PRIVACY AND SECURITY IN COMPUTER SYSTEMS

R. Turn and W. H. Ware

January 1975

P-5361

The Rand Paper Series

Papers are issued by The Rand Corporation as a service to its professional staff. Their purpose is to facilitate the exchange of ideas among those who share the author's research interests; Papers are not reports prepared in fulfillment of Rand's contracts or grants. Views expressed in a Paper are the author's own, and are not necessarily shared by Rand or its research sponsors.

The Rand Corporation
Santa Monica, California 90406

PRIVACY AND SECURITY IN COMPUTER SYSTEMS*

R. Turn and W. H. Ware

The Rand Corporation
Santa Monica, California

Computers and their applications in the 1970s differ dramatically from those visualized in the early 1950s when the computer age had its beginnings. Instead of remaining complex and esoteric computational aids to mathematicians and scientists, modern computers have found their most important function in general information processing--storing and manipulating strings of text. Their users need not be highly trained mathematicians but can be office workers, clerks, students, and, of course, researchers in all fields of science and engineering.

Modern computer systems serve many users simultaneously and permit on-line programming, job execution, and data file manipulation from remotely located terminals. The computer's capacity for time-sharing and multiprogramming gives each user the impression that the entire system is devoted to his own exclusive use. Data files and programs may be shared and users can interact with their programs as they are being processed. This mode of operation is controlled by the operating system software--a set of program modules that control the flow of users' programs and service requests, allocate system resources, schedule execution, handle errors, and keep users and their programs from interfering with each other [1]. An important function of the operating system is to protect users' programs and data files against each other and, indeed, against themselves.

It is necessary to isolate users' processes (programs in execution) and to protect their working memory space and permanent data files in order to prevent them from being destroyed or modified by inadvertent programming or operating errors or by deliberate actions of malicious users. In addition, access to any computer system must be controlled

*This paper was prepared for publication in the AMERICAN SCIENTIST.

to assure that proper charges are made for the use of the system resources so that no one receives free services at the expense of someone else. However, there are other reasons for providing protection to the computer and its data files.

In business and industry computers are employed to automate a variety of accounting and record-keeping applications. The information involved, detailing production, marketing, finances, and new product development and research, could be extremely valuable to competitors. Industrial espionage, or gathering "marketing intelligence" as it is sometimes called, has become a large-scale activity in the United States [2].

Computerization of daily business operations has also provided new opportunities for white-collar crime-embezzlement, falsification of records, and larceny by employees. Numerous case histories show [3] that employees who design the systems, write the programs, and operate the data processing equipment have many opportunities for such acts. Some abuses that the computer makes especially easy are payments for fictitious purchases or to fictitious employees, manipulation of credit levels, and deposits of nonexistent payments into various accounts. Business firms, too, may use computers to embezzle their customers or stockholders. For example, in the Equity Funding Corporation case [4] the company greatly inflated its financial report and, thus, the attractiveness of its stock, by listing fictitious assets and using its computerized accounting system to mislead the auditors. The loss to stockholders and financiers was many millions of dollars.

In government, business and industry, and educational institutions computerized personal information record-keeping systems are maintained for administrative, investigative, statistical, or research purposes [5,6]. Information in administrative databank systems is used to make routine decisions about individual data subjects (e.g., to grant or deny benefits, credit, employment, admission to a university), to

establish their connections with activities under investigation (e.g. organized crime), or to correlate and aggregate their characteristics or behavior patterns with those of other individuals to obtain statistical summaries, behavior profiles, and correlations. In these systems privacy and other individual rights of the data subjects may be violated by unwarranted collection, use, and dissemination of personal information.

Furthermore, consolidation of record-keeping into computerized systems sets up highly centralized, easily identifiable targets for disruption and sabotage by disgruntled employees or by those disagreeing violently with the policies or activities of the computer system owner or users. The acts themselves may range from firebombing of computer centers to "boobytrapping" of programs to destroy themselves in case the programmer is dismissed. Table 1 summarizes the history of computer abuse [3]. It is likely that there are many other cases that have not been discovered or were not reported.

Table 1
STATISTICS ON COMPUTER ABUSE*

Year	Financial Fraud	Information or property theft	Unauthorized Use	Vandalism	Total
1969	3	6	0	3	12
1970	7	5	9	8	29
1971	22	18	6	6	52
1972	12	15	16	12	55
1973	21	15	8	9	53

*Cases reported and verified.

