and range of outage durations depend on the location of the system, the details of its connection to the local electric power utility, the details of the internal power distribution system, and the character of other activities in the building that use electric power. The system's potential losses from service interruption depends on the details of the functions it performs. Two systems that are otherwise identical can support functions that have quite different degrees of urgency. Thus, two systems may have the same electric power failure threat and vulnerability parameters, yet entirely different loss potential parameters.

Furthermore, a number of different security measures are available to address electric power failures. These measures differ in both cost and performance. For example, the cost of an uninterruptible power supply (UPS) depends on the size of the electric load it can support, the number of minutes it can support the load, and the speed with which it assumes the load when the primary power source fails. An on-site power generator could also be installed either in place of a UPS (accepting the fact that a power failure will cause a brief service interruption) or in order to provide long-term backup to a UPS system. Design decisions include the magnitude of the load the generator will support, the size of the on-site fuel supply, and the details of the facilities to switch the load from the primary source or the UPS to the on-site generator.

This example shows systems with a wide range of risks and a wide range of available security measures (including, of course, no action), each with its own cost factors and performance parameters.

15.9 Interdependencies

Physical and environmental security measures rely on and support the proper functioning of many of the other areas discussed in this handbook. Among the most important are the following:

Logical Access Controls. Physical security controls augment technical means for controlling access to information and processing. Even if the most advanced and best-implemented logical access controls are in place, if physical security measures are inadequate, logical access controls may be circumvented by directly accessing the hardware and storage media. For example, a computer system may be rebooted using different software.

Contingency Planning. A large portion of the contingency planning process involves the failure of physical and environmental controls. Having sound controls, therefore, can help minimize losses from such contingencies.

Identification and Authentication (I&A). Many physical access control systems require that people be identified and authenticated. Automated physical security access controls can use the same types of I&A as other computer systems. In addition, it is possible to use the same tokens (e.g., badges) as those used for other computer-based I&A.

Other. Physical and environmental controls are also closely linked to the activities of the local guard force, fire house, life safety office, and medical office. These organizations should be consulted for their expertise in planning controls for the systems environment.

15.10 Cost Considerations

Costs associated with physical security measures range greatly. Useful generalizations about costs, therefore, are difficult make. Some measures, such as keeping a door locked, may be a trivial expense. Other features, such as fire-detection and -suppression systems, can be far more costly. Cost considerations should include operation. For example, adding controlled-entry doors requires persons using the door to stop and unlock it. Locks also require physical key management and accounting (and rekeying when keys are lost or stolen). Often these effects will be inconsequential, but they should be fully considered. As with other security measures, the objective is to select those that are cost-beneficial.

References

Alexander, M., ed. "Secure Your Computers and Lock Your Doors." *Infosecurity News*. 4(6), 1993. pp. 80-85.

Archer, R. "Testing: Following Strict Criteria." *Security Dealer*. 15(5), 1993. pp. 32-35.

Breese, H., ed. *The Handbook of Property Conservation*. Norwood, MA: Factory Mutual Engineering Corp.

Chanaud, R. "Keeping Conversations Confidential." *Security Management*. 37(3), 1993. pp. 43-48.

Miehl, F. "The Ins and Outs of Door Locks." *Security Management*. 37(2), 1993. pp. 48-53.

National Bureau of Standards. *Guidelines for ADP Physical Security and Risk Management*. Federal Information Processing Standard Publication 31. June 1974.

Peterson, P. "Infosecurity and Shrinking Media." *ISSA Access*. 5(2), 1992. pp. 19-22.

Roenne, G. "Devising a Strategy Keyed to Locks." *Security Management*. 38(4), 1994. pp. 55-56.

Zimmerman, J. "Using Smart Cards - A Smart Move." *Security Management*. 36(1), 1992. pp. 32-36.

IV. TECHNICAL CONTROLS

Chapter 16

IDENTIFICATION AND AUTHENTICATION

For most systems, identification and authentication (I&A) is the first line of defense. I&A is a technical measure that prevents unauthorized people (or unauthorized processes) from entering a computer system.

I&A is a critical building block of computer security since it is the basis for most types of access control and for establishing user accountability.[107] Access control often requires that the system be able to identify and differentiate among users. For example, access control is often based on *least privilege*, which refers to the granting to users of only those accesses required to perform their duties. User accountability requires the linking of activities on a computer system to specific individuals and, therefore, requires the system to identify users.

Identification is the means by which a user *provides* a claimed identity to the system. *Authentication*[108] is the means of establishing the *validity* of this claim.

This chapter discusses the basic means of identification and authentication, the current technology used to provide I&A, and some important implementation issues.

A typical user identification could be JSMITH (for Jane Smith). This information can be known by system administrators and other system users. A typical user authentication could be Jane Smith's password, which is kept secret. This way system administrators can set up Jane's access and see her activity on the audit trail, and system users can send her e-mail, but no one can pretend to be Jane.

Computer systems recognize people based on the authentication data the systems *receive*. Authentication presents several challenges: collecting authentication data, transmitting the data securely, and knowing whether the person who was originally authenticated is *still* the person using the computer system. For example, a user may walk away from a terminal while still logged on, and another person may start using it.

There are three means of authenticating a user's identity *which can be used alone or in combination*:

- something the individual *knows* (a secret– e.g., a password, Personal Identification Number (PIN), or cryptographic key);

[107] Not all types of access control require identification and authentication.

[108] Computers also use authentication to verify that a message or file has not been altered and to verify that a message originated with a certain person. This chapter only addresses user authentication. The other forms of authentication are addressed in the Chapter 19.

- something the individual *possesses* (a token – e.g., an ATM card or a smart card); and

- something the individual *is* (a biometric – e.g., such characteristics as a voice pattern, handwriting dynamics, or a fingerprint).

While it may appear that any of these means could provide strong authentication, there are problems associated with each. If people wanted to pretend to be someone else on a computer system, they can guess or learn that individual's password; they can also steal or fabricate tokens. Each method also has drawbacks for legitimate users and system administrators: users forget passwords and may lose tokens, and administrative overhead for keeping track of I&A data and tokens can be substantial. Biometric systems have significant technical, user acceptance, and cost problems as well.

For most applications, trade-offs will have to be made among security, ease of use, and ease of administration, especially in modern networked environments.

This section explains current I&A technologies and their benefits and drawbacks as they relate to the three means of authentication. Although some of the technologies make use of cryptography because it can significantly strengthen authentication, the explanations of cryptography appear in Chapter 19, rather than in this chapter.

16.1 I&A Based on Something the User Knows

The most common form of I&A is a user ID coupled with a password. This technique is based solely on something the user knows. There are other techniques besides *conventional* passwords that are based on knowledge, such as knowledge of a cryptographic key.

16.1.1 Passwords

In general, password systems work by requiring the user to enter a user ID and password (or passphrase or personal identification number). The system compares the password to a previously stored password for that user ID. If there is a match, the user is authenticated and granted access.

Benefits of Passwords. Passwords have been successfully providing security for computer systems for a long time. They are integrated into many operating systems, and users and system administrators are familiar with them. When properly managed in a controlled environment, they can provide effective security.

Problems With Passwords. The security of a password system is dependent upon keeping passwords secret. Unfortunately, there are many ways that the secret may be divulged. All of the problems discussed below can be significantly mitigated by improving password security, as discussed in the sidebar. However, there is no fix for the problem of electronic monitoring, except to use more advanced authentication (e.g., based on cryptographic techniques or tokens).

1. Guessing or finding passwords. If users select their own passwords, they tend to make them easy to remember. That often makes them easy to guess. The names of people's children, pets, or favorite sports teams are common examples. On the other hand, assigned passwords may be difficult to remember, so users are more likely to write them down. Many computer systems are shipped with administrative accounts that have preset passwords. Because these passwords are standard, they are easily "guessed." Although security practitioners have been warning about this problem for years, many system administrators still do not change default passwords. Another method of learning passwords is to observe someone entering a password or PIN. The observation can be done by someone in the same room or by someone some distance away using binoculars. This is often referred to as *shoulder surfing.*

2. Giving passwords away. Users may share their passwords. They may give their password to a co-worker in order to share files. In addition, people can be tricked into divulging their passwords. This process is referred to as *social engineering.*

Improving Password Security

Password generators. If users are not allowed to generate their own passwords, they cannot pick easy-to-guess passwords. Some generators create only pronounceable nonwords to help users remember them. However, users tend to write down hard-to-remember passwords.

Limits on log-in attempts. Many operating systems can be configured to lock a user ID after a set number of failed log-in attempts. This helps to prevent guessing of passwords.

Password attributes. Users can be instructed, or the system can force them, to select passwords (1) with a certain minimum length, (2) with special characters, (3) that are unrelated to their user ID, or (4) to pick passwords which are not in an on-line dictionary. This makes passwords more difficult to guess (but more likely to be written down).

Changing passwords. Periodic changing of passwords can reduce the damage done by stolen passwords and can make brute-force attempts to break into systems more difficult. Too frequent changes, however, can be irritating to users.

Technical protection of the password file. Access control and one-way encryption can be used to protect the password file itself.

Note: Many of these techniques are discussed in FIPS 112, *Password Usage* and FIPS 181, *Automated Password Generator.*

3. Electronic monitoring. When passwords are transmitted to a computer system, they can be electronically monitored. This can happen on the network used to transmit the password or on the computer system itself. Simple encryption of a password that will be used again does not solve this problem because encrypting the same password will create the same ciphertext; the ciphertext becomes the password.

> *4. Accessing the password file.* If the password file is not protected by strong access controls, the file can be downloaded. Password files are often protected with one-way encryption[109] so that plain-text passwords are not available to system administrators or hackers (if they successfully bypass access controls). Even if the file is encrypted, brute force can be used to learn passwords if the file is downloaded (e.g., by encrypting English words and comparing them to the file).

Passwords Used as Access Control. Some mainframe operating systems and many PC applications use passwords as a means of restricting access to specific resources within a system. Instead of using mechanisms such as access control lists (see Chapter 17), access is granted by entering a password. The result is a proliferation of passwords that can reduce the overall security of a system. While the use of passwords as a means of access control is common, it is an approach that is often less than optimal and not cost-effective.

16.1.2 Cryptographic Keys

Although the authentication derived from the knowledge of a cryptographic key may be based entirely on something the user knows, it is necessary for the user to also possess (or have access to) something that can perform the cryptographic computations, such as a PC or a smart card. For this reason, the protocols used are discussed in the Smart Tokens section of this chapter. However, it is possible to implement these types of protocols without using a smart token. Additional discussion is also provided under the Single Log-in section.

16.2 I&A Based on Something the User Possesses

Although some techniques are based solely on something the user possesses, most of the techniques described in this section are combined with something the user knows. This combination can provide significantly stronger security than either something the user knows or possesses alone.[110]

Objects that a user possesses for the purpose of I&A are called *tokens.* This section divides tokens into two categories: *memory tokens* and *smart tokens.*

[109] One-way encryption algorithms only provide for the encryption of data. The resulting ciphertext cannot be decrypted. When passwords are entered into the system, they are one-way encrypted, and the result is compared with the stored ciphertext. (See the Chapter 19.)

[110] For the purpose of understanding how possession-based I&A works, it is not necessary to distinguish whether possession of a token in various systems is identification or authentication.

16.2.1 Memory Tokens

Memory tokens store, but do not process, information. Special reader/writer devices control the writing and reading of data to and from the tokens. The most common type of memory token is a magnetic striped card, in which a thin stripe of magnetic material is affixed to the surface of a card (e.g., as on the back of credit cards). A common application of memory tokens for authentication to computer systems is the automatic teller machine (ATM) card. This uses a combination of something the user possesses (the card) with something the user knows (the PIN).

Some computer systems authentication technologies are based solely on possession of a token, but they are less common. Token-only systems are more likely to be used in other applications, such as for physical access. (See Chapter 15.)

Benefits of Memory Token Systems. Memory tokens when used with PINs provide significantly more security than passwords. In addition, memory cards are inexpensive to produce. For a hacker or other would-be masquerader to pretend to be someone else, the hacker must have both a valid token *and* the corresponding PIN. This is much more difficult than obtaining a valid password and user ID combination (especially since most user IDs are common knowledge).

Another benefit of tokens is that they can be used in support of log generation without the need for the employee to key in a user ID for each transaction or other logged event since the token can be scanned repeatedly. If the token is required for physical entry and exit, then people will be forced to remove the token when they leave the computer. This can help maintain authentication.

Problems With Memory Token Systems. Although sophisticated technical attacks are possible against memory token systems, most of the problems associated with them relate to their cost, administration, token loss, user dissatisfaction, and the compromise of PINs. Most of the techniques for increasing the security of memory token systems relate to the protection of PINs. Many of the techniques discussed in the sidebar on Improving Password Security apply to PINs.

1. *Requires special reader.* The need for a special reader increases the cost of using memory tokens. The readers used for memory tokens must include both the physical unit that reads the card and a processor that determines whether the card and/or the PIN entered with the card is valid. If the PIN or token is validated by a processor that is not physically located with the reader, then the authentication data is vulnerable to electronic monitoring (although cryptography can be used to solve this problem).

> Attacks on memory-card systems have sometimes been quite creative. One group stole an ATM machine that they installed at a local shopping mall. The machine collected valid account numbers and corresponding PINs, which the thieves used to forge cards. The forged cards were then used to withdraw money from legitimate ATMs.

2. *Token loss.* A lost token may prevent

the user from being able to log in until a replacement is provided. This can increase administrative overhead costs.

The lost token could be found by someone who wants to break into the system, or could be stolen or forged. If the token is also used with a PIN, any of the methods described above in password problems can be used to obtain the PIN. Common methods are finding the PIN taped to the card or observing the PIN being entered by the legitimate user. In addition, any information stored on the magnetic stripe that has not been encrypted can be read.

3. User Dissatisfaction. In general, users want computers to be easy to use. Many users find it inconvenient to carry and present a token. However, their dissatisfaction may be reduced if they see the need for increased security.

16.2.2 Smart Tokens

A smart token expands the functionality of a memory token by incorporating one or more integrated circuits into the token itself. When used for authentication, a smart token is another example of authentication based on something a user possesses (i.e., the token itself). A smart token typically requires a user also to provide something the user knows (i.e., a PIN or password) in order to "unlock" the smart token for use.

There are many different types of smart tokens. In general, smart tokens can be divided three different ways based on physical characteristics, interface, and protocols used. These three divisions are not mutually exclusive.

Physical Characteristics. Smart tokens can be divided into two groups: smart cards and other types of tokens. A smart card looks like a credit card, but incorporates an embedded microprocessor. Smart cards are defined by an International Standards Organization (ISO) standard. Smart tokens that are not smart cards can look like calculators, keys, or other small portable objects.

Interface. Smart tokens have either a manual or an electronic interface. Manual or human interface tokens have displays and/or keypads to allow humans to communicate with the card. Smart tokens with electronic interfaces must be read by special reader/writers. Smart cards, described above, have an electronic interface. Smart tokens that look like calculators usually have a manual interface.

Protocol. There are many possible protocols a smart token can use for authentication. In general, they can be divided into three categories: static password exchange, dynamic password generators, and challenge-response.

- *Static* tokens work similarly to memory tokens, except that the users authenticate themselves *to the token* and then the token authenticates the user to the computer.

- A token that uses a *dynamic password generator* protocol creates a unique value, for example, an eight-digit number, that changes periodically (e.g., every minute). If the token has a manual interface, the user simply reads the current value and then types it into the computer system for authentication. If the token has an electronic interface, the transfer is done automatically. If the correct value is provided, the log-in is permitted, and the user is granted access to the system.

- Tokens that use a *challenge-response* protocol work by having the computer generate a challenge, such as a random string of numbers. The smart token then generates a response based on the challenge. This is sent back to the computer, which authenticates the user based on the response. The challenge-response protocol is based on cryptography. Challenge-response tokens can use either electronic or manual interfaces.

There are other types of protocols, some more sophisticated and some less so. The three types described above are the most common.

Benefits of Smart Tokens

Smart tokens offer great flexibility and can be used to solve many authentication problems. The benefits of smart tokens vary, depending on the type used. In general, they provide greater security than memory cards. Smart tokens can solve the problem of electronic monitoring even if the authentication is done across an open network by using *one-time passwords.*

1. One-time passwords. Smart tokens that use either dynamic password generation or challenge-response protocols can create one-time passwords. Electronic monitoring is not a problem with one-time passwords because each time the user is authenticated to the computer, a different "password" is used. (A hacker could learn the one-time password through electronic monitoring, but would be of no value.)

2. Reduced risk of forgery. Generally, the memory on a smart token is not readable unless the PIN is entered. In addition, the tokens are more complex and, therefore, more difficult to forge.

3. Multi-application. Smart tokens with electronic interfaces, such as smart cards, provide a way for users to access many computers using many networks with only one log-in. This is further discussed in the Single Log-in section of this chapter. In addition, a single smart card can be used for multiple functions, such as physical access or as a debit card.

Problems with Smart Tokens

Like memory tokens, most of the problems associated with smart tokens relate to their cost, the administration of the system, and user dissatisfaction. Smart tokens are generally less vulnerable to the compromise of PINs because authentication usually takes place on the card. (It is possible, of course, for someone to watch a PIN being entered and steal that card.) Smart tokens cost more than memory cards because they are more complex, particularly challenge-response calculators.

1. *Need reader/writers or human intervention.* Smart tokens can use either an electronic or a human interface. An electronic interface requires a reader, which creates additional expense. Human interfaces require more actions from the user. This is especially true for challenge-response tokens with a manual interface, which require the user to type the challenge into the smart token and the response into the computer. This can increase user dissatisfaction.

Electronic reader/writers can take many forms, such as a slot in a PC or a separate external device. Most human interfaces consist of a keypad and display.

2. *Substantial Administration.* Smart tokens, like passwords and memory tokens, require strong administration. For tokens that use cryptography, this includes key management. (See Chapter 19.)

16.3 I&A Based on Something the User Is

Biometric authentication technologies use the unique characteristics (or attributes) of an individual to authenticate that person's identity. These include physiological attributes (such as fingerprints, hand geometry, or retina patterns) or behavioral attributes (such as voice patterns and hand-written signatures). Biometric authentication technologies based upon these attributes have been developed for computer log-in applications.

Biometric authentication is technically complex and expensive, and user acceptance can be difficult. However, advances continue to be made to make the technology more reliable, less costly, and more user-friendly.

Biometric authentication generally operates in the following manner:

Before any authentication attempts, a user is "enrolled" by creating a reference profile (or template) based on the desired physical attribute. The resulting template is associated with the identity of the user and stored for later use.

When attempting authentication, the user's biometric attribute is measured. The previously stored reference profile of the biometric attribute is compared with the measured profile of the attribute taken from the user. The result of the comparison is then used to either accept or reject the user.

Biometric systems can provide an increased level of security for computer systems, but the technology is still less mature than that of memory tokens or smart tokens. Imperfections in biometric authentication devices arise from technical difficulties in measuring and profiling physical attributes as well as from the somewhat variable nature of physical attributes. These may change, depending on various conditions. For example, a person's speech pattern may change under stressful conditions or when suffering from a sore throat or cold.

Due to their relatively high cost, biometric systems are typically used with other authentication means in environments requiring high security.

16.4 Implementing I&A Systems

Some of the important implementation issues for I&A systems include administration, maintaining authentication, and single log-in.

16.4.1 Administration

Administration of authentication data is a critical element for all types of authentication systems. The administrative overhead associated with I&A can be significant. I&A systems need to create, distribute, and store authentication data. For passwords, this includes creating passwords, issuing them to users, and maintaining a password file. Token systems involve the creation and distribution of tokens/PINs and data that tell the computer how to recognize valid tokens/PINs. For biometric systems, this includes creating and storing profiles.

The administrative tasks of creating and distributing authentication data and tokens can be a substantial. Identification data has to be kept current by adding new users and deleting former users. If the distribution of passwords or tokens is not controlled, system administrators will not know if they have been given to someone other than the legitimate user. It is critical that the distribution system ensure that authentication data is firmly linked with a given individual. Some of these issues are discussed in Chapter 10 under User Administration.

In addition, I&A administrative tasks should address lost or stolen passwords or tokens. It is often necessary to monitor systems to look for stolen or shared accounts.

One method of looking for improperly used accounts is for the computer to inform users when they last logged on. This allows users to check if someone else used their account.

Authentication data needs to be stored securely, as discussed with regard to accessing password files. The value of authentication data lies in the data's confidentiality, integrity, and availability. If confidentiality is compromised, someone may be able to use the information to masquerade as a legitimate user. If system administrators can read the authentication file, they can masquerade as another user. Many systems use encryption to hide the authentication data from the system administrators.[111] If integrity is compromised, authentication data can be added or the system can be disrupted. If availability is compromised, the system cannot authenticate users, and the users may not be able to work.

16.4.2 Maintaining Authentication

So far, this chapter has discussed initial authentication only. It is also possible for someone to use a legitimate user's account after log-in.[112] Many computer systems handle this problem by logging a user out or locking their display or session after a certain period of inactivity. However, these methods can affect productivity and can make the computer less user-friendly.

16.4.3 Single Log-in

From an efficiency viewpoint, it is desirable for users to authenticate themselves only once and then to be able to access a wide variety of applications and data available on local and remote systems, even if those systems require users to authenticate themselves. This is known as *single log-in*.[113] If the access is within the same host computer, then the use of a modern access control system (such as an access control list) should allow for a single log-in. If the access is across multiple platforms, then the issue is more complicated, as discussed below. There are three main techniques that can provide single log-in across multiple computers: host-to-host authentication, authentication servers, and user-to-host authentication.

Host-to-Host Authentication. Under a host-to-host authentication approach, users authenticate themselves once to a host computer. That computer then authenticates itself to other computers and vouches for the specific user. Host-to-host authentication can be done by passing an identification, a password, or by a challenge-response mechanism or other one-time password scheme. Under this approach, it is necessary for the computers to recognize each other and to trust each other.

Authentication Servers. When using authentication server, the users authenticate themselves to a special host computer (the authentication server). This computer then authenticates the user to

[111] Masquerading *by* system administrators cannot be prevented entirely. However, controls can be set up so that improper actions by the system administrator can be detected in audit records.

[112] After a user signs on, the computer treats all commands originating from the user's physical device (such as a PC or terminal) as being from that user.

[113] Single log-in is somewhat of a misnomer. It is currently not feasible to have one sign-on for every computer system a user might wish to access. The types of single log-in described apply mainly to groups of systems (e.g., within an organization or a consortium).

other host computers the user wants to access. Under this approach, it is necessary for the computers to trust the authentication server. (The authentication server need not be a separate computer, although in some environments this may be a cost-effective way to increase the security of the server.) Authentication servers can be distributed geographically or logically, as needed, to reduce workload.

Kerberos and SPX are examples of network authentication server protocols. They both use cryptography to authenticate users to computers on networks.

User-to-Host. A user-to-host authentication approach requires the user to log-in to each host computer. However, a smart token (such as a smart card) can contain all authentication data and perform that service for the user. To users, it looks as though they were only authenticated once.

16.5 Interdependencies

There are many interdependencies among I&A and other controls. Several of them have been discussed in the chapter.

Logical Access Controls. Access controls are needed to protect the authentication database. I&A is often the basis for access controls. Dial-back modems and firewalls, discussed in Chapter 17, can help prevent hackers from trying to log-in.

Audit. I&A is necessary if an audit log is going to be used for individual accountability.

Cryptography. Cryptography provides two basic services to I&A: it protects the confidentiality of authentication data, and it provides protocols for proving knowledge and/or possession of a token without having to transmit data that could be replayed to gain access to a computer system.

16.6 Cost Considerations

In general, passwords are the least expensive authentication technique and generally the least secure. They are already embedded in many systems. Memory tokens are less expensive than smart tokens, but have less functionality. Smart tokens with a human interface do not require readers, but are more inconvenient to use. Biometrics tend to be the most expensive.

For I&A systems, the cost of administration is often underestimated. Just because a system comes with a password system does not mean that using it is free. For example, there is significant overhead to administering the I&A system.

References

Alexander, M., ed. "Keeping the Bad Guys Off-Line." *Infosecurity News.* 4(6), 1993. pp. 54-65.

American Bankers Association. *American National Standard for Financial Institution Sign-On Authentication for Wholesale Financial Transactions.* ANSI X9.26-1990. Washington, DC, February 28, 1990.

CCITT Recommendation X.509. The Directory - Authentication Framework. November 1988 (Developed in collaboration, and technically aligned, with ISO 9594-8).

Department of Defense. Password Management Guideline. CSC-STD-002-85. April 12, 1985.

Feldmeier, David C., and Philip R. Kam. "UNIX Password Security - Ten Years Later." *Crypto '89 Abstracts.* Santa Barbara, CA: Crypto '89 Conference, August 20-24, 1989.

Haykin, Martha E., and Robert B. J. Warnar. *Smart Card Technology: New Methods for Computer Access Control.* Special Publication 500-157. Gaithersburg, MD: National Institute of Standards and Technology, September 1988.

Kay, R. "Whatever Happened to Biometrics?" *Infosecurity News.* 4(5), 1993. pp. 60-62.

National Bureau of Standards. *Password Usage.* Federal Information Processing Standard Publication 112. May 30, 1985.

National Institute of Standards and Technology. *Automated Password Generator.* Federal Information Processing Standard Publication 181. October, 1993.

National Institute of Standards and Technology. *Guideline for the Use of Advanced Authentication Technology Alternatives.* Federal Information Processing Standard Publication 190. October, 1994.

Salamone, S. "Internetwork Security: Unsafe at Any Node?" *Data Communications.* 22(12), 1993. pp. 61-68.

Sherman, R. "Biometric Futures." *Computers and Security.* 11(2), 1992. pp. 128-133.

Smid, Miles, James Dray, and Robert B. J. Warnar. "A Token-Based Access Control System for Computer Networks." *Proceedings of the 12th National Commuter Security Conference.* National Institute of Standards and Technology, October 1989.

Steiner, J.O., C. Neuman, and J. Schiller. "Kerberos: An Authentication Service for Open Network Systems." *Proceedings Winter USENIX*. Dallas, Texas, February 1988. pp. 191-202.

Troy, Eugene F. *Security for Dial-Up Lines*. Special Publication 500-137, Gaithersburg, MD: National Bureau of Standards, May 1986.

Chapter 17

LOGICAL ACCESS CONTROL

Logical access controls provide a technical means of controlling what information users can utilize, the programs they can run, and the modifications they can make.

On many multiuser systems, requirements for using (and prohibitions against the use of) various computer resources[114] vary considerably. Typically, for example, some information must be accessible to all users,[115] some may be needed by several groups or departments, and some should be accessed by only a few individuals. While it is obvious that users must have access to the information they need to do their jobs, it may also be required to deny access to non-job-related information. It may also be important to control the *kind of access* that is afforded (e.g., the ability for the average user to execute, but not change, system programs). These types of access restrictions enforce policy and help ensure that unauthorized actions are not taken.

The term *access* is often confused with *authorization* and *authentication.*

Access is the *ability* to do something with a computer resource. This usually refers to a technical ability (e.g., read, create, modify, or delete a file, execute a program, or use an external connection).

Authorization is the *permission* to use a computer resource. Permission is granted, directly or indirectly, by the application or system owner.

Authentication is proving (to some reasonable degree) that users are who they claim to be.

Access is the ability to do something with a computer resource (e.g., use, change, or view). *Access control* is the means by which the ability is explicitly enabled or restricted in some way (usually through physical and system-based controls). Computer-based access controls are called *logical access controls.* Logical access controls can prescribe not only who or what (e.g., in the case of a process) is to have access to a specific system resource but also the type of access that is permitted. These controls may be built into the operating system, may be incorporated into applications programs or major utilities (e.g., database management systems or communications systems), or may be implemented through add-on security packages. Logical access controls may be implemented internally to the computer system being protected or may be implemented in external devices.

[114] The term *computer resources* includes information as well as system resources, such as programs, subroutines, and hardware (e.g., modems, communications lines).

[115] *Users* need not be actual human users. They could include, for example, a program or another computer requesting use of a system resource.

Logical access controls can help protect:

- operating systems and other system software from unauthorized modification or manipulation (and thereby help ensure the system's integrity and availability);
- the integrity and availability of information by restricting the number of users and processes with access; and
- confidential information from being disclosed to unauthorized individuals.

Controlling access is normally thought of as applying to human users (e.g., will technical access be provided for user JSMITH to the file "payroll.dat") but access can be provided to other computer systems. Also, access controls are often incorrectly thought of as only applying to *files*. However, they also protect other system resources such as the ability to place an outgoing long-distance phone call though a system modem (as well as, perhaps, the information that can be sent over such a call). Access controls can also apply to specific functions within an application and to specific fields of a file.

This chapter first discusses basic criteria that can be used to decide whether a particular user should be granted access to a particular system resource. It then reviews the use of these criteria by those who set policy (usually system-specific policy), commonly used *technical mechanisms* for implementing logical access control, and issues related to administration of access controls.

17.1 Access Criteria

In deciding whether to permit someone to use a system resource logical access controls examine whether *the user is authorized for the type of access requested.* (Note that this inquiry is usually distinct from the question of whether the user is authorized to use the system *at all*, which is usually addressed in an identification and authentication process.)

The system uses various criteria to determine if a request for access will be granted. They are typically used in some combination. Many of the advantages and complexities involved in implementing and managing access control are related to the different kinds of user accesses supported.

When determining what kind of technical access to allow to specific data, programs, devices, and resources, it is important to consider who will have access and what kind of access they will be allowed. It may be desirable for everyone in the organization to have access to some information on the system, such as the data displayed on an organization's daily calendar of nonconfidential meetings. The program that formats and displays the calendar, however, might be modifiable by only a very few system administrators, while the operating system controlling that program might be directly accessible by still fewer.

17.1.1 Identity

It is probably fair to say that the majority of access controls are based upon the identity of the user (either human or process), which is usually obtained through identification and authentication (I&A). (See Chapter 16.) The identity is usually unique, to support individual accountability, but can be a group identification or can even be anonymous. For example, public information dissemination systems may serve a large group called "researchers" in which the individual researchers are not known.

17.1.2 Roles

Access to information may also be controlled by the job assignment or function (i.e., the *role*) of the user who is seeking access. Examples of roles include data entry clerk, purchase officer, project leader, programmer, and technical editor. Access rights are grouped by role name, and the use of resources is restricted to individuals authorized to assume the associated role. An individual may be authorized for more than one role, but may be required to act in only a single role at a time. Changing roles may require logging out and then in again, or entering a role-changing command. Note that use of roles is *not* the same as shared-use accounts. An individual may be assigned a standard set of rights of a shipping department data entry clerk, for example, but the account would still be tied to that individual's identity to allow for auditing. (See Chapter 18.)

Many systems already support a small number of special-purpose roles, such as System Administrator or Operator. For example, an individual who is logged on in the role of a System Administrator can perform operations that would be denied to the same individual acting in the role of an ordinary user.

Recently, the use of roles has been expanded beyond system tasks to application-oriented activities. For example, a user in a company could have an Order Taking role, and would be able to collect and enter customer billing information, check on availability of particular items, request shipment of items, and issue invoices. In addition, there could be an Accounts Receivable role, which would receive payments and credit them to particular invoices. A Shipping role, could then be responsible for shipping products and updating the inventory. To provide additional security, constraints could be imposed so a single user would never be simultaneously authorized to assume all three roles. Constraints of this kind are sometimes referred to as *separation of duty constraints.*

The use of roles can be a very effective way of providing access control. The process of defining roles should be based on a thorough analysis of how an organization operates and should include input from a wide spectrum of users in an organization.

17.1.3 Location

Access to particular system resources may also be based upon physical or logical location. For example, in a prison, all users in areas to which prisoners are physically permitted may be limited to read-only access. Changing or deleting is limited to areas to which prisoners are denied physical access. The same authorized users (e.g., prison guards) would operate under significantly different logical access controls, depending upon their physical location. Similarly, users can be restricted based upon network addresses (e.g., users from sites within a given organization may be permitted greater access than those from outside).

17.1.4 Time

Time-of-day or day-of-week restrictions are common limitations on access. For example, use of confidential personnel files may be allowed only during normal working hours – and maybe denied before 8:00 a.m. and after 6:00 p.m. and all day during weekends and holidays.

17.1.5 Transaction

Another approach to access control can be used by organizations handling transactions (e.g., account inquiries). Phone calls may first be answered by a computer that requests that callers key in their account number and perhaps a PIN. Some routine transactions can then be made directly, but more complex ones may require human intervention. In such cases, the computer, which already knows the account number, can grant a clerk, for example, access to a particular account *for the duration of the transaction.* When completed, the access authorization is terminated. This means that users have no choice in which accounts they have access to, and can reduce the potential for mischief. It also eliminates employee browsing of accounts (e.g., those of celebrities or their neighbors) and can thereby heighten privacy.

17.1.6 Service Constraints

Service constraints refer to those restrictions that depend upon the parameters that may arise during use of the application or that are preestablished by the resource owner/manager. For example, a particular software package may only be licensed by the organization for five users at a time. Access would be denied for a sixth user, even if the user were otherwise authorized to use the application. Another type of service constraint is based upon application content or numerical thresholds. For example, an ATM machine may restrict transfers of money between accounts to certain dollar limits or may limit maximum ATM withdrawals to $500 per day. Access may also be selectively permitted based on the type of service requested. For example, users of computers on a network may be permitted to exchange electronic mail but may not be allowed to log in to each others' computers.

17.1.7 Common Access Modes

In addition to considering criteria for *when* access should occur, it is also necessary to consider the *types* of access, or *access modes.* The concept of access modes is fundamental to access control. Common access modes, which can be used in both operating or application systems, include the following:[116]

[116] These access modes are described generically; exact definitions and capabilities will vary from implementation to implementation. Readers are advised to consult their system and application documentation.

Read access provides users with the capability to view information in a system resource (such as a file, certain records, certain fields, or some combination thereof), but not to *alter* it, such as delete from, add to, or modify in any way. One must assume that information can be copied and printed if it can be read (although perhaps only manually, such as by using a print screen function and retyping the information into another file).

Write access allows users to add to, modify, or delete information in system resources (e.g., files, records, programs). Normally user have read access to anything they have write access to.

Execute privilege allows users to run programs.

Delete access allows users to erase system resources (e.g., files, records, fields, programs).[117] Note that if users have write access but not delete access, they could overwrite the field or file with gibberish or otherwise inaccurate information and, in effect, delete the information.

Other specialized access modes (more often found in applications) include:

Create access allows users to create new files, records, or fields.

Search access allows users to list the files in a directory.

Of course, these criteria can be used in conjunction with one another. For example, an organization may give authorized individuals write access to an application at any time from within the office but only read access during normal working hours if they dial-in.

Depending upon the technical mechanisms available to implement logical access control, a wide variety of access permissions and restrictions are possible. No discussion can present all possibilities.

17.2 Policy: The Impetus for Access Controls

Logical access controls are a technical means of implementing *policy decisions*. Policy is made by a management official responsible for a particular system, application, subsystem, or group of systems. The development of an access control policy may not be an easy endeavor. It requires balancing the often-competing interests of security, operational requirements, and user-friendliness. In addition, technical constraints have to be considered.

[117] "Deleting" information does not necessarily physically remove the data from the storage media. This can have serious implications for information that must be kept confidential. See "Disposition of Sensitive Automated Information," CSL Bulletin, NIST, October 1992.

This chapter discusses issues relating to the technical implementation of logical access controls – not the actual policy decisions as to who *should* have what type of access. These decisions are typically included in system-specific policy, as discussed in Chapters 5 and 10.

Once these policy decisions have been made, they will be *implemented* (or *enforced*) through logical access controls. In doing so, it is important to realize that the capabilities of various types of technical mechanisms (for logical access control) vary greatly.[118]

A few *simple* examples of *specific policy issues* are provided below; it is important to recognize, however, that *comprehensive system-specific policy* is significantly more complex.

1. The director of an organization's personnel office could decide that all clerks can update all files, to increase the efficiency of the office. Or the director could decide that clerks can only view and update specific files, to help prevent information browsing.

2. In a disbursing office, a single individual is usually prohibited from both requesting and authorizing that a particular payment be made. This is a *policy decision* taken to reduce the likelihood of embezzlement and fraud.

3. Decisions may also be made regarding access to the system itself. In the government, for example, the senior information resources management official may decide that agency systems that process information protected by the Privacy Act may not be used to process public-access database applications.

17.3 Technical Implementation Mechanisms

Many mechanisms have been developed to provide internal and external access controls, and they vary significantly in terms of precision, sophistication, and cost. These methods are not mutually exclusive and are often employed in combination. Managers need to analyze their organization's protection requirements to select the most appropriate, cost-effective logical access controls.

17.3.1 Internal Access Controls

Internal access controls are a logical means of separating what defined users (or user groups) can or cannot do with system resources. Five methods of internal access control are discussed in this section: passwords, encryption, access control lists, constrained user interfaces, and labels.

17.3.1.1 Passwords

Passwords are most often associated with user authentication. (See Chapter 16.) However, they are also used to protect data and applications on many systems, including PCs. For instance, an accounting application may require a password to access certain financial data or to invoke a

[118] Some policies may not be technically implementable; appropriate technical controls may simply not exist.

restricted application (or function of an application).[119]

The use of passwords as a means of access control can result in a proliferation of passwords that can reduce overall security.

Password-based access control is often inexpensive because it is already included in a large variety of applications. However, users may find it difficult to remember additional application passwords, which, if written down or poorly chosen, can lead to their compromise. Password-based access controls for PC applications are often easy to circumvent if the user has access to the operating system (and knowledge of what to do). As discussed in Chapter 16, there are other disadvantages to using passwords.

17.3.1.2 Encryption

Another mechanism that can be used for logical access control is encryption. Encrypted information can only be decrypted by those possessing the appropriate cryptographic key. This is especially useful if strong physical access controls cannot be provided, such as for laptops or floppy diskettes. Thus, for example, if information is encrypted on a laptop computer, and the laptop is stolen, the information cannot be accessed. While encryption can provide strong access control, it is accompanied by the need for strong key management. Use of encryption may also affect availability. For example, lost or stolen keys or read/write errors may prevent the decryption of the information. (See the cryptography chapter.)

17.3.1.3 Access Control Lists

Access Control Lists (ACLs) refer to a register of: (1) users (including groups, machines, processes) who have been given permission to use a particular system resource, and (2) the types of access they have been permitted.

ACLs vary considerably in their capability and flexibility. Some only allow specifications for certain pre-set groups (e.g., owner, group, and world) while more advanced ACLs allow much more flexibility, such as *user-defined* groups. Also, more advanced ACLs can be used to explicitly *deny* access to a particular individual or group. With more advanced ACLs, access can be at the discretion of the policymaker (and implemented by the security administrator) or individual user, depending upon how the controls are technically implemented.

Elementary ACLs. Elementary ACLs (e.g., "permission bits") are a widely available means of providing access control on multiuser systems. In this scheme, a short, predefined list of the access rights to files or other system resources is maintained.

[119] Note that this password is normally *in addition* to the one supplied initially to log onto the system.

Elementary ACLs are typically based on the concepts of *owner, group,* and *world.* For each of these, a set of access modes (typically chosen from read, write, execute, and delete) is specified by the owner (or custodian) of the resource. The owner is usually its creator, though in some cases, ownership of resources may be automatically assigned to project administrators, regardless of the identity of the creator. File owners often have all privileges for their resources.

Example of Elementary ACL for the file "payroll":

Owner: PAYMANAGER
Access: Read, Write, Execute, Delete

Group: COMPENSATION-OFFICE
Access: Read, Write, Execute, Delete

"World"
Access: None

In addition to the privileges assigned to the owner, each resource is associated with *a named group of users.* Users who are members of the group can be granted modes of access distinct from nonmembers, who belong to the rest of the "world" that includes all of the system's users. User groups may be arranged according to departments, projects, or other ways appropriate for the particular organization. For example, groups may be established for members of the Personnel and Accounting departments. The system administrator is normally responsible for technically maintaining and changing the membership of a group, based upon input from the owners/custodians of the particular resources to which the groups may be granted access.

As the name implies, however, the technology is not particularly flexible. It may not be possible to explicitly deny access to an individual who is a member of the file's group. Also, it may not be possible for two groups to easily share information (without exposing it to the "world"), since the list is predefined to only include one group. If two groups wish to share information, an owner may make the file available to be read by "world." This may disclose information that should be restricted. Unfortunately, elementary ACLs have no mechanism to easily permit such sharing.

Since one would presume that no one would have access without being granted access, why would it be desirable to explicitly deny access? Consider a situation in which a group name has already been established for 50 employees. If it were desired to exclude five of the individuals from that group, it would be easier for the access control administrator to simply grant access to that group and take it away from the five rather than grant access to 45 people. Or, consider the case of a complex application in which many groups of users are defined. It may be desired, for some reason, to prohibit Ms. *X* from generating a particular report (perhaps she is under investigation). In a situation in which group names are used (and perhaps modified by others), this explicit denial may be a safety check to restrict Ms. *X*'s access – in case someone were to redefine a group (with access to the report generation function) to include Ms. *X*. She would still be denied access.

Advanced ACLs. Like elementary ACLs, advanced ACLs provide a form of access control based upon a logical registry. They do, however, provide *finer precision* in control.

Advanced ACLs can be very useful in many complex information sharing situations. They provide a great deal of flexibility in implementing system-specific policy and allow for customization to meet the security requirements of functional managers. Their flexibility also makes them more of a challenge to manage. The rules for determining access in the face of apparently conflicting ACL entries are not uniform across all implementations and can be confusing to security administrators. When such systems are introduced, they should be coupled with training to ensure their correct use.