Although manual record-keeping systems and data files are subject to similar threats, certain characteristics of information storage and processing in computer systems make threats to these systems more serious.

First, information is stored in the form of magnetization or voltage-level patterns that are not directly readable by users. These can be changed without a trace of evidence. Computerized records do not have signatures or seals to verify authenticity or to distinguish copies from originals, and they can be manipulated electronically from terminals remote from the physical storage of the data. Transactions can be performed automatically at high speed without human monitoring or intervention. Finally, processing rules are expressed in programs stored in the same devices and in the same manner as data and which can thus be changed easily and without trace. Such programs are complex and difficult to validate. On the other hand, a properly designed and implemented computerized information system can control errors and manage access to the records much more effectively than any manual record-keeping system.

In this paper we will examine the protection of privacy and other individual rights in personal information databank systems, maintenance of information confidentiality in statistical and research data bases, and implementation of data security techniques against malicious users and external penetrators.

PRIVACY PROTECTION

Let us turn first to the issue of privacy, which in the context of this discussion, refers to the rights of the individual regarding the collection, processing, storage, dissemination, and use of information about his personal attributes and activities. In one proposal, these rights are embodied in a Code of Fair Information Practices, conceived by the Special Advisory Committee on Automated Personal Data Systems to the Secretary of the Department of Health, Education and Welfare [7]. The Code rests on the following basic principles, which are equally applicable to personal information databank systems in the government and in the private sector:

- There must be no personal data record-keeping systems whose very existence is secret.

- There must be a way for an individual to find out what information about him is on record and how it is used.
- There must be a way for an individual to correct or amend a record of identifiable information about him.
- There must be a way for an individual to prevent information about him that was obtained for one purpose from being used or made available for other purposes without his consent.
- Any organization creating, maintaining, using, or disseminating records of identifiable personal data must guarantee the reliability of the data for their intended use and must take precautions to prevent misuse of the data.

All personal information databank systems would be subject to these requirements and must incorporate a corresponding set of safeguards. Enactment of the Code as a federal or state law specifying penalities for noncompliance and establishing an enforcement machinery would be an important step in securing individual rights. Indeed, the United States Congress recently enacted the first such law, the Privacy Act of 1974, signed by the President on January 1, 1975 [8]. However, this act applies only to databanks maintained by the Federal government. Similar bills are pending in many state legislatures.

Unfair information practices would be subject to criminal and civil sanctions, and victims could recover both punitive and actual damages and obtain injunctive relief. For example, the Privacy Act of 1974 specifies a fine not to exceed $5000 for any willful violation of the act by officers or employees of a Federal record-keeping organization, or by anyone requesting information from the organization under false pretenses.

The Code would require an annual public notice listing the names of record-keeping systems, their nature and purpose, their data sources, the categories of data maintained, the identities of agencies that routinely use the records, and each agency's policies regarding the storage, retrievability, access controls, retention, and disposal of the records.

The Code would regulate the behavior of organizations maintaining personal information record-keeping systems by requiring them to identify

an arbitrator to whom complaints could be directed; to take affirmative action to inform employees of safeguards and to specify penalities for any infraction of them; to take precautions against transferring identifiable personal information to other record-keeping systems that may not include adequate safeguards; and to maintain records with an accuracy, completeness, currency, and pertinence consistent with their intended use.

Individual rights would be explicitly guaranteed by the Code. When asked to supply personal data, an individual would be informed whether he is legally required to do so or whether he may refuse. Upon his request, he would be informed whether he is a subject in a given data system, and he would have the opportunity to inspect his record, challenge it, and cause corrections or amendments to be made. In the case of a dispute that cannot be resolved, he would have the right to submit a concise rebuttal, and the data items in question would be marked as being disputed. He would also receive assurance that data about himself would be used only for the stated purposes of the system, for the agency would be required to request permission for and to obtain records of any extraordinary uses. Nearly all proposed legislation and the enacted Privacy Act of 1974 have adopted the Code as the basic framework of privacy protection. Based on the use of existing legal and judicial institutions to implement privacy protection, the Code is consistent with the traditional approach in the United States to deterring societally or personally undesirable actions against the individual (e.g., criminal laws, and unfair labor practices laws). It would create a minimum of new bureaucratic functions. Through court decisions and interpretations, it would provide an adaptive solution to the issue of personal privacy as the attitudes of society and the needs for personal information change. However, like other social policies developed through legislative and judicial processes, the reforms imposed by the Code would proceed in a deliberate fashion which often seems too slow to advocates of privacy protection.