Example of Advanced ACL for the file "payroll"

PAYMGR:	R,	W,	E,	D
J. Anderson:	R,	W,	E,	-
L. Carnahan:	-,	-,	-,	-
B. Guttman:	R,	W,	E,	-
E. Roback:	R,	W,	E,	-
H. Smith:	R,	-,	-,	-
PAY-OFFICE:	R,	-,	-,	-
WORLD:	-,	-,	-,	-

17.3.1.4 Constrained User Interfaces

Often used in conjunction with ACLs are *constrained user interfaces*, which restrict users' access to specific functions by never allowing them to request the use of information, functions, or other specific system resources for which they do not have access. Three major types exist: (1) *menus*, (2) *database views*, and (3) *physically constrained user interfaces.*

Constrained user interfaces can provide a form of access control that closely models how an organization operates. Many systems allow administrators to restrict users' ability to use the operating system or application system directly. Users can only execute commands that are provided by the administrator, typically in the form of a *menu*. Another means of restricting users is through restricted *shells* which limit the system commands the user can invoke. The use of menus and shells can often make the system easier to use and can help reduce errors.

Menu-driven systems are a common constrained user interface, where different users are provided different menus on the same system.

Database views is a mechanism for restricting user access to data contained in a database. It may be necessary to allow a user to access a database, but that user may not need access to all the data in the database (e.g., not all fields of a record nor all records in the database). Views can be used to enforce complex access requirements that are often needed in database situations, such as those based on the content of a field. For example, consider the situation where clerks maintain personnel records in a database. Clerks are assigned a range of clients based upon last name (e.g., A-C, D-G). Instead of granting a user access to all records, the view can grant the user access to the record based upon the first letter of the last name field.

Physically constrained user interfaces can also limit a user's abilities. A common example is an ATM machine, which provides only a limited number of physical buttons to select options; no alphabetic keyboard is usually present.

17.3.1.5 Security Labels

A security label is a designation assigned to a resource (such as a file). Labels can be used for a variety of purposes, including controlling access, specifying protective measures, or indicating additional handling instructions. In many implementations, once this designator has been set, it cannot be changed (except perhaps under carefully controlled conditions that are subject to auditing).

Data Categorization

One tool that is used to increase the ease of security labelling is categorizing data by similar protection requirements. For example, a label could be developed for "organization proprietary data." This label would mark information that can be disclosed only to the organization's employees. Another label, "public data" could be used to mark information that is available to anyone.

When used for access control, labels are also assigned to *user sessions*. Users are permitted to initiate sessions with specific labels only. For example, a file bearing the label "Organization Proprietary Information" would not be accessible (readable) except during user sessions with the corresponding label. Moreover, only a restricted set of users would be able to initiate such sessions. The labels of the session and those of the files accessed during the session are used, in turn, to label output from the session. This ensures that information is uniformly protected throughout its life on the system.

Labels are a very strong form of access control; however, they are often inflexible and can be expensive to administer. Unlike permission bits or access control lists, labels cannot ordinarily be changed. Since labels are permanently linked to specific information, data cannot be disclosed by a user copying information and changing the access to that file so that the information is more accessible than the original owner intended. By removing users' ability to arbitrarily designate the accessibility of files they own, opportunities for certain kinds of human errors and malicious software problems are eliminated. In the example above, it would not be possible to copy Organization Proprietary Information into a file with a different label. This prevents inappropriate disclosure, but can interfere with legitimate extraction of some information.

For systems with stringent security requirements (such as those processing national security information), labels may be useful in access control.

Labels are well suited for consistently and uniformly enforcing access restrictions, although their administration and inflexibility can be a significant deterrent to their use.

17.3.2 External Access Controls

External access controls are a means of controlling interactions between the system and outside people, systems, and services. External access controls use a wide variety of methods, often including a separate physical device (e.g., a computer) that is between the system being protected and a network.

17.3.2.1 Port Protection Devices

Fitted to a communications port of a host computer, a port protection device (PPD) authorizes access to the port itself, prior to and independent of the computer's own access control functions. A PPD can be a separate device in the communications stream,[120] or it may be incorporated into a communications device (e.g., a modem). PPDs typically require a separate authenticator, such as a password, in order to access the communications port.

One of the most common PPDs is the *dial-back modem*. A typical dial-back modem sequence follows: a user calls the dial-back modem and enters a password. The modem hangs up on the user and performs a table lookup for the password provided. If the password is found, the modem places a return call to the user (at a previously specified number) to initiate the session. The return call itself also helps to protect against the use of lost or compromised accounts. This is, however, not always the case. Malicious hackers can use such advance functions as call forwarding to reroute calls.

17.3.2.2 Secure Gateways/Firewalls

Often called *firewalls*, secure gateways block or filter access between two networks, often between a private[121] network and a larger, more public network such as the Internet, which attract malicious hackers. Secure gateways allow internal users to connect to external networks and at the same time prevent malicious hackers from compromising the internal systems.[122]

Some secure gateways are set up to allow all traffic to pass through except for specific traffic which has known or suspected vulnerabilities or security problems, such as remote log-in services. Other secure gateways are set up to disallow all traffic except for specific types, such as e-mail. Some secure gateways can make access-control decisions based on the location of the requester. There are several technical approaches and mechanisms used to support secure gateways.

[120] Typically PPDs are found only in serial communications streams.

[121] *Private network* is somewhat of a misnomer. *Private* does not mean that the organization's network is totally inaccessible to outsiders or prohibits use of the outside network from insiders (or the network would be disconnected). It also does not mean that all the information on the network requires confidentiality protection. It does mean that a network (or part of a network) is, in some way, separated from another network.

[122] Questions frequently arise as to whether secure gateways help prevent the spread of viruses. In general, having a gateway scan transmitted files for viruses requires more system overhead than is practical, especially since the scanning would have to handle many different file formats. However, secure gateways may reduce the spread of network worms.

Because gateways provide security by restricting services or traffic, they can affect a system's usage. For this reason, firewall experts always emphasize the need for policy, so that appropriate officials decide how the organization will balance operational needs and security.

Types of Secure Gateways

There are many types of secure gateways. Some of the most common are packet filtering (or screening) routers, proxy hosts, bastion hosts, dual-homed gateways, and screened-host gateways.

In addition to reducing the risks from malicious hackers, secure gateways have several other benefits. They can reduce internal system security overhead, since they allow an organization to concentrate security efforts on a limited number of machines. (This is similar to putting a guard on the first floor of a building instead of needing a guard on every floor.)

A second benefit is the centralization of services. A secure gateway can be used to provide a central management point for various services, such as advanced authentication (discussed in Chapter 16), e-mail, or public dissemination of information. Having a central management point can reduce system overhead and improve service.

17.3.2.3 Host-Based Authentication

Host-based authentication grants access based upon the *identity of the host* originating the request, instead of the identity of the user making the request. Many network applications in use today use host-based authentication to determine whether access is allowed. Under certain circumstances it is fairly easy to masquerade as the legitimate host, especially if the masquerading host is physically located close to the host being impersonated. Security measures to protect against misuse of some host-based authentication systems are available (e.g., Secure RPC[123] uses DES to provide a more secure identification of the client host).

An example of host-based authentication is the Network File System (NFS) which allows a server to make file systems/directories available to specific machines.

17.4 Administration of Access Controls

One of the most complex and challenging aspects of access control, administration involves implementing, monitoring, modifying, testing, and terminating user accesses on the system. These can be demanding tasks, even though they typically do not include making the actual decisions as

[123] RPC, or Remote Procedure Call, is the service used to implement NFS.

to the type of access each user may have.[124] Decisions regarding accesses should be guided by organizational policy, employee job descriptions and tasks, information sensitivity, user "need-to-know" determinations, and many other factors.

There are three basic approaches to administering access controls: centralized, decentralized, or a combination of these. Each has relative advantages and disadvantages. Which is most appropriate in a given situation will depend upon the particular organization and its circumstances.

System and Security Administration

The administration of systems and security requires access to advanced functions (such as setting up a user account). The individuals who technically set up and modify who has access to what are very powerful users on the system; they are often called system or security administrators. On some systems, these users are referred to as having *privileged accounts.*

The type of access of these accounts varies considerably. Some administrator privileges, for example, may allow an individual to administer only one application or subsystem, while a higher level of privileges may allow for oversight and establishment of subsystem administrators.

Normally, users who are security administrators have two accounts: one for regular use and one for security use. This can help protect the security account from compromise. Furthermore, additional I&A precautions, such as ensuring that administrator passwords are robust and changed regularly, are important to minimize opportunities for unauthorized individuals to gain access to these functions.

17.4.1 Centralized Administration

Using centralized administration, one office or individual is responsible for configuring access controls. As users' information processing needs change, their accesses can be modified only through the central office, usually after requests have been approved by the appropriate official. This allows very strict control over information, because the ability to make changes resides with very few individuals. Each user's account can be centrally monitored, and closing all accesses for any user can be easily accomplished if that individual leaves the organization. Since relatively few individuals oversee the process, consistent and uniform procedures and criteria are usually not difficult to enforce. However, when changes are needed quickly, going through a central administration office can be frustrating and time-consuming.

17.4.2 Decentralized Administration

In decentralized administration, access is directly controlled by the owners or creators of the files, often the functional manager. This keeps control in the hands of those most accountable for the information, most familiar with it and its uses, and best able to judge who needs what kind of

[124] As discussed in the policy section earlier in this chapter, those decisions are usually the responsibility of the applicable application manager or cognizant management official. See also the discussion of system-specific policy in Chapters 5 and 10.

access. This may lead, however, to a lack of consistency among owners/creators as to procedures and criteria for granting user accesses and capabilities. Also, when requests are not processed centrally, it may be much more difficult to form a systemwide composite view of all user accesses on the system at any given time. Different application or data owners may inadvertently implement combinations of accesses that introduce conflicts of interest or that are in some other way not in the organization's best interest.[125] It may also be difficult to ensure that all accesses are properly terminated when an employee transfers internally or leaves an organization.

17.4.3 Hybrid Approach

A hybrid approach combines centralized and decentralized administration. One typical arrangement is that central administration is responsible for the broadest and most basic accesses, and the owners/creators of files control types of accesses or changes in users' abilities for the files under their control. The main disadvantage to a hybrid approach is adequately defining which accesses should be assignable locally and which should be assignable centrally.

17.5 Coordinating Access Controls

It is vital that access controls protecting a system work together. At a minimum, three basic types of access controls should be considered: physical, operating system, and application. In general, access controls within an application are the most specific. However, for application access controls to be fully effective they need to be supported by operating system access controls. Otherwise access can be made to application resources without going through the application.[126] Operating system and application access controls need to be supported by physical access controls.

17.6 Interdependencies

Logical access controls are closely related to many other controls. Several of them have been discussed in the chapter.

Policy and Personnel. The most fundamental interdependencies of logical access control are with policy and personnel. Logical access controls are the technical implementation of system-specific and organizational policy, which stipulates *who* should be able to access what kinds of information, applications, and functions. These decisions are normally based on the principles of

[125] Without necessary review mechanisms, central administration does not *a priori* preclude this.

[126] For example, logical access controls within an application block User A from viewing File F. However, if operating systems access controls do not also block User A from viewing File F, User A can use a utility program (or another application) to view the file.

separation of duties and least privilege.

Audit Trails. As discussed earlier, logical access controls can be difficult to implement correctly. Also, it is sometimes *not possible* to make logical access control as precise, or fine-grained, as would be ideal for an organization. In such situations, users may either deliberately or inadvertently abuse their access. For example, access controls cannot prevent a user from modifying data the user is authorized to modify, even if the modification is incorrect. Auditing provides a way to identify abuse of access permissions. It also provides a means to review the actions of system or security administrators.

Identification and Authentication. In most logical access control scenarios, the identity of the user must be established before an access control decision can be made. The access control process then associates the permissible forms of accesses with that identity. This means that access control can only be as effective as the I&A process employed for the system.

Physical Access Control. Most systems can be compromised if someone can physically access the machine (i.e., CPU or other major components) by, for example, restarting the system with different software. Logical access controls are, therefore, dependent on physical access controls (with the exception of encryption, which can depend solely on the strength of the algorithm and the secrecy of the key).

17.7 Cost Considerations

Incorporating logical access controls into a computer system involves the purchase or use of access control mechanisms, their implementation, and changes in user behavior.

Direct Costs. Among the direct costs associated with the use of logical access controls are the purchase and support of hardware, operating systems, and applications that provide the controls, and any add-on security packages. The most significant personnel cost in relation to logical access control is usually for administration (e.g., initially determining, assigning, and keeping access rights up to date). Label-based access control is available in a limited number of commercial products, but at greater cost and with less variety of selection. Role-based systems are becoming more available, but there are significant costs involved in customizing these systems for a particular organization. Training users to understand and use an access control system is another necessary cost.

Indirect Costs. The primary indirect cost associated with introducing logical access controls into a computer system is the effect on user productivity. There may be additional overhead involved in having individual users properly determine (when under their control) the protection attributes of information. Another indirect cost that may arise results from users not being able to immediately access information necessary to accomplish their jobs because the permissions were incorrectly assigned (or have changed). This situation is familiar to most organizations that put strong emphasis on logical access controls.

References

Abrams, M.D., et al. *A Generalized Framework for Access Control: An Informal Description.* McLean, VA: Mitre Corporation, 1990.

Baldwin, R.W. "Naming and Grouping Privileges to Simplify Security Management in Large Databases." *1990 IEEE Symposium on Security and Privacy Proceedings.* Oakland, CA: IEEE Computer Society Press, May 1990. pp. 116-132.

Caelli, William, Dennis Longley, and Michael Shain. *Information Security Handbook.* New York, NY: Stockton Press, 1991.

Cheswick, William, and Steven Bellovin. *Firewalls and Internet Security.* Reading, MA: Addison-Wesley Publishing Company, 1994.

Curry, D. *Improving the Security of Your UNIX System, ITSTD-721-FR-90-21.* Menlo Park, CA: SRI International, 1990.

Dinkel, Charles. *Secure Data Network System Access Control Documents.* NISTIR 90-4259. Gaithersburg, MD: National Institute of Standards and Technology, 1990.

Fites, P., and M. Kratz. *Information Systems Security: A Practitioner's Reference.* New York, NY: Van Nostrand Reinhold, 1993. Especially Chapters 1, 9, and 12.

Garfinkel, S., and Spafford, G. "UNIX Security Checklist." *Practical UNIX Security.* Sebastopol, CA: O'Riley & Associates. Inc., 1991. pp. 401-413.

Gasser, Morrie. *Building a Secure Computer System.* New York, NY: Van Nostrand Reinhold, 1988.

Haykin, M., and R. Warner. *Smart Card Technology: New Methods for Computer Access Control.* Spec Pub 500-157. Gaithersburg, MD: National Institute of Standards and Technology, 1988.

Landwehr, C., C. Heitmeyer, and J. McLean. "A Security Model for Military Message Systems." *ACM Transactions on Computer Systems*, Vol. 2, No. 3, August 1984.

National Bureau of Standards. *Guidelines for Security of Computer Applications.* Federal

Information Processing Standard Publication 73. June 1980.

Pfleeger, Charles. *Security in Computing*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1989.

President's Council on Integrity and Efficiency. *Review of General Controls in Federal Computer Systems*. Washington, DC: President's Council on Integrity and Efficiency, October 1988.

S. Salamone, "Internetwork Security: Unsafe at Any Node?" *Data Communications*. 22(12), 1993. pp. 61-68.

Sandhu, R. "Transaction Control Expressions for Separation of Duty." *Fourth Annual Computer Security Applications Conference Proceedings*. Orlando, FL, December 1988, pp. 282-286.

Thomsen, D.J. "Role-based Application Design and Enforcement." *Fourth IFIP Workshop on Database Security Proceedings*. International Federation for Information Processing, Halifax, England, September 1990.

T. Whiting. "Understanding VAX/VMS Security." *Computers and Security*. 11(8), 1992. pp. 695-698.

Chapter 18

AUDIT TRAILS

Audit trails maintain a record of system activity both by system and application processes and by user activity of systems and applications.[127] In conjunction with appropriate tools and procedures, audit trails can assist in detecting security violations, performance problems, and flaws in applications.[128]

The Difference Between Audit Trails and Auditing

An *audit trail* is a series of records of computer events, about an operating system, an application, or user activities. A computer system may have several audit trails, each devoted to a particular type of activity.

Auditing is the review and analysis of management, operational, and technical controls. The auditor can obtain valuable information about activity on a computer system from the audit trail. Audit trails improve the *auditability* of the computer system. Auditing is discussed in the assurance chapter.

Audit trails may be used as either a support for regular system operations or a kind of insurance policy or as both of these. As insurance, audit trails are maintained but are not used unless needed, such as after a system outage. As a support for operations, audit trails are used to help system administrators ensure that the system or resources have not been harmed by hackers, insiders, or technical problems.

This chapter focuses on audit trails as a technical control, rather than the process of security auditing, which is a review and analysis of the security of a system as discussed in Chapter 9. This chapter discusses the benefits and objectives of audit trails, the types of audit trails, and some common implementation issues.

18.1 Benefits and Objectives

An *event* is any action that happens on a computer system. Examples include logging into a system, executing a program, and opening a file.

Audit trails can provide a means to help accomplish *several* security-related objectives, including individual accountability,

[127] Some security experts make a distinction between an *audit trail* and an *audit log* as follows: a *log* is a record of events made by a particular software package, and an *audit trail* is an entire history of an event, possibly using several logs. However, common usage within the security community does not make use of this definition. Therefore, this document does not distinguish between trails and logs.

[128] The type and amount of detail recorded by audit trails vary by both the technical capability of the logging application and the managerial decisions. Therefore, when we state that "audit trails can...," the reader should be aware that capabilities vary widely.

reconstruction of events, intrusion detection, and problem analysis.

18.1.1 Individual Accountability

Audit trails are a technical mechanism that help managers maintain individual accountability. By advising users that they are personally accountable for their actions, which are tracked by an audit trail that logs user activities, managers can help promote proper user behavior.[129] Users are less likely to attempt to circumvent security policy if they know that their actions will be recorded in an audit log.

For example, audit trails can be used in concert with access controls to identify and provide information about users suspected of improper modification of data (e.g., introducing errors into a database). An audit trail may record "before" and "after" versions of records. (Depending upon the size of the file and the capabilities of the audit logging tools, this may be very resource-intensive.) Comparisons can then be made between the actual changes made to records and what was expected. This can help management determine if errors were made by the user, by the system or application software, or by some other source.

Audit trails work in concert with logical access controls, which restrict use of system resources. Granting users access to particular resources usually means that they need that access to accomplish their job. Authorized access, of course, can be misused, which is where audit trail analysis is useful. While users cannot be prevented from using resources to which they have legitimate access authorization, audit trail analysis is used to examine their actions. For example, consider a personnel office in which users have access to those personnel records for which they are responsible. Audit trails can reveal that an individual is printing far more records than the average user, which could indicate the selling of personal data. Another example may be an engineer who is using a computer for the design of a new product. Audit trail analysis could reveal that an outgoing modem was used extensively by the engineer the week before quitting. This could be used to investigate whether proprietary data files were sent to an unauthorized party.

18.1.2 Reconstruction of Events

Audit trails can also be used to reconstruct events after a problem has occurred. Damage can be more easily assessed by reviewing audit trails of system activity to pinpoint how, when, and why normal operations ceased. Audit trail analysis can often distinguish between operator-induced errors (during which the system may have performed exactly as instructed) or system-created errors (e.g., arising from a poorly tested piece of replacement code). If, for example, a system fails or the integrity of a file (either program or data) is questioned, an analysis of the audit trail

[129] For a fuller discussion of changing employee behavior, see Chapter 13.

can reconstruct the series of steps taken by the system, the users, and the application. Knowledge of the conditions that existed at the time of, for example, a system crash, can be useful in avoiding future outages. Additionally, if a technical problem occurs (e.g., the corruption of a data file) audit trails can aid in the recovery process (e.g., by using the record of changes made to reconstruct the file).

18.1.3 Intrusion Detection

Intrusion detection refers to the process of identifying attempts to penetrate a system and gain unauthorized access.

If audit trails have been designed and implemented to record appropriate information, they can assist in intrusion detection. Although normally thought of as a real-time effort, intrusions can be detected *in real time*, by examining audit records as they are created (or through the use of other kinds of warning flags/notices), or *after the fact* (e.g., by examining audit records in a batch process).

Real-time intrusion detection is primarily aimed at outsiders attempting to gain unauthorized access to the system. It may also be used to detect changes in the system's performance indicative of, for example, a virus or worm attack.[130] There may be difficulties in implementing real-time auditing, including unacceptable system performance.

After-the-fact identification may indicate that unauthorized access was attempted (or was successful). Attention can then be given to damage assessment or reviewing controls that were attacked.

18.1.4 Problem Analysis

Audit trails may also be used as on-line tools to help identify problems other than intrusions as they occur. This is often referred to as *real-time auditing* or monitoring. If a system or application is deemed to be critical to an organization's business or mission, real-time auditing may be implemented to monitor the status of these processes (although, as noted above, there can be difficulties with real-time analysis). An analysis of the audit trails may be able to verify that the *system* operated normally (i.e., that an error may have resulted from operator error, as opposed to a system-originated error). Such use of audit trails may be complemented by system performance logs. For example, a significant increase in the use of system resources (e.g., disk file space or outgoing modem use) *could* indicate a security problem.