CONFIDENTIALITY

In contrast to privacy, which refers to the rights of the individual, confidentiality implies that the data themselves and the information they contain must be protected, and that their use must be confined to authorized purposes by authorized people.

Certain categories of personal information are given a confidential status by statutes and laws. For example, the personal data gathered in the United States decennial census are required to be kept confidential by Federal law [9]. This means that no individually identified census responses may be disseminated to anyone outside the Census Bureau, and even within the Bureau only specially authorized employees are permitted access. Attorney-client information exchanges, certain medical and mental health information, and legal proceedings involving children and juveniles are other examples of information categories that are protected from general access by confidentiality provisions in Federal or state statutes [10].

Most categories of personal information do not enjoy any statutory protection, however; disclosure of such information may be compelled by legal process, such as subpoena issued by a court, legislative committee, or other official body that has jurisdiction in the locality where the data are kept. Personal information gathered by educational institutions and by research projects in social, political and behavioral sciences are very susceptible to these procedures.

The lack of statutory confidentiality of personal information gathered for research purposes is a serious concern to researchers whose studies require the gathering of sensitive personal information. While the researcher may have the best of intentions as far as preventing any dissemination of identified information (and may assure his respondents of its confidentiality) if faced with a subpoena he has the choice of either being in contempt and suffering the penalties, or else surrendering the data [11]. In either case his research project has been seriously damaged.

The reality of the subpoena threat against research data bases was demonstrated by two incidents in 1969. One involved an Office of Economic Opportunity-sponsored negative income tax experiment in New Jersey in which first the county prosecutor and then a grand jury subpoened the records [12]. During the same period the General Accounting Office and the Senate Finance Committee also sought access to the records. In the second case, an investigating commission demanded access to the data on an anti-poverty research project in Chicago [13]. The project collapsed.

The Code of Fair Information Practices addresses this problem by seeking Federal legislation to protect statistical reporting or research data against compulsory disclosure through legal process. Such legislation would include the following features:

- Protection should be limited to data identifiable with or traceable to specific individuals.
- Protection should be specific enough to qualify for non-disclosure exemption under the Freedom of Information Act [14].
- Protection should be available for data in the custody of all statistical reporting and research systems whether supported by federal funds or not.
- Federal law should be controlling; no state statute should interfere with the protection provided.
- Either the custodian or the individual about whom data are sought by legal process should be able to invoke the protection, but only the individual should be able to waive it.

Whether or not general statutory confidentiality protection is provided for statistical reporting or research data, the Code specifies that the data gathering organization is responsible for:

- Informing the individual subject whether he is legally required to supply the data requested or may refuse, and also of any specific consequences for him, which are known to the organization, of providing or not providing such data;
- Guarantee that no use of individually identifiable data will be made that is not within the stated purposes of the system

as reasonably understood by the individual, unless the informed consent of the individual has been explicitly obtained; and

- Guarantee that no data about an individual will be made available from the system in response to a compulsory legal process, unless the individual to whom the data pertains has been notified of the demand and has been afforded full access to the data before they are made available in response to the demand.

Until this blanket protection is achieved, there are several procedural and technical means available to reduce the likelihood of damage to research subjects in case the data files are subpoenaed or otherwise obtained by unauthorized persons.

One procedural means of protecting the confidentiality of the data is exposure reduction which involves limiting the amount and nature of the data collected. At the risk of reducing the utility of his studies, the researcher refrains from obtaining sensitive information or uses survey techniques that introduce uncertainty in the responses. In the case of "randomized response" approach [15], sensitive questions are answered only in a statistical manner. For example, a respondent may be given two questions, one sensitive and the other innocuous (e.g., Do you take drugs? Do you like opera?). He then selects one of the questions randomly and answers it truthfully but does not indicate which question he selected. Given that statistics regarding the innocuous question are available, the researcher can estimate the proportion of respondents affirming or denying the sensitive question.

Reduction of sensitivity of information and, hence, the possibility that it may be subpoenaed can also be accomplished by the randomized response approach, and by "inoculating" identifiable information with errors in an irreversible but controlled fashion, so that the statistics of the data ensemble remain unchanged but individual responses may be incorrect [16]. For example, the "yes-no" answers to a sensitive question may be changed at random. Of course it is essential to publicize widely that the data files have been contaminated in this way.

The anonymity of information in a statistical databank may also be ensured by removing or modifying a number of identifying data items sufficient to preclude the use of remaining data to identify a specific individual, even if the so-called "twenty questions game" is played against the file. For databanks where identification must be preserved for future updating of the information, a link-file system approach [17] may be appropriate. In such a scheme, the identifying data for each individual are separated from the rest of the record and assigned a random code number. The substantive data are assigned a different code number. A third data file that establishes the correspondence between the first two is kept at a location outside the jurisdiction of the authorities of the locality in which the data are kept, for example, in a foreign country.

Encoding and encryption techniques can be applied to conceal the stored information [18] or to perform data merging operations involving two databanks without revealing the information [19]. It is essential, of course, to assure that the keys to encoding and encryption operations are adequately safeguarded. Accountability procedures are also necessary to ensure that a specific data bank employee or user is responsible at all times for every sensitive data file or set of records. Finally, access control procedures must be enforced to prevent unauthorized access or dissemination of any sensitive data. These techniques are discussed in the following section.

COMPUTER SECURITY

Computer security encompasses the measures required to (1) protect a computer-based system, including its physical hardware, personnel, and data against deliberate or accidental damage; (2) protect the system against denial of use by its rightful owners; and (3) protect information or data against divulgence to unauthorized recipients.

Threats that must be averted by computer security measures include natural disasters, riots, equipment failures, negligent or maliciously motivated employees and users, and external intruders.

The physical security measures against these threats are well in hand [20], but their application in a computer facility requires careful analysis and engineering. For example, a ceiling sprinkler system is not a proper fire extinguishing system in a room housing a computer, and a tear gas dispensing system that deters a rioting mob also corrodes sensitive electronic components in the computer circuitry.

Different security measures are required to limit access to programs and data in the computer system to legitimate authorized users. These access control measures must be able to counter actions that covertly capitalize on weaknesses in the computer system that may be unknown even to the system's management. It may be very difficult to determine that such actions are in process or have been completed successfully. For example, if a penetrator succeeds in gaining control of the operating system, he could first disable the accounting and auditing programs and then proceed to read or modify any program or data file without being detected.

Unauthorized access may occur accidentally due to programming errors or malfunctioning equipment, or as a result of deliberately planned activity. In the latter case, the ability of an intruder to gain access to protected resources depends on the structure of the computer system [21,22] and on the opportunities it provides for interaction [23]. For example, a remotely-accessible, time-shared system where users can submit assembly language programs offers more opportunities for penetration than a system where the users are limited to performing a fixed set of transactions and cannot submit their own programs.

Data security techniques are implemented to prevent unauthorized access or, if absolute prevention is impossible or is unneeded, to increase the costs of intrusion to a level where the expected "profits" are unattractive. The objectives are [24,25]:

- Isolation of users and their processes from each other and from supervisory programs, to prevent users' processes from interfering with each other or the supervisor and from capturing control of the system.

- Positive identification of all users and authentication of their identity and attachment of unforgeable identifiers to all users' processes in the system (the time of creation of the process has been suggested as one such identifier).

- Total access control by the supervisory program over all shared system and user-owned resources (memory space, subroutines, data files, input-output devices, etc.).

- Concealment of information on removable storage media and in communication channels by privacy transformation (encryption) techniques.

- Integrity control, ensuring that the protective system is correctly implemented and is not weakened by modification of software or hardware.

Defensive design and application of the principle of least privilege are basic to any data security system. Examples of defensive design include the concentration of software-implemented security functions into security kernels--compact software modules whose correct operation can be proved by formal techniques or by testing, and compartmentation of the system to limit the damage an intruder can do if he does succeed in penetrating some part of the protected system. The principle of least privilege specifies that any user's or system's process be granted only those access rights and privileges it needs to perform its functions. Neither defensive design nor the principle of least privilege has been applied in the design of contemporary operating systems.

The simplest way to isolate a set of users is to process their programs one at a time completely erasing any portion of the memory space that is available to the subsequent user. This approach is still practiced in processing classified government data, but it is unnatural and wasteful in modern resource-sharing systems. In new systems, a common isolation technique is to determine for each user the bounds of the assigned memory space (the lowest and the highest address values he may use) and testing each memory reference to be sure that it falls within these bounds. In computers using the virtual memory concept, memory space is further protected by the user's inability to

generate addresses that are outside their own assigned memory space.