[130] Viruses and worms of forms of malicious code. A virus is a code segment that replicates by attaching copies of itself to existing executables. A worm is a self-replicating program.

18.2 Audit Trails and Logs

A system can maintain several different audit trails concurrently. There are typically two kinds of audit records, (1) an event-oriented log and (2) a record of every keystroke, often called keystroke monitoring. Event-based logs usually contain records describing *system* events, *application* events, or *user* events.

An audit trail should include sufficient information to establish what events occurred and who (or what) caused them. In general, an event record should specify when the event occurred, the user ID associated with the event, the program or command used to initiate the event, and the result. Date and time can help determine if the user was a masquerader or the actual person specified.

18.2.1 Keystroke Monitoring[131]

Keystroke monitoring is the process used to view or record both the keystrokes entered by a computer user and the computer's response during an interactive session. Keystroke monitoring is usually considered a special case of audit trails. Examples of keystroke monitoring would include viewing characters as they are typed by users, reading users' electronic mail, and viewing other recorded information typed by users.

Some forms of routine system maintenance may record user keystrokes. This could constitute keystroke monitoring if the keystrokes are preserved along with the user identification so that an administrator could determine the keystrokes entered by specific users. Keystroke monitoring is conducted in an effort to protect systems and data from intruders who access the systems without authority or in excess of their assigned authority. Monitoring keystrokes typed by intruders can help administrators assess and repair damage caused by intruders.

18.2.2 Audit Events

System audit records are generally used to monitor and fine-tune system performance. *Application audit trails* may be used to discern flaws in applications, or violations of security policy committed within an application. *User audits records* are generally used to hold individuals accountable for their actions. An analysis of user audit records may expose a variety

[131] The Department of Justice has advised that an ambiguity in U.S. law makes it unclear whether keystroke monitoring is considered equivalent to an unauthorized telephone wiretap. The ambiguity results from the fact that current laws were written years before such concerns as keystroke monitoring or system intruders became prevalent. Additionally, no legal precedent has been set to determine whether keystroke monitoring is legal or illegal. System administrators conducting such monitoring might be subject to criminal and civil liabilities. The Department of Justice advises system administrators to protect themselves by giving notice to system users if keystroke monitoring is being conducted. Notice should include agency/organization policy statements, training on the subject, and a banner notice on each system being monitored. [NIST, *CSL Bulletin*, March 1993]

of security violations, which might range from simple browsing to attempts to plant Trojan horses or gain unauthorized privileges.

Sample System Log File Showing Authentication Messages

```
Jan 27 17:14:04 host1 login: ROOT LOGIN console
Jan 27 17:15:04 host1 shutdown: reboot by root
Jan 27 17:18:38 host1 login: ROOT LOGIN console
Jan 27 17:19:37 host1 reboot: rebooted by root
Jan 28 09:46:53 host1 su: 'su root' succeeded for user1 on /dev/ttyp0
Jan 28 09:47:35 host1 shutdown: reboot by user1
Jan 28 09:53:24 host1 su: 'su root' succeeded for user1 on /dev/ttyp1
Feb 12 08:53:22 host1 su: 'su root' succeeded for user1 on /dev/ttyp1
Feb 17 08:57:50 host1 date: set by user1
Feb 17 13:22:52 host1 su: 'su root' succeeded for user1 on /dev/ttyp0
```

The system itself enforces certain aspects of policy (particularly *system-specific* policy) such as access to files and access to the system itself. Monitoring the alteration of systems configuration files that implement the policy is important. If special accesses (e.g., security administrator access) have to be used to alter configuration files, the system should generate audit records whenever these accesses are used.

Application-Level Audit Record for a Mail Delivery System

```
Apr  9 11:20:22 host1 AA06370: from=<user2@host2>, size=3355, class=0
Apr  9 11:20:23 host1 AA06370: to=<user1@host1>, delay=00:00:02,
stat=Sent
Apr  9 11:59:51 host1 AA06436: from=<user4@host3>, size=1424, class=0
Apr  9 11:59:52 host1 AA06436: to=<user1@host1>, delay=00:00:02,
stat=Sent
Apr  9 12:43:52 host1 AA06441: from=<user2@host2>, size=2077, class=0
Apr  9 12:43:53 host1 AA06441: to=<user1@host1>, delay=00:00:01,
stat=Sent
```

Sometimes a finer level of detail than system audit trails is required. *Application audit trails* can provide this greater level of recorded detail. If an application is critical, it can be desirable to record not only who invoked the application, but certain details specific to each use. For example, consider an e-mail application. It may be desirable to record who sent mail, as well as to whom they sent mail and the length of messages. Another example would be that of a database application. It may be useful to record who accessed what database as well as the individual rows or columns of a table that were read (or changed or deleted), instead of just recording the execution of the database program.

User Log Showing a Chronological List of Commands Executed by Users

```
rcp         user1     ttyp0     0.02 secs Fri Apr  8 16:02
ls          user1     ttyp0     0.14 secs Fri Apr  8 16:01
clear       user1     ttyp0     0.05 secs Fri Apr  8 16:01
rpcinfo     user1     ttyp0     0.20 secs Fri Apr  8 16:01
nroff       user2     ttyp2     0.75 secs Fri Apr  8 16:00
sh          user2     ttyp2     0.02 secs Fri Apr  8 16:00
mv          user2     ttyp2     0.02 secs Fri Apr  8 16:00
sh          user2     ttyp2     0.03 secs Fri Apr  8 16:00
col         user2     ttyp2     0.09 secs Fri Apr  8 16:00
man         user2     ttyp2     0.14 secs Fri Apr  8 15:57
```

A *user audit trail* monitors and logs user activity in a system or application by recording events initiated by the user (e.g., access of a file, record or field, use of a modem).

Flexibility is a critical feature of audit trails. Ideally (from a security point of view), a system administrator would have the ability to monitor all system and user activity, but could choose to log only certain functions at the system level, and within certain applications. The decision of how much to log and how much to review should be a function of application/data sensitivity and should be decided by each functional manager/application owner with guidance from the system administrator and the computer security manager/officer, weighing the costs and benefits of the logging.[132]

18.2.2.1 System-Level Audit Trails

A system audit trail should be able to identify failed log-on attempts, especially if the system does not limit the number of failed log-on attempts. Unfortunately, some system-level audit trails cannot detect attempted log-ons, and therefore, cannot log them for later review. These audit trails can only monitor and log successful log-ons and subsequent activity. To effectively detect intrusion, a record of failed log-on attempts is required.

If a system-level audit capability exists, the audit trail should capture, at a minimum, any attempt to log on (successful or unsuccessful), the log-on ID, date and time of each log-on attempt, date and time of each log-off, the devices used, and the function(s) performed once logged on (e.g., the applications that the user tried, successfully or unsuccessfully, to

[132] In general, audit logging can have privacy implications. Users should be aware of applicable privacy laws, regulations, and policies that may apply in such situations.

invoke). System-level logging also typically includes information that is not specifically security-related, such as system operations, cost-accounting charges, and network performance.

18.2.2.2 Application-Level Audit Trails

System-level audit trails may not be able to track and log events *within* applications, or may not be able to provide the level of detail needed by application or data owners, the system administrator, or the computer security manager. In general, application-level audit trails monitor and log user activities, including data files opened and closed, specific actions, such as reading, editing, and deleting records or fields, and printing reports. Some applications may be sensitive enough from a data availability, confidentiality, and/or integrity perspective that a "before" and "after" picture of each modified record (or the data element(s) changed within a record) should be captured by the audit trail.

Audit Logs for Physical Access

Physical access control systems (e.g., a card/key entry system or an alarm system) use software and audit trails similar to general-purpose computers. The following are *examples* of criteria that may be used in selecting which events to log:

The date and time the access was attempted or made should be logged, as should the gate or door through which the access was attempted or made, and the individual (or user ID) making the attempt to access the gate or door.

Invalid attempts should be monitored and logged by noncomputer audit trails just as they are for computer-system audit trails. Management should be made aware if someone attempts to gain access during unauthorized hours.

Logged information should also include attempts to add, modify, or delete physical access privileges (e.g., granting a new employee access to the building or granting transferred employees access to their new office [and, of course, deleting their old access, as applicable]).

As with system and application audit trails, auditing of noncomputer functions can be implemented to send messages to security personnel indicating valid or invalid attempts to gain access to controlled spaces. In order not to desensitize a guard or monitor, all access should not result in messages being sent to a screen. Only exceptions, such as failed access attempts, should be highlighted to those monitoring access.

18.2.2.3 User Audit Trails

User audit trails can usually log:

- all commands directly initiated by the user;
- all identification and authentication attempts; and
- files and resources accessed.

It is most useful if options and parameters are also recorded from commands. It is much more useful to know that a user tried to delete a log file (e.g., to hide unauthorized actions) than to know the user merely issued the delete command, possibly for a personal data file.

18.3 Implementation Issues

Audit trail data requires protection, since the data should be available for use when needed and is not useful if it is not accurate. *Also, the* *best planned and implemented audit trail is of limited value without timely review of the logged data.* Audit trails may be reviewed periodically, as needed (often triggered by occurrence of a security event), automatically in realtime, or in some combination of these. System managers and administrators, with guidance from computer security personnel, should determine how long audit trail data will be maintained – either on the system or in archive files.

Following are examples of implementation issues that may have to be addressed when using audit trails.

18.3.1 Protecting Audit Trail Data

Access to on-line audit logs should be strictly controlled. Computer security managers and system administrators or managers should have access for review purposes; however, security and/or administration personnel who maintain logical access functions may have no need for access to audit logs.

It is particularly important to ensure the *integrity* of audit trail data against modification. One way to do this is to use digital signatures. (See Chapter 19.) Another way is to use write-once devices. The audit trail files needs to be protected since, for example, intruders may try to "cover their tracks" by modifying audit trail records. Audit trail records should be protected by strong access controls to help prevent unauthorized access. The integrity of audit trail information may be particularly important when legal issues arise, such as when audit trails are used as legal evidence. (This may, for example, require daily printing and signing of the logs.) Questions of such legal issues should be directed to the cognizant legal counsel.

The confidentiality of audit trail information may also be protected, for example, if the audit trail is recording information about users that may be disclosure-sensitive such as transaction data containing personal information (e.g., "before" and "after" records of modification to income tax data). Strong access controls and encryption can be particularly effective in preserving confidentiality.

18.3.2 Review of Audit Trails

Audit trails can be used to review what occurred after an event, for periodic reviews, and for real-time analysis. Reviewers should know what to look for to be effective in spotting unusual activity. They need to understand what normal activity looks like. Audit trail review can be easier if the audit trail function can be queried by user ID, terminal ID, application name, date and time, or some other set of parameters to run reports of selected information.

Audit Trail Review After an Event. Following a known system or application software problem, a known violation of existing requirements by a user, or some unexplained system or user problem, the appropriate system-level or application-level administrator should review the audit trails.

Review by the application/data owner would normally involve a separate report, based upon audit trail data, to determine if their resources are being misused.

Periodic Review of Audit Trail Data. Application owners, data owners, system administrators, data processing function managers, and computer security managers should determine how much review of audit trail records is necessary, based on the importance of identifying unauthorized activities. This determination should have a direct correlation to the frequency of periodic reviews of audit trail data.

Real-Time Audit Analysis. Traditionally, audit trails are analyzed in a batch mode at regular intervals (e.g., daily). Audit records are archived during that interval for later analysis. Audit analysis tools can also be used in a real-time, or near real-time fashion. Such intrusion detection tools are based on audit reduction, attack signature, and variance techniques. Manual review of audit records in real time is almost never feasible on large multiuser systems due to the volume of records generated. However, it might be possible to view all records associated with a particular user or application, and view them in real time.[133]

18.3.3 Tools for Audit Trail Analysis

Many types of tools have been developed to help to reduce the amount of information contained in audit records, as well as to distill useful information from the raw data. Especially on larger systems, audit trail software can create very large files, which can be extremely difficult to analyze manually. The use of automated tools is likely to be the difference between unused audit trail data and a robust program. Some of the types of tools include:

Audit reduction tools are preprocessors designed to reduce the volume of audit records to facilitate manual review. Before a security review, these tools can remove many audit records known to have little security significance. (This alone may cut in half the number of records in the audit trail.) These tools generally remove records generated by specified classes of events, such as records generated by nightly backups might be removed.

Trends/variance-detection tools look for anomalies in user or system behavior. It is possible to construct more sophisticated processors that monitor usage trends and detect major variations. For example, if a user typically logs in at 9 a.m., but appears at 4:30 a.m. one morning, this may indicate a security problem that may need to be investigated.

Attack signature-detection tools look for an *attack signature*, which is a specific sequence of events indicative of an unauthorized access attempt. A simple example would be repeated failed log-in attempts.

[133] This is similar to keystroke monitoring, though, and may be legally restricted.

18.4 Interdependencies

The ability to audit supports many of the controls presented in this handbook. The following paragraphs describe some of the most important interdependencies.

Policy. The most fundamental interdependency of audit trails is with policy. Policy dictates who is authorized access to what system resources. Therefore it specifies, directly or indirectly, what violations of policy should be identified through audit trails.

Assurance. System auditing is an important aspect of operational assurance. The data recorded into an audit trail is used to support a system audit. The analysis of audit trail data and the process of auditing systems are closely linked; in some cases, they may even be the same thing. In most cases, the analysis of audit trail data is a critical part of maintaining operational assurance.

Identification and Authentication. Audit trails are tools often used to help hold users accountable for their actions. To be held accountable, the users must be known to the system (usually accomplished through the identification and authentication process). However, as mentioned earlier, audit trails record events and associate them with the *perceived* user (i.e., the user ID). If a user is impersonated, the audit trail will establish events but *not* the identity of the user.

Logical Access Control. Logical access controls restrict the use of system resources to authorized users. Audit trails complement this activity in two ways. First, they may be used to identify breakdowns in logical access controls or to verify that access control restrictions are behaving as expected, for example, if a particular user is erroneously included in a group permitted access to a file. Second, audit trails are used *to audit use of resources by those who have legitimate access.* Additionally, to protect audit trail files, access controls are used to ensure that audit trails are not modified.

Contingency Planning. Audit trails assist in contingency planning by leaving a record of activities performed on the system or within a specific application. In the event of a technical malfunction, this log can be used to help reconstruct the state of the system (or specific files).

Incident Response. If a security incident occurs, such as hacking, audit records and other intrusion detection methods can be used to help determine the extent of the incident. For example, was just one file browsed, or was a Trojan horse planted to collect passwords?

Cryptography. Digital signatures can be used to protect audit trails from undetected modification. (This does not prevent deletion or modification of the audit trail, but will provide an alert that the audit trail has been altered.) Digital signatures can also be used in conjunction with adding secure time stamps to audit records. Encryption can be used if confidentiality of audit trail information is important.

18.5 Cost Considerations

Audit trails involve many costs. First, some system overhead is incurred recording the audit trail. Additional system overhead will be incurred storing and processing the records. The more detailed the records, the more overhead is required. Another cost involves human and machine time required to do the analysis. This can be minimized by using tools to perform most of the analysis. Many simple analyzers can be constructed quickly (and cheaply) from system utilities, but they are limited to audit reduction and identifying particularly sensitive events. More complex tools that identify trends or sequences of events are slowly becoming available as off-the-shelf software. (If complex tools are not available for a system, development may be prohibitively expensive. Some intrusion detection systems, for example, have taken years to develop.)

The final cost of audit trails is the cost of investigating anomalous events. If the system is identifying too many events as suspicious, administrators may spend undue time reconstructing events and questioning personnel.

References

Fites, P., and M. Kratz. *Information Systems Security: A Practitioner's Reference*. New York: Van Nostrand Reinhold, 1993, (especially Chapter 12, pp. 331 - 350).

Kim, G., and E. Spafford, "Monitoring File System Integrity on UNIX Platforms." *Infosecurity News*. 4(4), 1993. pp. 21-22.

Lunt, T. "Automated Audit Trail Analysis for Intrusion Detection," *Computer Audit Update*, April 1992. pp. 2-8.

National Computer Security Center. *A Guide to Understanding Audit in Trusted Systems*. NCSC-TG-001, Version-2. Ft. Meade, MD, 1988.

National Institute of Standards and Technology. "Guidance on the Legality of Keystroke Monitoring." *CSL Bulletin*. March 1993.

Phillips, P. W. "New Approach Identifies Malicious System Activity." *Signal*. 46(7), 1992. pp. 65-66.

Ruthberg, Z., et al. *Guide to Auditing for Controls and Security: A System Development Life Cycle Approach*. Special Publication 500-153. Gaithersburg, MD: National Bureau of Standards, 1988.

Stoll, Clifford. *The Cuckoo's Egg*. New York, NY: Doubleday, 1989.

Chapter 19

CRYPTOGRAPHY

Cryptography is a branch of mathematics based on the transformation of data. It provides an important tool for protecting information and is used in many aspects of computer security. For example, cryptography can help provide data confidentiality, integrity, electronic signatures, and advanced user authentication. Although modern cryptography relies upon advanced mathematics, users can reap its benefits without understanding its mathematical underpinnings.

This chapter describes cryptography as a tool for satisfying a wide spectrum of computer security needs and requirements. It describes fundamental aspects of the basic cryptographic technologies and some specific ways cryptography can be applied to improve security. The chapter also explores some of the important issues that should be considered when incorporating cryptography into computer systems.

Cryptography is traditionally associated only with keeping data secret. However, modern cryptography can be used to provide many security services, such as electronic signatures and ensuring that data has not been modified.

19.1 Basic Cryptographic Technologies

Cryptography relies upon two basic components: an *algorithm* (or cryptographic methodology) and a *key*. In modern cryptographic systems, algorithms are complex mathematical formulae and keys are strings of bits. For two parties to communicate, they must use the same algorithm (or algorithms that are designed to work together). In some cases, they must also use the same key. Many cryptographic keys must be kept secret; sometimes algorithms are also kept secret.

There are two basic types of cryptography: "secret key" and "public key."

There are two basic types of cryptography: *secret key systems* (also called symmetric systems) and *public key systems* (also called asymmetric systems). Table 19.1 compares some of the distinct features of secret and public key systems. Both types of systems offer advantages and disadvantages. Often, the two are combined to form a *hybrid system* to exploit the strengths of each type. To determine which type of cryptography best meets its needs, an organization first has to identify its security requirements and operating environment.

DISTINCT FEATURES	SECRET KEY CRYPTOGRAPHY	PUBLIC KEY CRYPTOGRAPHY
NUMBER OF KEYS	Single key.	Pair of keys.
TYPES OF KEYS	Key is secret.	One key is private, and one key is public.
PROTECTION OF KEYS	Disclosure and modification.	Disclosure and modification for private keys and modification for public keys.
RELATIVE SPEEDS	Faster.	Slower.

Table 19.1

19.1.1 Secret Key Cryptography

In secret key cryptography, two (or more) parties share the same key, and that key is used to encrypt and decrypt data. As the name implies, secret key cryptography relies on keeping the key secret. If the key is compromised, the security offered by cryptography is severely reduced or eliminated. Secret key cryptography assumes that the parties who share a key rely upon each other not to disclose the key and protect it against modification.

The best known secret key system is the *Data Encryption Standard* (DES), published by NIST as Federal Information Processing Standard (FIPS) 46-2. Although the adequacy of DES has at times been questioned, these claims remain unsubstantiated, and DES remains strong. It is the most widely accepted, publicly available cryptographic system today. The American National Standards Institute (ANSI) has adopted DES as the basis for encryption, integrity, access control, and key management standards.

> Secret key cryptography has been in use for centuries. Early forms merely transposed the written characters to hide the message.

The *Escrowed Encryption Standard*, published as FIPS 185, also makes use of a secret key system. (See the discussion of Key Escrow Encryption in this chapter.)

19.1.2 Public Key Cryptography

Whereas secret key cryptography uses a single key shared by two (or more) parties, public key cryptography uses a pair of keys for *each* party. One of the keys of the pair is "public" and the other is "private." The public key can be made known to other parties; the private key must be kept confidential and must be known only to its owner. Both keys, however, need to be protected against modification.

Public key cryptography is a modern invention and requires the use of advanced mathematics.

Public key cryptography is particularly useful when the parties wishing to communicate cannot rely upon each other or do not share a common key. There are several public key cryptographic systems. One of the first public key systems is RSA, which can provide many different security services. The Digital Signature Standard (DSS), described later in the chapter, is another example of a public key system.

19.1.3 Hybrid Cryptographic Systems

Public and secret key cryptography have relative advantages and disadvantages. Although public key cryptography does not require users to share a common key, secret key cryptography is much faster: equivalent implementations of secret key cryptography can run 1,000 to 10,000 times faster than public key cryptography.