The operating system and supervisory programs can be isolated from users by providing two or more system states for the processor (a "user state" and one or more "supervisor states") and a set of privileged instructions. These instructions are used by the operating system for allocating resources, establishing access control privileges, and requesting input-output operations, and they can be executed only when the system is in the appropriate supervisory state. As an illustration, assume that a user process needs to read a set of records from a particular data file. It issues a request specifying the file name and the records involved. The operating system will change the system state to "supervisory" and will test whether the process is authorized to have access to the file and the records involved. If authorization is permitted, it will transfer the requested data into the user's memory space and return the system "user" state.

IDENTIFICATION

In a remotely accessible computer system, it is necessary for the system to identify positively each user and each terminal. In a multicomputer network, participating systems must be identified to each other and to the users. The need for precautions is demonstrated by the so-called "piggy-back" system penetration threat, in which an illicit minicomputer-equipped terminal is inserted into a communication line to "manage" a user's interactions with the computer system [26,27]. The user's sign-on procedure is intercepted by the intruder who generates the correct responses until the user transmits his password at which point he is informed of a system failure and disconnected. The intruder then signs on himself using the intercepted password.

The most common technique for identifying a user to the system employs his name, man-number, or account number. The authenticity of the identification is verified by a password.

Three approaches to authentication of identity [28] are by means of something the user knows, something the user is, or something the

user has. Popular in the last category are badges or cards bearing a magnetic stripe, which can be inserted into a terminal for identification. These cards can be designed to resist forgers and, although they can be given to others, their possession can be made mandatory and is easy to check. For example, the cards may be assigned additional functions, such as operating a card-key lock to gain entry to the terminal room or they may be required for presentation when submitting computing jobs.

The use of personal characteristics, such as fingerprints, voice prints, or hand dimensions is attractive but involves the use of complex devices for extracting the physical variables and formulating them for transmission. Moreover, considerable processing time and storage space may be required. At present these techniques are considered too costly for general application, but future advances in hardware technology may permit manufacture of inexpensive special-purpose processors for fingerprint analysis or voice print generation. Thus these identification/authentication methods may become economically attractive.

ACCESS CONTROL

A major advantage of many modern computer systems is the ability of users to share programs and data among themselves. However, in order to control this function, the owners of the shared resources must be able to specify to the system who is to have access to data and what processing actions they may taken. In return, the system must be able to enforce the rules rigidly not only under static, predetermined conditions, but also under dynamic conditions when authorization changes are frequent. In a dynamic situation an authorized user may generate new processes and data files and wish to pass to others selected access rights, to retract previously granted rights, or to specify the rights-passing conditions within the new processes themselves. Clearly, management of access rights is a complicated task that must be implemented in the system's operating software.

Several conceptual models have been developed for the implementation of access control procedures in operating systems. The basic elements are subjects seeking to gain access (processes), protected objects (data files, other processes), and access modes (the operation that processes may perform on objects, such as "read" or "modify" data, "execute," a program, etc.). The access control matrix specifies the access rights and modes. Table 2 illustrates a hypothetical example.

Since an access control matrix is likely to be sparse, it will be uneconomical to store in matrix form. Instead, each object can be provided a list of processes that are to be afforded access to it (i.e., the columns of the access control matrix will be associated with the objects) or, alternatively, each process can have information that will allow it access to those objects it is authorized to obtain (i.e., the rows of the access control matrix will be associated with subjects).

Table 2 An Illustrative Access Control Matrix

	Objects						
	Processes			Files		Devices	
Subjects	P_1	P_2	P_3	F_1	F_2	D_1	D_2
S_1	control*	owner execute	owner control	read write		seek	owner
S_2		control	stop	owner	update	owner	seek
S_3			control	delete	read		

*Entries in the matrix show the access mode that subjects are afforded.

In the latter case, a process will have a set of "keys" to open "locks" on protected objects or, viewed alternatively, it has a set of access-granting "tokens" for presentation to the access control mechanism. Each of these approaches has been implemented in an experimental design [29-32].

The situation is more complex if access granting depends on the data that are being requested. For example, a user may be allowed to process salary information only for employees who earn less than $20,000, who are members of a particular department or project, or who satisfy some other specified criteria. Thus, each access-granting decision may require a computational-logical procedure of considerable complexity. Further, if proprietary programs are to be used, the owner must be assured that the user does not make a copy for himself, and the user must be assured that the proprietary program does not keep a copy of his data. A mutually suspicious situation such as this may occur with a program for preparing income tax returns. A more detailed discussion of this and other complex access control models is contained in the literature [30-32].

In computer-terminal and computer-computer communications links, and in removable storage devices, data can be protected against wire-tapping and theft by concealment techniques including the use of cryptographic transformations. While the basic principles of cryptography [33-34] formulated for the concealment of natural language still apply, there are certain qualitative and quantitative differences in the application of cryptographic techniques to protect information in computers, and in the use of cryptanalytic techniques by an intruder to gain access to information so protected. Computer-stored data are unlike written or telegraphed messages in ways that may both enhance and diminish the protection provided by encryption. For example:

- Most of the stored data are numerical values, codes, names and addresses, or statements in programming language. These tend to have a more uniform character and polygram frequency distributions than natural languages.

- Data and expressions in computers tend to have rigid formats and to follow strict syntactic rules.

- Large amounts of data are stored, and sizable fragments of material known to occur also in the encrypted files can be expected to be available to the cryptanalyst for use in formulating and testing his hypotheses.

Given these differences and the availability of computers for cryptanalytic purposes, standard cryptographic transformations (both polyalphabetic substitutions and simple transpositions) can be solved very quickly with the help of powerful mathematical techniques [35]. On the other hand, the rapidly decreasing cost of digital hardware will make it economically feasible to construct special-purpose processors for applications of combined substitution and transposition transformations that approximate the "mixing transformations" recommended by Shannon [34]. For example, the IBM "Lucifer" system [36] could be built with only 4 monolithic microcircuit chips (each containing 280 logic gates) to apply a complex sequence of transformations.

INTEGRITY CONTROL

A comprehensive system of security safeguards is effective only if it is correctly designed and implemented and operates correctly thereafter. The major problem in resource-sharing systems is the design of the operating system software. Large operating systems contain hundreds of program modules and hundreds of thousands of instructions. For example, the MULTICS operating system [37] which was designed and implemented with data security and access control in mind, contains over 2000 program modules. Some 400 of these implement functions are involved in or critical to the system's security; on the average, each module contains 200 lines of program statements. Despite more than nine years of operational use and continuous testing, occasional errors in MULTICS are still found. Not every error in the operating system software is also a security vulnerability, but many are. Indeed, every operating system now in use that has been tested has been found to contain numerous errors.

Software errors are a general concern, and techniques are being developed to produce more reliable software. But the need for security adds a new dimension to the pursuit or error-free operation: not only should programs perform correctly all tasks they are designed to do, but they should also not be able to perform any other tasks.

Anomalous behavior is very difficult to validate and may require formal proofs of program correctness in addition to vulnerability analyses and penetration testing. All current validation approaches have shortcomings: only very small programs can be handled by mathematical-logical program proving methods [38], and penetration testing shows only whether or not a particular test team can defeat the security system. To top it off, all approaches are quite expensive. Vulnerability analysis of a modest-sized operating system may require 3-6 months of work by a team of 3-4 analysts. The average cost per design error discovered ranges from $100 to $1000 [39].

Among the techniques for assuring that security mechanisms continue to function properly are auditing and threat monitoring. Auditing involves continuous recording of information about access control requests and corresponding system decisions for analysis after the fact by systems' security personnel or external auditors. Threat monitoring programs assemble information on the operation of the security system in real time to detect unusual activities. In present systems threat monitoring is at a very primitive level. Essentially, it counts the number of times a user fails to provide the correct password. More sophisticated threat monitoring instrumentation presupposes the ability to characterize penetration activities in terms of sets of measurable system variables, plus the ability to distinguish penetration attempts from other unusual but legitimate data processing activities.

PROTECTION COSTS

Protection costs include the initial cost of establishing a protection system and recurring operational costs. Typical initial cost items are:

- Analysis and specification of protection requirements.
- Design and implementation of policies, regulations and procedures for providing data security and privacy and confidentiality safeguards.

- Acquisition of protection-oriented equipment and facilities and provisions for physical security.
- Generation, validation, and testing of the system software.
- Conversion of personal information data files to incorporate protection features.

The initial investment in design will greatly influence the quality of protection achieved. In particular, expenditures for software design, implementation and validation are the key to the system's effectiveness against possible penetration attempts by malicious users. Experience shows that currently these types of intrusions are more likely than penetrations from outside [3].