Secret key systems are often used for bulk data encryption and public key systems for automated key distribution.

To maximize the advantages and minimize the disadvantages of both secret and public key cryptography, a computer system can use both types in a complementary manner, with each performing different functions. Typically, the speed advantage of secret key cryptography means that it is used for encrypting data. Public key cryptography is used for applications that are less demanding to a computer system's resources, such as encrypting the keys used by secret key cryptography (for distribution) or to sign messages.

19.1.4 Key Escrow

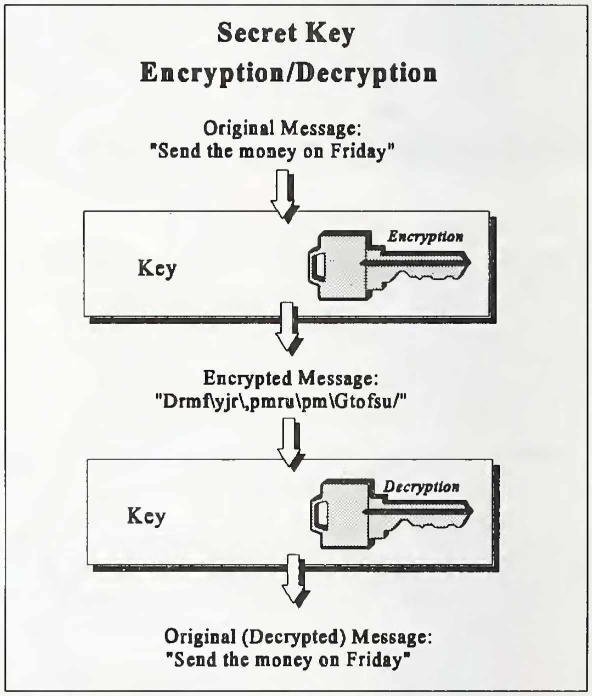
Because cryptography can provide extremely strong encryption, it can thwart the government's efforts to lawfully perform electronic surveillance. For example, if strong cryptography is used to encrypt a phone conversation, a court-authorized wiretap will not be effective. To meet the needs of the government *and* to provide privacy, the federal government has adopted voluntary key escrow cryptography. This technology allows the use of strong encryption, but also allows the government when legally authorized to obtain decryption keys held by escrow agents. NIST has published the *Escrowed Encryption Standard* as FIPS 185. Under the Federal Government's voluntary key escrow initiative, the decryption keys are split into parts and given to separate escrow authorities. Access to one part of the key does *not* help decrypt the data; both keys must be obtained.

19.2 Uses of Cryptography

Cryptography is used to protect data *both* inside and outside the boundaries of a computer system. Outside the computer system, cryptography is sometimes the *only* way to protect data. While in a computer system, data is normally protected with logical and physical access controls (perhaps supplemented by cryptography). However, when in transit across communications lines or resident on someone else's computer, data cannot be protected by the originator's[134] logical or physical access controls. Cryptography provides a solution by protecting data even when the data is no longer in the control of the originator.

19.2.1 Data Encryption

One of the best ways to obtain cost-effective data confidentiality is through the use of encryption. Encryption transforms intelligible data, called *plaintext,*[135] into an unintelligible form, called *ciphertext*. This process is reversed through the process of decryption. Once data is encrypted, the ciphertext does not have to be protected against disclosure. However, if ciphertext is modified, it will not decrypt correctly.



Secret Key Encryption/Decryption

Original Message:
"Send the money on Friday"
Key
Encryption
Encrypted Message:
"Drmf\yjr\,pmru\pm\Gtofsu/"
Key
Decryption
Original (Decrypted) Message:
"Send the money on Friday"

Both secret key and public key cryptography can be used for data encryption although not all public key algorithms provide for data encryption.

To use a secret key algorithm, data is encrypted using a key. The same key must be used to

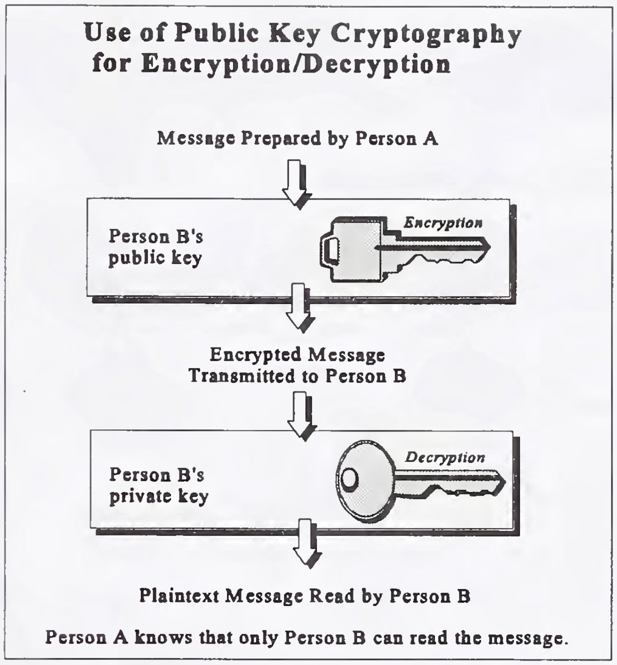
[134] The originator does not have to be the original creator of the data. It can also be a guardian or custodian of the data.

[135] Plaintext can be intelligible to a human (e.g., a novel) or to a machine (e.g., executable code).

decrypt the data.

When public key cryptography is used for encryption, any party may use any other party's public key to encrypt a message; however, only the party with the corresponding private key can decrypt, and thus read, the message.

Since secret key encryption is typically much faster, it is normally used for encrypting larger amounts of data.


Use of Public Key Cryptography for Encryption/Decryption
Message Prepared by Person A
Person B's public key
Encryption
Encrypted Message Transmitted to Person B
Person B's private key
Decryption
Plaintext Message Read by Person B
Person A knows that only Person B can read the message.

19.2.2 Integrity

In computer systems, it is not always possible for humans to scan information to determine if data has been erased, added, or modified. Even if scanning were possible, the individual may have no way of knowing what the correct data should be. For example, "do" may be changed to "do not," or $1,000 may be changed to $10,000. It is therefore desirable to have an automated means of detecting *both* intentional and unintentional modifications of data.

While error detecting codes have long been used in communications protocols (e.g., parity bits), these are more effective in detecting (and correcting) unintentional modifications. They can be defeated by adversaries. Cryptography can effectively detect both intentional and unintentional modification; however, cryptography does not protect files from being modified. Both secret key and public key cryptography can be used to ensure integrity. Although newer public key methods may offer more flexibility than the older secret key method, secret key integrity verification systems have been successfully integrated into many applications.

When secret key cryptography is used, a message authentication code (MAC) is calculated from and appended to the data. To verify that the data has not been modified at a later time, any party with access to the correct secret key can recalculate the MAC. The new MAC is compared with the original MAC, and if they are identical, the verifier has confidence that the data has not been modified by an unauthorized party. FIPS 113, *Computer Data Authentication*, specifies a standard technique for calculating a MAC for integrity verification.

Public key cryptography verifies integrity by using of public key signatures and secure hashes. A secure hash algorithm is used to create a message digest. The message digest, called a hash, is a short form of the message that changes if the message is modified. The hash is then signed with a private key. Anyone can recalculate the hash and use the corresponding public key to verify the integrity of the message.[136]

19.2.3 Electronic Signatures

Today's computer systems store and process increasing numbers of paper-based documents in electronic form. Having documents in electronic form permits rapid processing and transmission and improves overall efficiency. However, approval of a paper document has traditionally been indicated by a written signature. What is needed, therefore, is the electronic equivalent of a written signature that can be recognized as having the same legal status as a written signature. In addition to the integrity protections, discussed above, cryptography can provide a means of linking a document with a particular person, as is done with a written signature. Electronic signatures can use either secret key or public key cryptography; however, public key methods are generally easier to use.

What Is an Electronic Signature?

An electronic signature is a cryptographic mechanism that performs a similar function to a written signature. It is used to verify the origin and contents of a message. For example, a recipient of data (e.g., an e-mail message) can verify who signed the data and that the data was not modified after being signed. This also means that the originator (e.g., sender of an e-mail message) cannot falsely deny having signed the data.

Cryptographic signatures provide extremely strong proof that a message has not been altered and was signed by a specific key.[137] However, there are other mechanisms besides cryptographic-based electronic signatures that perform a *similar* function. These mechanisms provide some assurance of the origin of a message, some verification of the message's integrity, or both.[138]

[136] Sometimes a secure hash is used for integrity verification. However, this can be defeated if the hash is not stored in a secure location, since it may be possible for someone to change the message and then replace the old hash with a new one based on the modified message.

[137] Electronic signatures rely on the secrecy of the keys and the link or binding between the owner of the key and the key itself. If a key is compromised (by theft, coercion, or trickery), then the electronic originator of a message may not be the same as the owner of the key. Although the binding of cryptographic keys to actual people is a significant problem, it does not necessarily make electronic signatures less secure than written signatures. Trickery and coercion are problems for written signatures as well. In addition, written signatures are easily forged.

[138] The strength of these mechanisms relative to electronic signatures varies depending on the specific implementation; however, in general, electronic signatures are stronger and more flexible. These mechanisms may be used in conjunction with electronic signatures or separately, depending upon the system's specific needs and limitations.

- Examination of the transmission path of a message. When messages are sent across a network, such as the Internet, the message source and the physical path of the message are recorded as a part of the message. These can be examined electronically or manually to help ascertain the origin of a message.

- Use of a value-added network provider. If two or more parties are communicating via a third party network, the network provider may be able to provide assurance that messages originate from a given source and have not been modified.

- Acknowledgment statements. The recipient of an electronic message may confirm the message's origin and contents by sending back an acknowledgement statement.

- Use of audit trails. Audit trails can track the sending of messages and their contents for later reference.

Simply taking a digital picture of a written signature does not provide adequate security. Such a *digitized* written signature could easily be copied from one electronic document to another with no way to determine whether it is legitimate. Electronic signatures, on the other hand, are unique to the message being signed and will not verify if they are copied to another document.

19.2.3.1 Secret Key Electronic Signatures

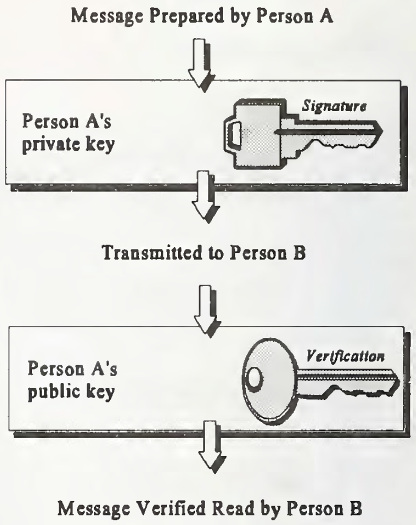
An electronic signature can be implemented using secret key message authentication codes (MACs). For example, if two parties share a secret key, and one party receives data with a MAC that is correctly verified using the shared key, that party may assume that the other party signed the data. This assumes, however, that the two parties trust each other. Thus, through the use of a MAC, in addition to data integrity, a form of electronic signature is obtained. Using additional controls, such as key notarization and key attributes, it is possible to provide an electronic signature even if the two parties do not trust each other.

> Systems incorporating message authentication technology have been approved for use by the federal government as a replacement for written signatures on electronic documents.

19.2.3.2 Public Key Electronic Signatures

Another type of electronic signature called a *digital signature* is implemented using public key cryptography. Data is electronically signed by applying the originator's private key to the data. (The exact mathematical process for doing this is not important for this discussion.) To increase the speed of the process, the private key is applied to a shorter form of the data, called a "hash" or "message digest," rather than to the entire set of data. The resulting digital signature can be stored or transmitted along with the data. The signature can be verified by any party using the public key of the signer. This feature is very useful, for example, when distributing signed copies of virus-free software. Any recipient can verify that the program remains virus-free. If the signature verifies properly, then the verifier has confidence that the data was not modified after being signed and that the owner of the public key was the signer.

Use of Public Key Cryptography for Digital Signature



Person B knows that only Person A could have sent the message.

NIST has published standards for a digital signature and a secure hash for use by the federal government in FIPS 186, *Digital Signature Standard* and FIPS 180, *Secure Hash Standard.*

19.2.4 User Authentication

Cryptography can increase security in user authentication techniques. As discussed in Chapter 16, cryptography is the basis for several advanced authentication methods. Instead of communicating passwords over an open network, authentication can be performed by demonstrating knowledge of a cryptographic key. Using these methods, a one-time password, which is not susceptible to eavesdropping, can be used. User authentication can use either secret or public key cryptography.

19.3 Implementation Issues

This section explores several important issues that should be considered when using (e.g., designing, implementing, integrating) cryptography in a computer system.

19.3.1 Selecting Design and Implementation Standards

Applicable security standards provide a common level of security and interoperability among users.

NIST and other organizations have developed numerous standards for designing, implementing, and using cryptography and for integrating it into automated systems. By using these standards, organizations can reduce costs and protect their investments in technology. Standards provide solutions that have been accepted by a wide community and that have been reviewed by experts in relevant areas. Standards help ensure interoperability among different vendors' equipment, thus allowing an organization to select from among various products in order to find cost-effective equipment.

Managers and users of computer systems will have to select among various standards when deciding to use cryptography. Their selection should be based on cost-effectiveness analysis, trends in the standard's acceptance, and interoperability requirements. In addition, each standard should be carefully analyzed to determine if it is applicable to the organization and the desired application. For example, the Data Encryption Standard and the Escrowed Encryption Standard are both applicable to certain applications involving communications of data over commercial modems. Some federal standards are mandatory for federal computer systems, including DES (FIPS 46-2) and the DSS (FIPS 181).

19.3.2 Deciding on Hardware vs. Software Implementations

The trade-offs among security, cost, simplicity, efficiency, and ease of implementation need to be studied by managers acquiring various security products meeting a standard. Cryptography can be implemented in either hardware or software. Each has its related costs and benefits.

In general, software is less expensive and slower than hardware, although for large applications, hardware may be less expensive. In addition, software may be less secure, since it is more easily modified or bypassed than equivalent hardware products. Tamper resistance is usually considered better in hardware.

In many cases, cryptography is implemented in a hardware device (e.g., electronic chip, ROM-protected processor) but is controlled by software. This software requires integrity protection to ensure that the hardware device is provided with correct information (i.e., controls, data) and is not bypassed. Thus, a hybrid solution is generally provided, even when the basic cryptography is implemented in hardware. Effective security requires the correct management of the entire hybrid solution.

19.3.3 Managing Keys

The proper management of cryptographic keys is essential to the effective use of cryptography for security. Ultimately, the security of information protected by cryptography directly depends upon the protection afforded to keys.

All keys need to be protected against modification, and secret keys and private keys need protection against unauthorized disclosure. Key management involves the procedures and protocols, both manual and automated, used throughout the entire life cycle of the keys. This includes the generation, distribution, storage, entry, use, destruction, and archiving of cryptographic keys.

With secret key cryptography, the secret key(s) should be securely distributed (i.e., safeguarded against unauthorized replacement, modification, and disclosure) to the parties wishing to communicate. Depending upon the number and location of users, this task may not be trivial. Automated techniques for generating and distributing cryptographic keys can ease overhead costs of key management, but some resources have to be devoted to this task. FIPS 171, *Key Management Using ANSI X9.17*, provides key management solutions for a variety of operational environments.

Public key cryptography users also have to satisfy certain key management requirements. For example, since a private-public key pair is associated with (i.e., generated or held by) a specific user, it is necessary to *bind* the public part of the key pair to the user.[139]

In a small community of users, public keys and their "owners" can be strongly bound by simply exchanging public keys (e.g., putting them on a CD-ROM or other media). However, conducting electronic business on a larger scale, potentially involving geographically and organizationally distributed users, necessitates a means for obtaining public keys electronically with a high degree of confidence in their integrity and binding to individuals. The support for the binding between a key and its owner is generally referred to as a *public key infrastructure.*

Users also need to be able enter the community of key holders, generate keys (or have them generated on their behalf), disseminate public keys, revoke keys (in case, for example, of compromise of the private key), and change keys. In addition, it may be necessary to build in time/date stamping and to archive keys for verification of old signatures.

19.3.4 Security of Cryptographic Modules

Cryptography is typically implemented in a *module* of software, firmware, hardware, or some combination thereof. This module contains the cryptographic algorithm(s), certain control parameters, and temporary storage facilities for the key(s) being used by the algorithm(s). The proper functioning of the cryptography requires the secure design, implementation, and use of the cryptographic module. This includes protecting the module against tampering.

FIPS 140-1, *Security Requirements for Cryptographic Modules*, specifies the physical and logical security requirements for cryptographic modules. The standard defines four security levels for cryptographic modules, with each level providing a significant increase in security over the preceding level. The four levels allow for cost-effective solutions that are appropriate for different degrees of data sensitivity and different application environments. The user can select the best module for any given application or system, avoiding the cost of unnecessary security features.

[139] In some cases, the key may be bound to a position or an organization, rather than to an individual user.

19.3.5 Applying Cryptography to Networks

The use of cryptography within networking applications often requires special considerations. In these applications, the suitability of a cryptographic module may depend on its capability for handling special requirements imposed by locally attached communications equipment or by the network protocols and software.

Encrypted information, MACs, or digital signatures may require transparent communications protocols or equipment to avoid being misinterpreted by the communications equipment or software as control information. It may be necessary to format the encrypted information, MAC, or digital signature to ensure that it does not confuse the communications equipment or software. It is essential that cryptography satisfy the requirements imposed by the communications equipment and does not interfere with the proper and efficient operation of the network.

Data is encrypted on a network using either link or end-to-end encryption. In general, *link encryption* is performed by service providers, such as a data communications provider. Link encryption encrypts all of the data along a communications path (e.g., a satellite link, telephone circuit, or T1 line). Since link encryption also encrypts routing data, communications nodes need to decrypt the data to continue routing. *End-to-end encryption* is generally performed by the end-user organization. Although data remains encrypted when being passed through a network, routing information remains visible. It is possible to combine both types of encryption.

19.3.6 Complying with Export Rules

The U.S. Government controls the export of cryptographic implementations. The rules governing export can be quite complex, since they consider multiple factors. In addition, cryptography is a rapidly changing field, and rules may change from time to time. Questions concerning the export of a particular implementation should be addressed to appropriate legal counsel.

19.4 Interdependencies

There are many interdependencies among cryptography and other security controls highlighted in this handbook. Cryptography both depends on other security safeguards and assists in providing them.

Physical Security. Physical protection of a cryptographic module is required to prevent – or at least detect – physical replacement or modification of the cryptographic system and the keys within it. In many environments (e.g., open offices, portable computers), the cryptographic module itself has to provide the desired levels of physical security. In other environments (e.g., closed communications facilities, steel-encased Cash-Issuing Terminals), a cryptographic module may be safely employed within a secured facility.

User Authentication. Cryptography can be used both to protect passwords that are stored in computer systems and to protect passwords that are communicated between computers. Furthermore, cryptographic-based authentication techniques may be used in conjunction with, or in place of, password-based techniques to provide stronger authentication of users.

Logical Access Control. In many cases, cryptographic software may be embedded within a host system, and it may not be feasible to provide extensive physical protection to the host system. In these cases, logical access control may provide a means of isolating the cryptographic software from other parts of the host system and for protecting the cryptographic software from tampering and the keys from replacement or disclosure. The use of such controls should provide the equivalent of physical protection.

Audit Trails. Cryptography may play a useful role in audit trails. For example, audit records may need to be signed. Cryptography may also be needed to protect audit records stored on computer systems from disclosure or modification. Audit trails are also used to help support electronic signatures.

Assurance. Assurance that a cryptographic module is properly and securely implemented is essential to the effective use of cryptography. NIST maintains validation programs for several of its standards for cryptography. Vendors can have their products validated for conformance to the standard through a rigorous set of tests. Such testing provides increased assurance that a module meets stated standards, and system designers, integrators, and users can have greater confidence that validated products conform to accepted standards.

NIST maintains validation programs for several of its cryptographic standards.

A cryptographic system should be monitored and periodically audited to ensure that it is satisfying its security objectives. All parameters associated with correct operation of the cryptographic system should be reviewed, and operation of the system itself should be periodically tested and the results audited. Certain information, such as secret keys or private keys in public key systems, should not be subject to audit. However, nonsecret or nonprivate keys could be used in a simulated audit procedure.

19.5 Cost Considerations

Using cryptography to protect information has both direct and indirect costs. Cost is determined in part by product availability; a wide variety of products exist for implementing cryptography in integrated circuits, add-on boards or adapters, and stand-alone units.

19.5.1 Direct Costs

The direct costs of cryptography include:

- Acquiring or implementing the cryptographic module and integrating it into the computer system. The medium (i.e., hardware, software, firmware, or combination) and various other issues such as level of security, logical and physical configuration, and special processing requirements will have an impact on cost.

- Managing the cryptography and, in particular, managing the cryptographic keys, which includes key generation, distribution, archiving, and disposition, as well as security measures to protect the keys, as appropriate.