The operational cost of protection includes the usual overhead, such as salaries, equipment rental, and expendable supplies. It may also include the following:

- Processing time for user identification and authentication, application of access control procedures, and recording transaction logs and audit trails.
- Main- and secondary-storage requirements for the protection-oriented programs, tables and data fields, and for transaction logs and audit trails.
- Computer and personnel time for testing and revalidation of system's hardware and software after modifications, repairs, or restarts.
- Personnel training and education in protection oriented policies, procedures, and attitudes.

Further, if processing time and storage requirements for the safeguards are substantial, the computer system may be unable to meet its peak workload demand; the organization may be compelled to reduce service or to acquire more or larger processors or more on-line storage. In general, security mechanisms tend to reduce the general availability of a system to its users, and thus they are in conflict with the traditional goals of systems managers and users--economy, increased availability, and easier access.

The frequency and complexity of the access control decisions that must be made are important variables in the cost of protection. If the access rights of each user are tested only at the initial log-in time, or at the time of the initial file opening request, the processing requirements may be small relative to the normal file processing operations. However, cost will escalate when access-control tests are applied each time a record is retrieved from protected data files. Furthermore, if the access control decisions are data-dependent, the cost of processing time will be even greater, for every data field must be tested to determine its value and then compared with the parameters specified for the access test. In general, it is estimated that access control features tend to increase the overall processing time by 5 to 10 percent, the operating system software size by 10 percent, and the main memory requirement for the operating system by 10 to 20 percent [40].

CONCLUDING REMARKS

This paper has presented a broad overview of the three topics important in the design, operation, and use of computerized information systems and personal information databanks: safeguarding of individual rights of data subjects, providing confidentiality to identifiable personal information in statistical and research-oriented information systems, and protecting computer resources and data files against unauthorized use by maliciously-inclined insiders or by intruders from outside.

In the protection of individual rights of data subjects an important advance was made by enacting the Privacy Act of 1974. Still needed is privacy legislation to cover the sphere. However, no law can be expected to be perfect, and we must depend on the American political and judicial system to continue evolutionary improvement through amendments and legal interpretations. Extension of the Code to cover personal information databanks in non-governmental establishments is clearly the next move.

The need to provide statutory confidentiality protection to personal information on identifiable individuals in regards to data-banks is also recognized in the Code of Fair Information Practices and recommendations to this end have been made in some of the pending bills. Until these recommendations can be enacted, technical and procedural means must be used to reduce the dangers of accidental, malicious, or legally forced disclosure of such information even though this may reduce the research value of the information.

Techniques for data security are evolving rapidly but much research and development remains. Implementation in an existing system is often excessively costly or even infeasible. Moreover, not all computer systems will require the same level of protection. For example, those containing personal information that is already publicly available need implement only those features that protect data from accidental modification and prevent users from interfering with each other. More sensitive information in on-line, shared, or integrated databanks systems may require all the known protective features and more. In fact, extremely sensitive information should not be stored in any contemporary resource-sharing computerized databank system.

REFERENCES

1. Denning, P. J., "Third Generation Computer Systems," ACM Computing Surveys, Vol. 3, No. 4, October 1971, pp. 175-216.

2. Greene, R. M., Jr., Business Intelligence and Espionage, Dow Jones-Irving, Inc., 1966.

3. Parker, D. B., S. Nycum, and S. S. Oura, Computer Abuse, Stanford Research Institute, Menlo Park, Calif., 1973.

4. McLaughlin, R. A., "Equity Funding: Everyone Is Pointing at the Computer," Datamation, June 1973, pp. 88-91.

5. Westin, A. F., and M. A. Baker, Databanks in a Free Society, Quadrangle Books, New York, N. Y., 1972.

6. Wheeler, S. (ed.), On Record: Files and Dossiers in American Life, Russell Sage Foundation, New York, N. Y., 1969.

7. "Records, Computers, and the Rights of Citizens" a report of the Secretary's Advisory Committee on Automated Personal Data Systems. U. S. Department of Health, Education, and Welfare, July, 1973. DHEW Publication No. (OS) 73-94. See "Summary and Recommendations," p.xxiii.

8. Privacy Act of 1974, Title 5, U. S. Code, Section 552a.

9. Title 13, U. S. Code.

10. "The Computerization of Government Files: What Impact on the Individual?" UCLA Law Review, 15.5, Sept. 1968, pp. 1374-1498.

11. Nejelski, P., and L. M. Lerman, "A Research-Subject Testimonial Privilege: What To Do Before the Subpoena Arrives," Wisconsin Law Review, No. 4, Fall 1971, pp. 1085-1148.