19.5.2 Indirect Costs

The indirect costs of cryptography include:

- A decrease in system or network performance, resulting from the additional overhead of applying cryptographic protection to stored or communicated data.

- Changes in the way users interact with the system, resulting from more stringent security enforcement. However, cryptography can be made nearly transparent to the users so that the impact is minimal.

References

Alexander, M., ed. "Protecting Data With Secret Codes," *Infosecurity News.* 4(6), 1993. pp. 72-78.

American Bankers Association. *American National Standard for Financial Institution Key Management (Wholesale).* ANSI X9.17-1985. Washington, DC., 1985.

Denning, P., and D. Denning, "The Clipper and Capstone Encryption Systems." *American Scientist.* 81(4), 1993. pp. 319-323.

Diffie, W., and M. Hellman. "New Directions in Cryptography." *IEEE Transactions on Information Theory.* Vol. IT-22, No. 6, November 1976. pp. 644-654.

Duncan, R. "Encryption ABCs." *Infosecurity News.* 5(2), 1994. pp. 36-41.

International Organization for Standardization. *Information Processing Systems - Open Systems*

Interconnection Reference Model - Part 2: Security Architecture. ISO 7498/2. 1988.

Meyer, C.H., and S. M. Matyas. *Cryptography: A New Dimension in Computer Data Security.* New York, NY: John Wiley & Sons, 1982.

Nechvatal, James. *Public-Key Cryptography.* Special Publication 800-2. Gaithersburg, MD: National Institute of Standards and Technology, April 1991.

National Bureau of Standards. *Computer Data Authentication.* Federal Information Processing Standard Publication 113. May 30, 1985.

National Institute of Standards and Technology. "Advanced Authentication Technology." *Computer Systems Laboratory Bulletin.* November 1991.

National Institute of Standards and Technology. *Data Encryption Standard.* Federal Information Processing Standard Publication 46-2. December 30, 1993.

National Institute of Standards and Technology. "Digital Signature Standard." *Computer Systems Laboratory Bulletin.* January 1993.

National Institute of Standards and Technology. *Digital Signature Standard.* Federal Information Processing Standard Publication 186. May 1994.

National Institute of Standards and Technology. *Escrowed Encryption Standard.* Federal Information Processing Standard Publication 185. 1994.

National Institute of Standards and Technology. *Key Management Using ANSI X9.17.* Federal Information Processing Standard Publication 171. April 27, 1992.

National Institute of Standards and Technology. *Secure Hash Standard.* Federal Information Processing Standard Publication 180. May 11, 1993.

National Institute of Standards and Technology. *Security Requirements for Cryptographic Modules.* Federal Information Processing Standard Publication 140-1. January 11, 1994.

Rivest, R., A. Shamir, and L. Adleman. "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems." *Communications of the ACM.*, Vol. 21, No. 2, 1978. pp. 120-126.

Saltman, Roy G., ed. *Good Security Practices for Electronic Commerce, Including Electronic Data interchange.* Special Publication 800-9. Gaithersburg, MD: National Institute of Standards and Technology. December 1993.

Schneier, B. "A Taxonomy of Encryption Algorithms." *Computer Security Journal*. 9(1), 1193. pp. 39-60.

Schneier, B. "Four Crypto Standards." *Infosecurity News*. 4(2), 1993. pp. 38-39.

Schneier, B. *Applied Cryptography: Protocols, Algorithms, and Source Code in C*. New York, NY: John Wiley & Sons, Inc., 1994.

U.S. Congress, Office of Technology Assessment. "Security Safeguards and Practices." *Defending Secrets, Sharing Data: New Locks and Keys for Electronic Information*. Washington, DC: 1987, pp. 54-72.

V. EXAMPLE

Chapter 20

ASSESSING AND MITIGATING THE RISKS TO A HYPOTHETICAL COMPUTER SYSTEM

This chapter illustrates how a hypothetical government agency (HGA) deals with computer security issues in its operating environment.[140] It follows the evolution of HGA's initiation of an assessment of the threats to its computer security system all the way through to HGA's recommendations for mitigating those risks. In the real world, many solutions exist for computer security problems. No single solution can solve similar security problems in all environments. Likewise, the solutions presented in this example may not be appropriate for all environments.

This case study is provided for illustrative purposes only, and should not be construed as guidance or specific recommendations to solving specific security issues. Because a comprehensive example attempting to illustrate all handbook topics would be inordinately long, this example necessarily simplifies the issues presented and omits many details. For instance, to highlight the similarities and differences among controls in the different processing environments, it addresses some of the major types of processing platforms linked together in a distributed system: personal computers, local-area networks, wide-area networks, and mainframes; it does not show how to secure these platforms.

This example can be used to help understand how security issues are examined, how some potential solutions are analyzed, how their cost and benefits are weighed, and ultimately how management accepts responsibility for risks.

This section also highlights the importance of management's acceptance of a particular level of risk—this will, of course, vary from organization to organization. It is management's prerogative to decide what level of risk is appropriate, given operating and budget environments and other applicable factors.

20.1 Initiating the Risk Assessment

HGA has information systems that comprise and are intertwined with several different kinds of assets valuable enough to merit protection. HGA's systems play a key role in transferring U.S. Government funds to individuals in the form of paychecks; hence, financial resources are among the assets associated with HGA's systems. The system components owned and operated by HGA

[140] While this chapter draws upon many actual systems, details and characteristics were changed and merged. Although the chapter is arranged around an agency, the case study could also apply to a large division or office within an agency.

are also assets, as are personnel information, contracting and procurement documents, draft regulations, internal correspondence, and a variety of other day-to-day business documents, memos, and reports. HGA's assets include intangible elements as well, such as reputation of the agency and the confidence of its employees that personal information will be handled properly and that the wages will be paid on time.

A recent change in the directorship of HGA has brought in a new management team. Among the new Chief Information Officer's first actions was appointing a Computer Security Program Manager who immediately initiated a comprehensive risk analysis to assess the soundness of HGA's computer security program in protecting the agency's assets and its compliance with federal directives. This analysis drew upon prior risk assessments, threat studies, and applicable internal control reports. The Computer Security Program Manager also established a timetable for periodic reassessments.

Since the wide-area network and mainframe used by HGA are owned and operated by other organizations, they were not treated in the risk assessment as HGA's assets. And although HGA's personnel, buildings, and facilities are essential assets, the Computer Security Program Manager considered them to be outside the scope of the risk analysis.

After examining HGA's computer system, the risk assessment team identified specific threats to HGA's assets, reviewed HGA's and national safeguards against those threats, identified the vulnerabilities of those policies, and recommended specific actions for mitigating the remaining risks to HGA's computer security. The following sections provide highlights from the risk assessment. The assessment addressed many other issues at the programmatic and system levels. However, this chapter focuses on security issues related to the time and attendance application. (Other issues are discussed in Chapter 6.)

20.2 HGA's Computer System

HGA relies on the distributed computer systems and networks shown in Figure 20.1. They consist of a collection of components, some of which are systems in their own right. Some belong to HGA, but others are owned and operated by other organizations. This section describes these components, their role in the overall distributed system architecture, and how they are used by HGA.

20.2.1 System Architecture

Most of HGA's staff (a mix of clerical, technical, and managerial staff) are provided with personal computers (PCs) located in their offices. Each PC includes hard-disk and floppy-disk drives.

The PCs are connected to a local area network (LAN) so that users can exchange and share

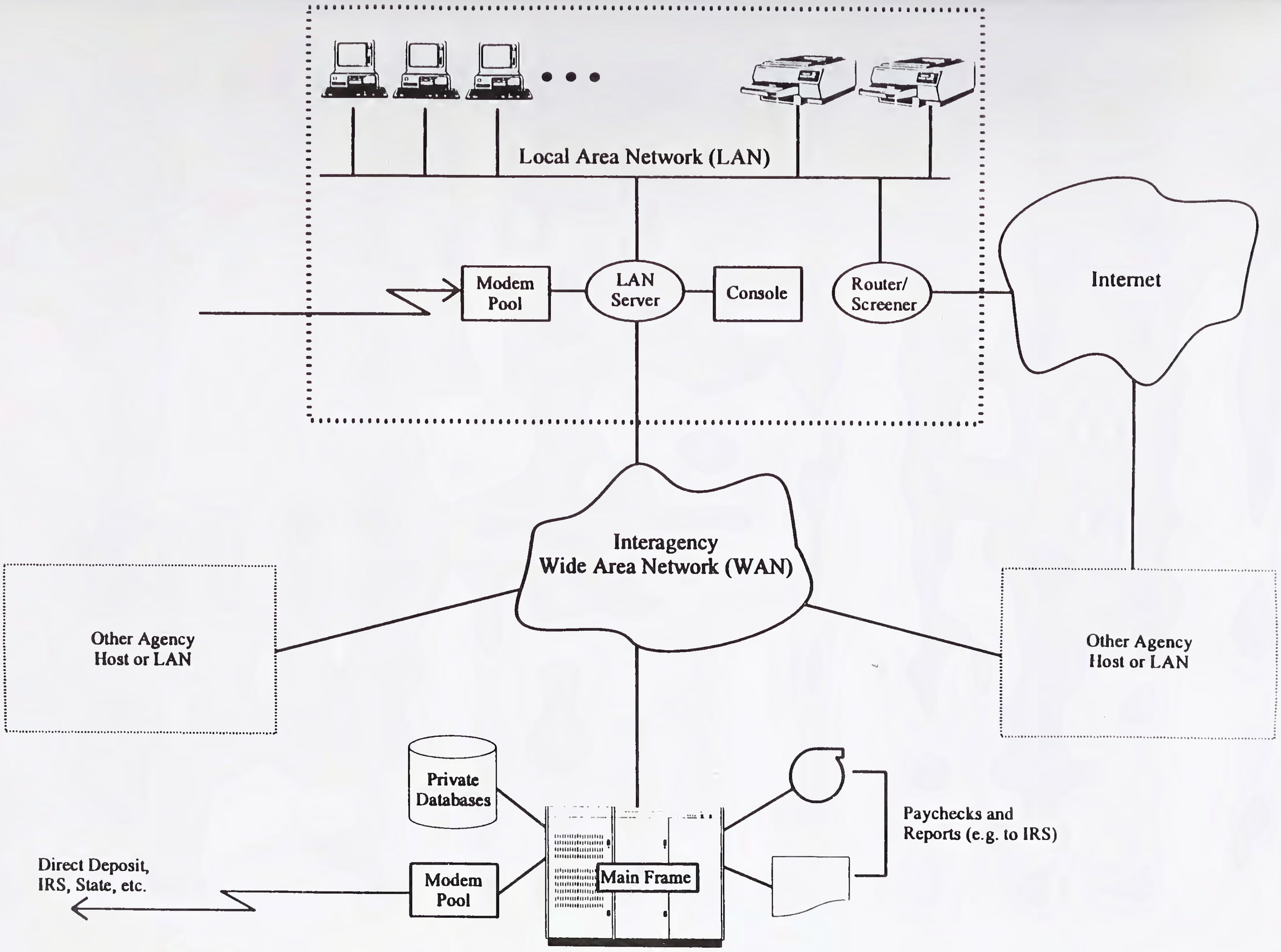

Local Area Network (LAN)
Modem Pool
LAN Server
Console
Router/ Screener
Internet
Interagency Wide Area Network (WAN)
Other Agency Host or LAN
Other Agency Host or LAN
Private Databases
Main Frame
Modem Pool
Direct Deposit, IRS, State, etc.
Paychecks and Reports (e.g. to IRS)

Figure 20.1

information. The central component of the LAN is a *LAN server*, a more powerful computer that acts as an intermediary between PCs on the network and provides a large volume of disk storage for shared information, including shared application programs. The server provides logical access controls on potentially sharable information via elementary access control lists. These access controls can be used to limit user access to various files and programs stored on the server. Some programs stored on the server can be retrieved via the LAN and executed on a PC; others can only be executed on the server.

To initiate a session on the network or execute programs on the server, users at a PC must log into the server and provide a user identifier and password known to the server. Then they may use files to which they have access.

One of the applications supported by the server is *electronic mail* (e-mail), which can be used by all PC users. Other programs that run on the server can only be executed by a limited set of PC users.

Several printers, distributed throughout HGA's building complex, are connected to the LAN. Users at PCs may direct printouts to whichever printer is most convenient for their use.

Since HGA must frequently communicate with industry, the LAN also provides a connection to the Internet via a *router*. The router is a network interface device that translates between the protocols and addresses associated with the LAN and the Internet. The router also performs *network packet filtering*, a form of network access control, and has recently been configured to disallow non–e-mail (e.g., file transfer, remote log-in) between LAN and Internet computers.

The LAN server also has connections to several other devices.

- A *modem pool* is provided so that HGA's employees on travel can "dial up" via the public switched (telephone) network and read or send e-mail. To initiate a dial-up session, a user must successfully log in. During dial-up sessions, the LAN server provides access only to e-mail facilities; no other functions can be invoked.

- A *special console* is provided for the server administrators who configure the server, establish and delete user accounts, and have other special privileges needed for administrative and maintenance functions. These functions can only be invoked from the *administrator console*; that is, they cannot be invoked from a PC on the network or from a dial-up session.

- A *connection to a government agency X.25-based wide-area network* (WAN) is provided so that information can be transferred to or from other agency systems. One of the other hosts on the WAN is a large multiagency mainframe system. This mainframe is used to collect and process information from a large number of agencies while providing a range of access controls.

20.2.2 System Operational Authority/Ownership

The system components contained within the large dashed rectangle shown in Figure 20.1 are managed and operated by an organization within HGA known as the Computer Operations Group (COG). This group includes the PCs, LAN, server, console, printers, modem pool, and router. The WAN is owned and operated by a large commercial telecommunications company that provides WAN services under a government contract. The mainframe is owned and operated by a federal agency that acts as a service provider for HGA and other agencies connected to the WAN.

20.2.3 System Applications

PCs on HGA's LAN are used for word processing, data manipulation, and other common applications, including spreadsheet and project management tools. Many of these tasks are concerned with data that are sensitive with respect to confidentiality or integrity. Some of these documents and data also need to be available in a timely manner.

The mainframe also provides storage and retrieval services for other databases belonging to individual agencies. For example, several agencies, including HGA, store their personnel databases on the mainframe; these databases contain dates of service, leave balances, salary and W-2 information, and so forth.

In addition to their time and attendance application, HGA's PCs and the LAN server are used to manipulate other kinds of information that may be sensitive with respect to confidentiality or integrity, including personnel-related correspondence and draft contracting documents.

20.3 Threats to HGA's Assets

Different assets of HGA are subject to different kinds of threats. Some threats are considered less likely than others, and the potential impact of different threats may vary greatly. The likelihood of threats is generally difficult to estimate accurately. Both HGA and the risk assessment's authors have attempted to the extent possible to base these estimates on historical data, but have also tried to anticipate new trends stimulated by emerging technologies (e.g., external networks).

20.3.1 Payroll Fraud

As for most large organizations that control financial assets, attempts at fraud and embezzlement are likely to occur. Historically, attempts at payroll fraud have almost always come from within HGA or the other agencies that operate systems on which HGA depends. Although HGA has thwarted many of these attempts, and some have involved relatively small sums of money, it considers preventing financial fraud to be a *critical* computer security priority, particularly in light of the potential financial losses and the risks of damage to its reputation with Congress, the public, and other federal agencies.

Attempts to defraud HGA have included the following:

- Submitting fraudulent time sheets for hours or days not worked, or for pay periods following termination or transfer of employment. The former may take the form of overreporting compensatory or overtime hours worked, or underreporting vacation or sick leave taken. Alternatively, attempts have been made to modify time sheet data after being entered and approved for submission to payroll.

- Falsifying or modifying dates or data on which one's "years of service" computations are based, thereby becoming eligible for retirement earlier than allowed, or increasing one's pension amount.

- Creating employee records and time sheets for fictitious personnel, and attempting to obtain their paychecks, particularly after arranging for direct deposit.

20.3.2 Payroll Errors

Of greater likelihood, but of perhaps lesser potential impact on HGA, are errors in the entry of time and attendance data; failure to enter information describing new employees, terminations, and transfers in a timely manner; accidental corruption or loss of time and attendance data; or errors in interagency coordination and processing of personnel transfers.

Errors of these kinds can cause financial difficulties for employees and accounting problems for HGA. If an employee's vacation or sick leave balance became negative erroneously during the last pay period of the year, the employee's last paycheck would be automatically reduced. An individual who transfers between HGA and another agency may risk receiving duplicate paychecks or no paychecks for the pay periods immediately following the transfer. Errors of this sort that occur near the end of the year can lead to errors in W-2 forms and subsequent difficulties with the tax collection agencies.

20.3.3 Interruption of Operations

HGA's building facilities and physical plant are several decades old and are frequently under repair or renovation. As a result, power, air conditioning, and LAN or WAN connectivity for the server are typically interrupted several times a year for periods of up to one work day. For example, on several occasions, construction workers have inadvertently severed power or network cables. Fires, floods, storms, and other natural disasters can also interrupt computer operations, as can equipment malfunctions.

Another threat of small likelihood, but significant potential impact, is that of a malicious or disgruntled employee or outsider seeking to disrupt time-critical processing (e.g., payroll) by deleting necessary inputs or system accounts, misconfiguring access controls, planting computer viruses, or stealing or sabotaging computers or related equipment. Such interruptions, depending upon when they occur, can prevent time and attendance data from getting processed and transferred to the mainframe before the payroll processing deadline.

20.3.4 Disclosure or Brokerage of Information

Other kinds of threats may be stimulated by the growing market for information about an organization's employees or internal activities. Individuals who have legitimate work-related reasons for access to the master employee database may attempt to disclose such information to other employees or contractors or to sell it to private investigators, employment recruiters, the press, or other organizations. HGA considers such threats to be moderately likely and of low to high potential impact, depending on the type of information involved.

20.3.5 Network-Related Threats

Most of the human threats of concern to HGA originate from insiders. Nevertheless, HGA also recognizes the need to protect its assets from outsiders. Such attacks may serve many different purposes and pose a broad spectrum of risks, including unauthorized disclosure or modification of information, unauthorized use of services and assets, or unauthorized denial of services.

As shown in Figure 20.1, HGA's systems are connected to the three external networks: (1) the Internet, (2) the Interagency WAN, and (3) the public-switched (telephone) network. Although these networks are a source of security risks, connectivity with them is essential to HGA's mission and to the productivity of its employees; connectivity cannot be terminated simply because of security risks.

In each of the past few years before establishing its current set of network safeguards, HGA had detected several attempts by outsiders to penetrate its systems. Most, but not all of these, have come from the Internet, and those that succeeded did so by learning or guessing user account passwords. In two cases, the attacker deleted or corrupted significant amounts of data, most of which were later restored from backup files. In most cases, HGA could detect no ill effects of the attack, but concluded that the attacker may have browsed through some files. HGA also conceded that its systems did not have audit logging capabilities sufficient to track an attacker's activities. Hence, for most of these attacks, HGA could not accurately gauge the extent of penetration.

In one case, an attacker made use of a bug in an e-mail utility and succeeded in acquiring System Administrator privileges on the server—a significant breach. HGA found no evidence that the attacker attempted to exploit these privileges before being discovered two days later. When the attack was detected, COG immediately contacted the HGA's Incident Handling Team, and was told that a bug fix had been distributed by the server vendor several months earlier. To its embarrassment, COG discovered that it had already received the fix, which it then promptly installed. It now believes that no subsequent attacks of the same nature have succeeded.

Although HGA has no evidence that it has been significantly harmed to date by attacks via external networks, it believes that these attacks have great potential to inflict damage. HGA's management considers itself lucky that such attacks have not harmed HGA's reputation and the confidence of the citizens its serves. It also believes the likelihood of such attacks via external networks will increase in the future.

20.3.6 Other Threats

HGA's systems also are exposed to several other threats that, for reasons of space, cannot be fully enumerated here. Examples of threats and HGA's assessment of their probabilities and impacts include those listed in Table 20.1.

20.4 Current Security Measures

HGA has numerous policies and procedures for protecting its assets against the above threats. These are articulated in HGA's *Computer Security Manual*, which implements and synthesizes the requirements of many federal directives, such as Appendix III to OMB Circular A-130, the Computer Security Act of 1987, and the Privacy Act. The manual also includes policies for automated financial systems, such as those based on OMB Circulars A-123 and A-127, as well as the Federal Managers' Financial Integrity Act.

Several examples of those policies follow, as they apply generally to the use and administration of HGA's computer system and specifically to security issues related to time and attendance, payroll, and continuity of operations.

20.4.1 General Use and Administration of HGA's Computer System

HGA's Computer Operations Group (COG) is responsible for controlling, administering, and maintaining the computer resources owned and operated by HGA. These functions are depicted in Figure 20.1 enclosed in the large, dashed rectangle. Only individuals holding the job title System Administrator are authorized to establish log-in IDs and passwords on multiuser HGA systems (e.g., the LAN server). Only HGA's employees and contract personnel may use the system, and only after receiving written authorization from the department supervisor (or, in the case of contractors, the contracting officer) to whom these individuals report.

COG issues copies of all relevant security policies and procedures to new users. Before activating a system account for a new users, COG requires that they (1) attend a security awareness and training course or complete an interactive computer-aided-instruction training session and (2) sign an acknowledgment form indicating that they understand their security responsibilities.

Authorized users are assigned a secret log-in ID and password, which they must not share with anyone else. They are expected to comply with all of HGA's password selection and security procedures (e.g., periodically changing passwords). Users who fail to do so are subject to a range of penalties.

Examples of Threats to HGA Systems

Potential Threat	**Probability**	**Impact**
Accidental Loss/Release of Disclosure-Sensitive Information	Medium	Low/Medium
Accidental Destruction of Information	High	Medium
Loss of Information due to Virus Contamination	Medium	Medium
Misuse of System Resources	Low	Low
Theft	High	Medium
Unauthorized Access to Telecommunications Resources[*]	Medium	Medium
Natural Disaster	Low	High

[*] HGA operates a PBX system, which may be vulnerable to (1) hacker disruptions of PBX availability and, consequently, agency operations, (2) unauthorized access to outgoing phone lines for long-distance services, (3) unauthorized access to stored voice-mail messages, and (4) surreptitious access to otherwise private conversations/data transmissions.

Table 20.1

Users creating data that are sensitive with respect to disclosure or modification are expected to make effective use of the automated access control mechanisms available on HGA computers to reduce the risk of exposure to unauthorized individuals. (Appropriate training and education are in place to help users do this.) In general, access to disclosure-sensitive information is to be granted only to individuals whose jobs require it.

20.4.2 Protection Against Payroll Fraud and Errors: Time and Attendance Application

The time and attendance application plays a major role in protecting against payroll fraud and errors. Since the time and attendance application is a component of a larger automated payroll process, many of its functional and security requirements have been derived from both governmentwide and HGA-specific policies related to payroll and leave. For example, HGA must protect personal information in accordance with the Privacy Act. Depending on the specific type of information, it should normally be viewable only by the individual concerned, the individual's supervisors, and personnel and payroll department employees. Such information should also be timely and accurate.

Each week, employees must sign and submit a time sheet that identifies the number of hours they have worked and the amount of leave they have taken. The Time and Attendance Clerk enters the data for a given group of employees and runs an application on the LAN server to verify the data's validity and to ensure that only authorized users with access to the Time and Attendance Clerk's functions can enter time and attendance data. The application performs these security checks by using the LAN server's access control and identification and authentication (I&A) mechanisms. The application compares the data with a limited database of employee information to detect incorrect employee identifiers, implausible numbers of hours worked, and so forth. After correcting any detected errors, the clerk runs another application that formats the time and attendance data into a report, flagging exception/out-of-bound conditions (e.g., negative leave balances).

Department supervisors are responsible for reviewing the correctness of the time sheets of the employees under their supervision and indicating their approval by initialing the time sheets. If they detect significant irregularities and indications of fraud in such data, they must report their findings to the Payroll Office before submitting the time sheets for processing. In keeping with the principle of separation of duty, all data on time sheets and corrections on the sheets that may affect pay, leave, retirement, or other benefits of an individual must be reviewed for validity by at least two authorized individuals (other than the affected individual).

Protection Against Unauthorized Execution

Only users with access to Time and Attendance Supervisor functions may approve and submit time and attendance data — or subsequent corrections thereof — to the mainframe. Supervisors may not approve their own time and attendance data.

Only the System Administrator has been granted access to assign a special access control privilege to server programs. As a result, the server's operating system is designed to prevent a bogus time and attendance application created by any other user from communicating with the WAN and, hence, with the mainframe.

The time and attendance application is supposed to be configured so that the clerk and supervisor functions can only be carried out from specific PCs attached to the LAN and only during normal working hours. Administrators are not authorized to exercise functions of the time and attendance application apart from those concerned with configuring the accounts, passwords, and access permissions for clerks and supervisors. Administrators are expressly prohibited by policy from entering, modifying, or submitting time and attendance data via the time and attendance application or other mechanisms.[141]

Protection against unauthorized execution of the time and attendance application depends on I&A and access controls. While the time and attendance application is accessible from any PC, unlike most programs run by PC users, it does not execute directly on the PC's processor. Instead, it executes on the server, while the PC behaves as a terminal, relaying the user's keystrokes to the server and displaying text and graphics sent from the server. The reason for this approach is that common PC systems do not provide I&A and access controls and, therefore, cannot protect against unauthorized time and attendance program execution. *Any* individual who has access to the PC could run any program stored there.

Another possible approach is for the time and attendance program to perform I&A and access control on its own by requesting and validating a password before beginning each time and attendance session. This approach, however, can be defeated easily by a moderately skilled programming attack, and was judged inadequate by HGA during the application's early design phase.

Recall that the server is a more powerful computer equipped with a multiuser operating system that includes password-based I&A and access controls. Designing the time and attendance application program so that it executes on the server under the control of the server's operating system provides a more effective safeguard against unauthorized execution than executing it on the user's PC.

Protection Against Payroll Errors

The frequency of data entry errors is reduced by having Time and Attendance clerks enter each time sheet into the time and attendance application twice. If the two copies are identical, both are considered error free, and the record is accepted for subsequent review and approval by a supervisor. If the copies are not identical, the discrepancies are displayed, and for each discrepancy, the clerk determines which copy is correct. The clerk then incorporates the corrections into one of the copies, which is then accepted for further processing. If the clerk

[141] Technically, Systems Administrators may still have the ability to do so. This highlights the importance of adequate managerial reviews, auditing, and personnel background checks.

makes the same data-entry error twice, then the two copies will match, and one will be accepted as correct, even though it is erroneous. To reduce this risk, the time and attendance application could be configured to require that the two copies be entered by different clerks.

In addition, each department has one or more Time and Attendance Supervisors who are authorized to review these reports for accuracy and to approve them by running another server program that is part of the time and attendance application. The data are then subjected to a collection of "sanity checks" to detect entries whose values are outside expected ranges. Potential anomalies are displayed to the supervisor prior to allowing approval; if errors are identified, the data are returned to a clerk for additional examination and corrections.

When a supervisor approves the time and attendance data, this application logs into the interagency mainframe via the WAN and transfers the data to a payroll database on the mainframe. The mainframe later prints paychecks or, using a pool of modems that can send data over phone lines, it may transfer the funds electronically into employee-designated bank accounts. Withheld taxes and contributions are also transferred electronically in this manner.

The Director of Personnel is responsible for ensuring that forms describing significant payroll-related personnel actions are provided to the Payroll Office at least one week before the payroll processing date for the first affected pay period. These actions include hiring, terminations, transfers, leaves of absences and returns from such, and pay raises.

The Manager of the Payroll Office is responsible for establishing and maintaining controls adequate to ensure that the amounts of pay, leave, and other benefits reported on pay stubs and recorded in permanent records and those distributed electronically are accurate and consistent with time and attendance data and with other information provided by the Personnel Department. In particular, paychecks must never be provided to anyone who is not a bona fide, active-status employee of HGA. Moreover, the pay of any employee who terminates employment, who transfers, or who goes on leave without pay must be suspended as of the effective date of such action; that is, extra paychecks or excess pay must not be dispersed.

Protection Against Accidental Corruption or Loss of Payroll Data

The same mechanisms used to protect against fraudulent modification are used to protect against accidental corruption of time and attendance data — namely, the access-control features of the server and mainframe operating systems.

COG's nightly backups of the server's disks protect against loss of time and attendance data. To a limited extent, HGA also relies on mainframe administrative personnel to back up time and attendance data stored on the mainframe, even though HGA has no direct control over these individuals. As additional protection against loss of data at the mainframe, HGA retains copies of all time and attendance data on line on the server for at least one year, at which time the data are archived and kept for three years. The server's access controls for the on-line files are automatically set to read-only access by the time and attendance application at the time of submission to the mainframe. The integrity of time and attendance data will be protected by digital signatures as they are implemented.

The WAN's communications protocols also protect against loss of data during transmission from the server to the mainframe (e.g., error checking). In addition, the mainframe payroll application includes a program that is automatically run 24 hours before paychecks and pay stubs are printed. This program produces a report identifying agencies from whom time and attendance data for the current pay period were expected but not received. Payroll department staff are responsible for reviewing the reports and immediately notifying agencies that need to submit or resubmit time and attendance data. If time and attendance input or other related information is not available on a timely basis, pay, leave, and other benefits are temporarily calculated based on information estimated from prior pay periods.

20.4.3 Protection Against Interruption of Operations

HGA's policies regarding continuity of operations are derived from requirements stated in OMB Circular A-130. HGA requires various organizations within it to develop contingency plans, test them annually, and establish appropriate administrative and operational procedures for supporting them. The plans must identify the facilities, equipment, supplies, procedures, and personnel needed to ensure reasonable continuity of operations under a broad range of adverse circumstances.

COG Contingency Planning

COG is responsible for developing and maintaining a contingency plan that sets forth the procedures and facilities to be used when physical plant failures, natural disasters, or major equipment malfunctions occur sufficient to disrupt the normal use of HGA's PCs, LAN, server, router, printers, and other associated equipment.

The plan prioritizes applications that rely on these resources, indicating those that should be suspended if available automated functions or capacities are temporarily degraded. COG personnel have identified system software and hardware components that are compatible with those used by two nearby agencies. HGA has signed an agreement with those agencies, whereby they have committed to reserving spare computational and storage capacities sufficient to support HGA's system-based operations for a few days during an emergency.

No communication devices or network interfaces may be connected to HGA's systems without written approval of the COG Manager. The COG staff is responsible for installing all known security-related software patches in a timely manner and for maintaining spare or redundant PCs, servers, storage devices, and LAN interfaces to ensure that at least 100 people can simultaneously perform word processing tasks at all times.

To protect against accidental corruption or loss of data, COG personnel back up the LAN server's disks onto magnetic tape every night and transport the tapes weekly to a sister agency for storage. HGA's policies also stipulate that all PC users are responsible for backing up weekly any significant data stored on their PC's local hard disks. For the past several years, COG has issued a yearly memorandum reminding PC users of this responsibility. COG also strongly encourages them to store significant data on the LAN server instead of on their PC's hard disk so that such data will be backed up automatically during COG's LAN server backups.

To prevent more limited computer equipment malfunctions from interrupting routine business operations, COG maintains an inventory of approximately ten fully equipped spare PC's, a spare LAN server, and several spare disk drives for the server. COG also keeps thousands of feet of LAN cable on hand. If a segment of the LAN cable that runs through the ceilings and walls of HGA's buildings fails or is accidentally severed, COG technicians will run temporary LAN cabling along the floors of hallways and offices, typically restoring service within a few hours for as long as needed until the cable failure is located and repaired.

To protect against PC virus contamination, HGA authorizes only System Administrators approved by the COG Manager to install licensed, copyrighted PC software packages that appear on the COG-approved list. PC software applications are generally installed only on the server. (These stipulations are part of an HGA assurance strategy that relies on the quality of the engineering practices of vendors to provide software that is adequately robust and trustworthy.) Only the COG Manager is authorized to add packages to the approved list. COG procedures also stipulate that every month System Administrators should run virus-detection and other security-configuration validation utilities on the server and, on a spot-check basis, on a number of PCs. If they find a virus, they must immediately notify the agency team that handles computer security incidents.

COG is also responsible for reviewing audit logs generated by the server, identifying audit records indicative of security violations, and reporting such indications to the Incident-Handling Team. The COG Manager assigns these duties to specific members of the staff and ensures that they are implemented as intended.

The COG Manager is responsible for assessing adverse circumstances and for providing recommendations to HGA's Director. Based on these and other sources of input, the Director will determine whether the.circumstances are dire enough to merit activating various sets of procedures called for in the contingency plan.

Division Contingency Planning

HGA's divisions also must develop and maintain their own contingency plans. The plans must identify critical business functions, the system resources and applications on which they depend, and the maximum acceptable periods of interruption that these functions can tolerate without significant reduction in HGA's ability to fulfill its mission. The head of each division is responsible for ensuring that the division's contingency plan and associated support activities are adequate.

For each major application used by multiple divisions, a chief of a single division must be designated as the *application owner.* The designated official (supported by his or her staff) is responsible for addressing that application in the contingency plan and for coordinating with other divisions that use the application.

If a division relies exclusively on computer resources maintained by COG (e.g., the LAN), it need not duplicate COG's contingency plan, but is responsible for reviewing the adequacy of that plan. If COG's plan does not adequately address the division's needs, the division must communicate its concerns to the COG Director. In either situation, the division must make known the criticality of its applications to the COG. If the division relies on computer resources or services that are *not* provided by COG, the division is responsible for (1) developing its own contingency plan or (2) ensuring that the contingency plans of other organizations (e.g., the WAN service provider) provide adequate protection against service disruptions.

20.4.4 Protection Against Disclosure or Brokerage of Information

HGA's protection against information disclosure is based on a need-to-know policy and on personnel hiring and screening practices. The need-to-know policy states that time and attendance information should be made accessible only to HGA employees and contractors whose assigned professional responsibilities require it. Such information must be protected against access from all other individuals, including other HGA employees. Appropriate hiring and screening practices can lessen the risk that an untrustworthy individual will be assigned such responsibilities.

The need-to-know policy is supported by a collection of physical, procedural, and automated safeguards, including the following:

- Time and attendance paper documents are must be stored securely when not in use, particularly during evenings and on weekends. Approved storage containers include locked file cabinets and desk drawers—to which only the owner has the keys. While storage in a container is preferable, it is also permissible to leave time and attendance documents on top of a desk or other exposed surface in a locked office (with the realization that the guard force has keys to the office). (This is a judgment left to local discretion.) Similar rules apply to disclosure-sensitive information stored on floppy disks and other removable magnetic media.

- Every HGA PC is equipped with a key lock that, when locked, disables the PC.

When information is stored on a PC's local hard disk, the user to whom that PC was assigned is expected to (1) lock the PC at the conclusion of each work day and (2) lock the office in which the PC is located.

- The LAN server operating system's access controls provide extensive features for controlling access to files. These include group-oriented controls that allow teams of users to be assigned to named groups by the System Administrator. Group members are then allowed access to sensitive files not accessible to nonmembers. Each user can be assigned to several groups according to need to know. (The reliable functioning of these controls is assumed, perhaps incorrectly, by HGA.)

- All PC users undergo security awareness training when first provided accounts on the LAN server. Among other things, the training stresses the necessity of protecting passwords. It also instructs users to log off the server before going home at night or before leaving the PC unattended for periods exceeding an hour.

20.4.5 Protection Against Network-Related Threats

HGA's current set of external network safeguards has only been in place for a few months. The basic approach is to tightly restrict the kinds of external network interactions that can occur by funneling all traffic to and from external networks through two interfaces that filter out unauthorized kinds of interactions. As indicated in Figure 20.1, the two interfaces are the network router and the LAN server. The only kinds of interactions that these interfaces allow are (1) e-mail and (2) data transfers from the server to the mainframe controlled by a few special applications (e.g., the time and attendance application).

Figure 20.1 shows that the network router is the only direct interface between the LAN and the Internet. The router is a dedicated special-purpose computer that translates between the protocols and addresses associated with the LAN and the Internet. Internet protocols, unlike those used on the WAN, specify that packets of information coming from or going to the Internet must carry an indicator of the kind of service that is being requested or used to process the information. This makes it possible for the router to distinguish e-mail packets from other kinds of packets—for example, those associated with a remote log-in request.[142] The router has been configured by COG to discard all packets coming from or going to the Internet, except those associated with e-mail. COG personnel believe that the router effectively eliminates Internet-based attacks on HGA user accounts because it disallows all remote log-in sessions, even those accompanied by a legitimate password.

[142] Although not discussed in this example, recognize that technical "spoofing" can occur.

The LAN server enforces a similar type of restriction for dial-in access via the public-switched network. The access controls provided by the server's operating system have been configured so that during dial-in sessions, only the e-mail utility can be executed. (HGA policy, enforced by periodic checks, prohibits installation of modems on PCs, so that access must be through the LAN server.) In addition, the server's access controls have been configured so that its WAN interface device is accessible only to programs that possess a special access-control privilege. Only the System Administrator can assign this privilege to server programs, and only a handful of special-purpose applications, like the time and attendance application, have been assigned this privilege.

20.4.6 Protection Against Risks from Non–HGA Computer Systems

HGA relies on systems and components that it cannot control directly because they are owned by other organizations. HGA has developed a policy to avoid undue risk in such situations. The policy states that system components controlled and operated by organizations other than HGA may not be used to process, store, or transmit HGA information without obtaining explicit permission from the application owner and the COG Manager. Permission to use such system components may not be granted without written commitment from the controlling organization that HGA's information will be safeguarded commensurate with its value, as designated by HGA. This policy is somewhat mitigated by the fact that HGA has developed an issue-specific policy on the use of the Internet, which allows for its use for e-mail with outside organizations and access to other resources (but not for transmission of HGA's proprietary data).

20.5 Vulnerabilities Reported by the Risk Assessment Team

The risk assessment team found that many of the risks to which HGA is exposed stem from (1) the failure of individuals to comply with established policies and procedures or (2) the use of automated mechanisms whose assurance is questionable because of the ways they have been developed, tested, implemented, used, or maintained. The team also identified specific vulnerabilities in HGA's policies and procedures for protecting against payroll fraud and errors, interruption of operations, disclosure and brokering of confidential information, and unauthorized access to data by outsiders.

20.5.1 Vulnerabilities Related to Payroll Fraud

Falsified Time Sheets

The primary safeguards against falsified time sheets are review and approval by supervisory personnel, who are not permitted to approve their own time and attendance data. The risk assessment has concluded that, while imperfect, these safeguards are adequate. The related requirement that a clerk and a supervisor must cooperate closely in creating time and attendance data and submitting the data to the mainframe also safeguards against other kinds of illicit manipulation of time and attendance data by clerks or supervisors acting independently.

Unauthorized Access

When a PC user enters a password to the server during I&A, the password is sent to the server by broadcasting it over the LAN "in the clear." This allows the password to be intercepted easily by any other PC connected to the LAN. In fact, so-called "password sniffer" programs that capture passwords in this way are widely available. Similarly, a malicious program planted on a PC could also intercept passwords before transmitting them to the server. An unauthorized individual who obtained the captured passwords could then run the time and attendance application in place of a clerk or supervisor. Users might also store passwords in a log-on script file.

Bogus Time and Attendance Applications

The server's access controls are probably adequate for protection against bogus time and attendance applications that run on the server. However, the server's operating system and access controls have only been in widespread use for a few years and contain a number of security-related bugs. And the server's access controls are ineffective if not properly configured, and the administration of the server's security features in the past has been notably lax.

Unauthorized Modification of Time and Attendance Data

Protection against unauthorized modification of time and attendance data requires a variety of safeguards because each system component on which the data are stored or transmitted is a potential source of vulnerabilities.

First, the time and attendance data are entered on the server by a clerk. On occasion, the clerk may begin data entry late in the afternoon, and complete it the following morning, storing it in a temporary file between the two sessions. One way to avoid unauthorized modification is to store the data on a diskette and lock it up overnight. After being entered, the data will be stored in another temporary file until reviewed and approved by a supervisor. These files, now stored on the system, must be protected against tampering. As before, the server's access controls, if reliable and properly configured, can provide such protection (as can digital signatures, as discussed later) in conjunction with proper auditing.

Second, when the Supervisor approves a batch of time and attendance data, the time and attendance application sends the data over the WAN to the mainframe. The WAN is a collection of communications equipment and special-purpose computers called "switches" that act as relays, routing information through the network from source to destination. Each switch is a potential site at which the time and attendance data may be fraudulently modified. For example, an HGA PC user might be able to intercept time and attendance data and modify the data enroute to the payroll application on the mainframe. Opportunities include tampering with incomplete time and attendance input files while stored on the server, interception and tampering during WAN transit, or tampering on arrival to the mainframe prior to processing by the payroll application.

Third, on arrival at the mainframe, the time and attendance data are held in a temporary file on the mainframe until the payroll application is run. Consequently, the mainframe's I&A and access controls must provide a critical element of protection against unauthorized modification of the data.

According to the risk assessment, the server's access controls, with prior caveats, probably provide acceptable protection against unauthorized modification of data stored on the server. The assessment concluded that a WAN-based attack involving collusion between an employee of HGA and an employee of the WAN service provider, although unlikely, should not be dismissed entirely, especially since HGA has only cursory information about the service provider's personnel security practices and no contractual authority over how it operates the WAN.

The greatest source of vulnerabilities, however, is the mainframe. Although its operating system's access controls are mature and powerful, it uses password-based I&A. This is of particular concern, because it serves a large number of federal agencies via WAN connections. A number of these agencies are known to have poor security programs. As a result, one such agency's systems could be penetrated (e.g., from the Internet) and then used in attacks on the mainframe via the WAN. In fact, time and attendance data awaiting processing on the mainframe would probably not be as attractive a target to an attacker as other kinds of data or, indeed, disabling the system, rendering it unavailable. For example, an attacker might be able to modify the employee data base so that it disbursed paychecks or pensions checks to fictitious employees. Disclosure-sensitive law enforcement databases might also be attractive targets.

The access control on the mainframe is strong and provides good protection against intruders breaking into a second application after they have broken into a first. However, previous audits have shown that the difficulties of system administration may present some opportunities for intruders to defeat access controls.

20.5.2 Vulnerabilities Related to Payroll Errors

HGA's management has established procedures for ensuring the timely submission and interagency coordination of paperwork associated with personnel status changes. However, an unacceptably large number of troublesome payroll errors during the past several years has been traced to the late submission of personnel paperwork. The risk assessment documented the adequacy of HGA's safeguards, but criticized the managers for not providing sufficient incentives for compliance.

20.5.3 Vulnerabilities Related to Continuity of Operations

COG Contingency Planning

The risk assessment commended HGA for many aspects of COG's contingency plan, but pointed out that many COG personnel were completely unaware of the responsibilities the plan assigned to them. The assessment also noted that although HGA's policies require annual testing of contingency plans, the capability to resume HGA's computer-processing activities at another cooperating agency has never been verified and may turn out to be illusory.

Division Contingency Planning

The risk assessment reviewed a number of the application-oriented contingency plans developed by HGA's divisions (including plans related to time and attendance). Most of the plans were cursory and attempted to delegate nearly all contingency planning responsibility to COG. The assessment criticized several of these plans for failing to address potential disruptions caused by lack of access to (1) computer resources not managed by COG and (2) nonsystem resources, such as buildings, phones, and other facilities. In particular, the contingency plan encompassing the time and attendance application was criticized for not addressing disruptions caused by WAN and mainframe outages.

Virus Prevention

The risk assessment found HGA's virus-prevention policy and procedures to be sound, but noted that there was little evidence that they were being followed. In particular, no COG personnel interviewed had ever run a virus scanner on a PC on a routine basis, though several had run them during publicized virus scares. The assessment cited this as a significant risk item.

Accidental Corruption and Loss of Data

The risk assessment concluded that HGA's safeguards against accidental corruption and loss of time and attendance data were adequate, but that safeguards for some other kinds of data were not. The assessment included an informal audit of a dozen randomly chosen PCs and PC users in the agency. It concluded that many PC users store significant data on their PC's hard disks, but do not back them up. Based on anecdotes, the assessment's authors stated that there appear to have been many past incidents of loss of information stored on PC hard disks and predicted that such losses would continue.

20.5.4 Vulnerabilities Related to Information Disclosure/Brokerage

HGA takes a conservative approach toward protecting information about its employees. Since information brokerage is more likely to be a threat to large collections of data, HGA risk assessment focused primarily, but not exclusively, on protecting the mainframe.

The risk assessment concluded that significant, avoidable information brokering vulnerabilities were present—particularly due to HGA's lack of compliance with its own policies and procedures. Time and attendance documents were typically not stored securely after hours, and few PCs containing time and attendance information were routinely locked. Worse yet, few were routinely powered down, and many were left logged into the LAN server overnight. These practices make it easy for an HGA employee wandering the halls after hours to browse or copy time and attendance information on another employee's desk, PC hard disk, or LAN server directories.

The risk assessment pointed out that information sent to or retrieved from the server is subject to eavesdropping by other PCs on the LAN. The LAN hardware transmits information by broadcasting it to all connection points on the LAN cable. Moreover, information sent to or retrieved from the server is transmitted in the clear—that is, without encryption. Given the widespread availability of LAN "sniffer" programs, LAN eavesdropping is trivial for a prospective information broker and, hence, is likely to occur.

Last, the assessment noted that HGA's employee master database is stored on the mainframe, where it might be a target for information brokering by employees of the agency that owns the mainframe. It might also be a target for information brokering, fraudulent modification, or other illicit acts by any outsider who penetrates the mainframe via another host on the WAN.

20.5.5 Network-Related Vulnerabilities

The risk assessment concurred with the general approach taken by HGA, but identified several vulnerabilities. It reiterated previous concerns about the lack of assurance associated with the server's access controls and pointed out that these play a critical role in HGA's approach. The assessment noted that the e-mail utility allows a user to include a copy of *any* otherwise accessible file in an outgoing mail message. If an attacker dialed in to the server and succeeded in logging in as an HGA employee, the attacker could use the mail utility to export copies of all the files accessible to that employee. In fact, copies could be mailed to any host on the Internet.

The assessment also noted that the WAN service provider may rely on microwave stations or satellites as relay points, thereby exposing HGA's information to eavesdropping. Similarly, any information, including passwords and mail messages, transmitted during a dial-in session is subject to eavesdropping.

20.6 Recommendations for Mitigating the Identified Vulnerabilities

The discussions in the following subsections were chosen to illustrate a *broad sampling*[143] of handbook topics. Risk management and security program management themes are integral throughout, with particular emphasis given to the selection of risk-driven safeguards.

20.6.1 Mitigating Payroll Fraud Vulnerabilities

To remove the vulnerabilities related to payroll fraud, the risk assessment team recommended[144] the use of stronger authentication mechanisms based on smart tokens to generate one-time passwords that cannot be used by an interloper for subsequent sessions. Such mechanisms would make it very difficult for outsiders (e.g., from the Internet) who penetrate systems on the WAN to use them to attack the mainframe. The authors noted, however, that the mainframe serves many different agencies, and HGA has no authority over the way the mainframe is configured and operated. Thus, the costs and procedural difficulties of implementing such controls would be substantial. The assessment team also recommended improving the server's administrative procedures and the speed with which security-related bug fixes distributed by the vendor are installed on the server.

After input from COG security specialists and application owners, HGA's managers accepted most of the risk assessment team's recommendations. They decided that since the residual risks from the falsification of time sheets were acceptably low, no changes in procedures were necessary. However, they judged the risks of payroll fraud due to the interceptability of LAN server passwords to be unacceptably high, and thus directed COG to investigate the costs and procedures associated with using one-time passwords for Time and Attendance Clerks and supervisor sessions on the server. Other users performing less sensitive tasks on the LAN would continue to use password-based authentication.

While the immaturity of the LAN server's access controls was judged a significant source of risk, COG was only able to identify one other PC LAN product that would be significantly better in this respect. Unfortunately, this product was considerably less friendly to users and application developers, and incompatible with other applications used by HGA. The negative impact of changing PC LAN products was judged too high for the potential incremental gain in security benefits. Consequently, HGA decided to accept the risks accompanying use of the current product, but directed COG to improve its monitoring of the server's access control configuration

[143] Some of the controls, such as auditing and access controls, play an important role in many areas. The limited nature of this example, however, prevents a broader discussion.

[144] Note that, for the sake of brevity, the process of evaluating the cost-effectiveness of various security controls is not specifically discussed.

and its responsiveness to vendor security reports and bug fixes.

HGA concurred that risks of fraud due to unauthorized modification of time and attendance data at or in transit to the mainframe should not be accepted unless no practical solutions could be identified. After discussions with the mainframe's owning agency, HGA concluded that the owning agency was unlikely to adopt the advanced authentication techniques advocated in the risk assessment. COG, however, proposed an alternative approach that did not require a major resource commitment on the part of the mainframe owner.

The alternative approach would employ digital signatures based on public key cryptographic techniques to detect unauthorized modification of time and attendance data. The data would be *digitally signed* by the supervisor using a private key prior to transmission to the mainframe. When the payroll application program was run on the mainframe, it would use the corresponding public key to validate the correspondence between the time and attendance data and the signature. Any modification of the data during transmission over the WAN or while in temporary storage at the mainframe would result in a mismatch between the signature and the data. If the payroll application detected a mismatch, it would reject the data; HGA personnel would then be notified and asked to review, sign, and send the data again. If the data and signature matched, the payroll application would process the time and attendance data normally.

HGA's decision to use advanced authentication for time and attendance Clerks and Supervisors can be combined with digital signatures by using smart tokens. Smart tokens are programmable devices, so they can be loaded with private keys and instructions for computing digital signatures without burdening the user. When supervisors approve a batch of time and attendance data, the time and attendance application on the server would instruct the supervisor to insert their token in the token reader/writer device attached to the supervisors' PC. The application would then send a special "hash" (summary) of the time and attendance data to the token via the PC. The token would generate a digital signature using its embedded secret key, and then transfer the signature back to the server, again via the PC. The time and attendance application running on the server would append the signature to the data before sending the data to the mainframe and, ultimately, the payroll application.

Although this approach did not address the broader problems posed by the mainframe's I&A vulnerabilities, it does provide a reliable means of detecting time and attendance data tampering. In addition, it protects against bogus time and attendance submissions from systems connected to the WAN because individuals who lack a time and attendance supervisor's smart token will be unable to generate valid signatures. (Note, however, that the use of digital signatures does require increased administration, particularly in the area of key management.) In summary, digital signatures mitigate risks from a number of different kinds of threats.

HGA's management concluded that digitally signing time and attendance data was a practical, cost-effective way of mitigating risks, and directed COG to pursue its implementation. (They also noted that it would be useful as the agency moved to use of digital signatures in other applications.) This is an example of developing and providing a solution in an environment over which no single entity has overall authority.

20.6.2 Mitigating Payroll Error Vulnerabilities

After reviewing the risk assessment, HGA's management concluded that the agency's current safeguards against payroll errors and against accidental corruption and loss of time and attendance data were adequate. However, the managers also concurred with the risk assessment's conclusions about the necessity for establishing incentives for complying (and penalties for not complying) with these safeguards. They thus tasked the Director of Personnel to ensure greater compliance with paperwork-handling procedures and to provide quarterly compliance audit reports. They noted that the digital signature mechanism HGA plans to use for fraud protection can also provide protection against payroll errors due to accidental corruption.

20.6.3 Mitigating Vulnerabilities Related to the Continuity of Operations

The assessment recommended that COG institute a program of periodic internal training and awareness sessions for COG personnel having contingency plan responsibilities.The assessment urged that COG undertake a rehearsal during the next three months in which selected parts of the plan would be exercised. The rehearsal should include attempting to initiate some aspect of processing activities at one of the designated alternative sites. HGA's management agreed that additional contingency plan training was needed for COG personnel and committed itself to its first plan rehearsal within three months.

After a short investigation, HGA divisions owning applications that depend on the WAN concluded that WAN outages, although inconvenient, would not have a major impact on HGA. This is because the few time-sensitive applications that required WAN-based communication with the mainframe were originally designed to work with magnetic tape instead of the WAN, and could still operate in that mode; hence courier-delivered magnetic tapes could be used as an alternative input medium in case of a WAN outage. The divisions responsible for contingency planning for these applications agreed to incorporate into their contingency plans both descriptions of these procedures and other improvements.

With respect to mainframe outages, HGA determined that it could not easily make arrangements for a suitable alternative site. HGA also obtained and examined a copy of the mainframe facility's own contingency plan. After detailed study, including review by an outside consultant, HGA concluded that the plan had major deficiencies and posed significant risks because of HGA's reliance on it for payroll and other services. This was brought to the attention of the Director of HGA, who, in a formal memorandum to the head of the mainframe's owning agency, called for (1) a high-level interagency review of the plan by all agencies that rely on the mainframe, and (2) corrective action to remedy any deficiencies found.

HGA's management agreed to improve adherence to its virus-prevention procedures. It agreed (from the point of view of the entire agency) that information stored on PC hard disks is frequently lost. It estimated, however, that the labor hours lost as a result would amount to less than a person year—which HGA management does *not* consider to be unacceptable. After reviewing options for reducing this risk, HGA concluded that it would be cheaper to accept the associated loss than to commit significant resources in an attempt to avoid it. COG volunteered, however, to set up an automated program on the LAN server that e-mails backup reminders to all PC users once each quarter. In addition, COG agreed to provide regular backup services for about 5 percent of HGA's PCs; these will be chosen by HGA's management based on the information stored on their hard disks.

20.6.4 Mitigating Threats of Information Disclosure/Brokering

HGA concurred with the risk assessment's conclusions about its exposure to information-brokering risks, and adopted most of the associated recommendations.

The assessment recommended that HGA improve its security awareness training (e.g., via mandatory refresher courses) and that it institute some form of compliance audits. The training should be sure to stress the penalties for noncompliance. It also suggested installing "screen lock" software on PCs that automatically lock a PC after a specified period of idle time in which no keystrokes have been entered; unlocking the screen requires that the user enter a password or reboot the system.

The assessment recommended that HGA modify its information-handling policies so that employees would be required to store some kinds of disclosure-sensitive information only on PC local hard disks (or floppies), but not on the server. This would eliminate or reduce risks of LAN eavesdropping. It was also recommended that an activity log be installed on the server (and regularly reviewed). Moreover, it would avoid unnecessary reliance on the server's access-control features, which are of uncertain assurance. The assessment noted, however, that this strategy conflicts with the desire to store most information on the server's disks so that it is backed up routinely by COG personnel. (This could be offset by assigning responsibility for someone other than the PC owner to make backup copies.) Since the security habits of HGA's PC users have generally been poor, the assessment also recommended use of hard-disk encryption utilities to protect disclosure-sensitive information on unattended PCs from browsing by unauthorized individuals. Also, ways to encrypt information on the server's disks would be studied.

The assessment recommended that HGA conduct a thorough review of the mainframe's safeguards in these respects, and that it regularly review the mainframe audit log, using a query package, with particular attention to records that describe user accesses to HGA's employee master database.

20.6.5 Mitigating Network-Related Threats

The assessment recommended that HGA:

- require stronger I&A for dial-in access or, alternatively, that a restricted version of the mail utility be provided for dial-in, which would prevent a user from including files in outgoing mail messages;

- replace its current modem pool with encrypting modems, and provide each dial-in user with such a modem; and

- work with the mainframe agency to install a similar encryption capability for server-to-mainframe communications over the WAN.

As with previous risk assessment recommendations, HGA's management tasked COG to analyze the costs, benefits, and impacts of addressing the vulnerabilities identified in the risk assessment. HGA eventually adopted some of the risk assessment's recommendations, while declining others. In addition, HGA decided that its policy on handling time and attendance information needed to be clarified, strengthened, and elaborated, with the belief that implementing such a policy would help reduce risks of Internet and dial-in eavesdropping. Thus, HGA developed and issued a revised policy, stating that users are individually responsible for ensuring that they do not transmit disclosure-sensitive information outside of HGA's facilities via e-mail or other means. It also prohibited them from examining or transmitting e-mail containing such information during dial-in sessions and developed and promulgated penalties for noncompliance.

20.7 Summary

This chapter has illustrated how many of the concepts described in previous chapters might be applied in a federal agency. An integrated example concerning a Hypothetical Government Agency (HGA) has been discussed and used as the basis for examining a number of these concepts. HGA's distributed system architecture and its uses were described. The time and attendance application was considered in some detail.

For context, some national and agency-level policies were referenced. Detailed operational policies and procedures for computer systems were discussed and related to these high-level policies. HGA assets and threats were identified, and a detailed survey of selected safeguards, vulnerabilities, and risk mitigation actions were presented. The safeguards included a wide variety of procedural and automated techniques, and were used to illustrate issues of assurance, compliance, security program oversight, and inter-agency coordination.

As illustrated, effective computer security requires clear direction from upper management.

Upper management must assign security responsibilities to organizational elements and individuals and must formulate or elaborate the security policies that become the foundation for the organization's security program. These policies must be based on an understanding of the organization's mission priorities and the assets and business operations necessary to fulfill them. They must also be based on a pragmatic assessment of the threats against these assets and operations. A critical element is assessment of threat likelihoods. These are most accurate when derived from historical data, but must also anticipate trends stimulated by emerging technologies.

A good security program relies on an integrated, cost-effective collection of physical, procedural, and automated controls. Cost-effectiveness requires targeting these controls at the threats that pose the highest risks while accepting other residual risks. The difficulty of applying controls properly and in a consistent manner over time has been the downfall of many security programs. This chapter has provided numerous examples in which major security vulnerabilities arose from a lack of assurance or compliance. Hence, periodic compliance audits, examinations of the effectiveness of controls, and reassessments of threats are essential to the success of any organization's security program.

Cross Reference and Index

Interdependencies Cross Reference

The following is a cross reference of the interdependencies sections. Note that the references only include specific controls. Some controls were referenced in groups, such as technical controls and occasionally interdependencies were noted for all controls.

Control	Chapters Where It Is Cited
Policy	Program Management
	Life Cycle
	Personnel/User
	Contingency
	Awareness and Training
	Logical Access
	Audit
Program Management	Policy
	Awareness and Training
Risk Management	Life Cycle
	Contingency
	Incident
Life Cycle	Program Management
	Assurance
Assurance	Life Cycle
	Support and Operations
	Audit
	Cryptography
Personnel	Training and Awareness
	Support and Operations
	Access
Training and Awareness	Personnel/User
	Incident
	Support and Operations

Contingency	Incident Support and Operations Physical and Environmental Audit
Incident	Contingency Support and Operations Audit
Physical and Environment	Contingency Support and Operations Logical Access Cryptography
Support and Operations	Contingency Incident
Identification and Authentication	Personnel/User Physical and Environmental Logical Access Audit Cryptography
Access Controls	Policy Personnel/User Physical and Environmental Identification and Authentication Audit Cryptography
Audit	Identification and Authentication Logical Access Cryptography
Cryptography	Identification and Authentication

General Index

☆U.S. GOVERNMENT PRINTING OFFICE: 1995-404-525/47912

ANNOUNCEMENT OF NEW PUBLICATIONS ON COMPUTER SECURITY

Superintendent of Documents
Government Printing Office
Washington, DC 20402

Dear Sir:

Please add my name to the announcement list of new publications to be issued in the series: National Institute of Standards and Technology Special Publication 800-.

Name ____

Company ____

Address ____

City ________________ State ____________ Zip Code __________

(Notification key N-503)

NIST Technical Publications

Periodical

Journal of Research of the National Institute of Standards and Technology—Reports NIST research and development in those disciplines of the physical and engineering sciences in which the Institute is active. These include physics, chemistry, engineering, mathematics, and computer sciences. Papers cover a broad range of subjects, with major emphasis on measurement methodology and the basic technology underlying standardization. Also included from time to time are survey articles on topics closely related to the Institute's technical and scientific programs. Issued six times a year.

Nonperiodicals

Monographs—Major contributions to the technical literature on various subjects related to the Institute's scientific and technical activities.

Handbooks—Recommended codes of engineering and industrial practice (including safety codes) developed in cooperation with interested industries, professional organizations, and regulatory bodies.

Special Publications—Include proceedings of conferences sponsored by NIST, NIST annual reports, and other special publications appropriate to this grouping such as wall charts, pocket cards, and bibliographies.

National Standard Reference Data Series—Provides quantitative data on the physical and chemical properties of materials, compiled from the world's literature and critically evaluated. Developed under a worldwide program coordinated by NIST under the authority of the National Standard Data Act (Public Law 90-396). NOTE: The Journal of Physical and Chemical Reference Data (JPCRD) is published bimonthly for NIST by the American Chemical Society (ACS) and the American Institute of Physics (AIP). Subscriptions, reprints, and supplements are available from ACS, 1155 Sixteenth St., NW, Washington, DC 20056.

Building Science Series—Disseminates technical information developed at the Institute on building materials, components, systems, and whole structures. The series presents research results, test methods, and performance criteria related to the structural and environmental functions and the durability and safety characteristics of building elements and systems.

Technical Notes—Studies or reports which are complete in themselves but restrictive in their treatment of a subject. Analogous to monographs but not so comprehensive in scope or definitive in treatment of the subject area. Often serve as a vehicle for final reports of work performed at NIST under the sponsorship of other government agencies.

Voluntary Product Standards—Developed under procedures published by the Department of Commerce in Part 10, Title 15, of the Code of Federal Regulations. The standards establish nationally recognized requirements for products, and provide all concerned interests with a basis for common understanding of the characteristics of the products. NIST administers this program in support of the efforts of private-sector standardizing organizations.

Order the **following** *NIST publications—FIPS and NISTIRs—from the National Technical Information Service, Springfield, VA 22161.*

Federal Information Processing Standards Publications (FIPS PUB)—Publications in this series collectively constitute the Federal Information Processing Standards Register. The Register serves as the official source of information in the Federal Government regarding standards issued by NIST pursuant to the Federal Property and Administrative Services Act of 1949 as amended, Public Law 89-306 (79 Stat. 1127), and as implemented by Executive Order 11717 (38 FR 12315, dated May 11, 1973) and Part 6 of Title 15 CFR (Code of Federal Regulations).

NIST Interagency Reports (NISTIR)—A special series of interim or final reports on work performed by NIST for outside sponsors (both government and nongovernment). In general, initial distribution is handled by the sponsor; public distribution is by the National Technical Information Service, Springfield, VA 22161, in paper copy or microfiche form.

U.S. Department of Commerce
National Institute of Standards
and Technology
Gaithersburg, MD 20899–0001

Official Business
Penalty for Private Use $300