12. Kershaw, D. N., and J. C. Small, "Data Confidentiality and Privacy: Lessons from the New Jersey Negative Income Tax Experiment," Public Policy, Vol. 20, No. 2, Spring 1972, pp. 257-280.

13. Walsh, J., "Anti-Poverty R&D: Chicago Debacle Suggests Pitfalls Facing OEO," Science, Vol. 165, September 19, 1969, pp. 1243-1245.

14. Freedom of Information Act, Title 5, U. S. Code, Section 552.

15. Boruch, R. F., "Assuring Confidentiality of Responses in Social Research: A Note on Strategies," The American Sociologist, November 1971, pp. 308-311.

16. Boruch, R. F., "Maintaining Confidentiality of Data in Educational Research: A Systemic Analysis," American Psychologist, Vol. 26, No. 5, May 1971, pp. 413-430.

17. Astin, A. W., and R. F. Boruch, A Link System for Assuring Confidentiality of Research Data in Longitudinal Studies, ACE Research Reports, Vol. 5, No. 3, American Council on Education, Washington, D. C., 1970.

18. Turn, R., "Privacy Transformations for Databank Systems," AFIPS Conference Proceedings, Volume 42, 1973 National Computer Conference AFIPS Press, Montvale, N. J., 1973, pp. 589-601.

19. Boruch, R. F., "Strategies for Eliciting and Merging Confidential Social Research Data," Policy Sciences, Vol. 3, 1972, pp. 275-297.

20. Guidelines for Automatic Data Processing: Physical Security and Risk Management, U.S. Department of Commerce, National Bureau of Standards, FIPS Pub. 31, Washington, D. C., June 1974.

21. Government Looks at Privacy and Security in Computer Systems, Technical Note No. 809, U. S. Department of Commerce, National Bureau of Standards, Washington, D. C., November 1973.

22. Turn, R., Privacy and Security in Personal Information Databank Systems, R-1044-NSF, The Rand Corporation, Santa Monica, Calif., March 1974.

23. Anderson, J. P., Computer Security Planning Study, Vol. 2, ESD-TR-72-51, U. S. Air Force Systems Command, Electronic Systems Division, L. G. Hanscom Field, Bedford, Mass., October 1972.

24. Controlled Accessibility Workshop Report, Technical Note 827, U. S. Department of Commerce, National Bureau of Standards, Washington, D. C., May 1974.

25. Data Security and Data Processing, Volume 5, Study Results: TRW Systems, Inc., Publication G320-1375, International Business Machines Corporation, White Plains, N. Y., June 1974.

26. Petersen, H. E., and R. Turn, "Systems Implications of Information Privacy," AFIPS Conference Proceedings, Vol. 30, 1967, Spring Joint Computer Conference, Thompson Book Co., Washington, D. C., 1967, pp. 291-300.

27. Carroll, J. M., and P. Reeves, "Security of Data Communications: A Realization of Piggyback Infiltration," Infor, October 1973, pp. 226-231.

28. Approaches to Privacy and Security in Computer Systems, Special Publication 404, U. S. Department of Commerce, National Bureau of Standards, Washington, D. C., September 1974, pp. 26-31.

29. Graham, G. S., and P. J. Denning, "Protection -- Principles and Practice," AFIPS Conference Proceedings, Vol. 40, 1972 Spring Joint Computer Conference, AFIPS Press, Montvale, N. J., 1972, pp. 417-429.

30. Schroeder, M. D., Cooperation of Mutually Suspicious Subsystems in a Computer Utility, Report MAC-TR-104, Project MAC, Massachusetts Institute of Technology, Cambridge, Mass., September 1972.

31. Jones, A. K., Protection in Programmed Systems, Department of Computer Science, Carnegie-Mellon University, Pittsburgh, Pa., June 1973.

32. Hoffman, L. J. (ed.), Security and Privacy in Computer Systems, Melville Publishing Co., Los Angeles, Calif., 1973.

33. Gaines, H. F., Cryptanalysis: A Study of Ciphers and Their Solution, Dover Publications, Inc., New York, N. Y., 1956.

34. Shannon, C. E., "Communications Theory of Secrecy Systems," Bell System Technical Journal, 1949, pp. 656-715.

35. Tuckerman, B., A Study of the Vigenere-Vernam Single and Multiple Loop Enciphering Systems, Report RC-2879, IBM Research Laboratory, Yorktown Heights, N. Y., May 14, 1970.

36. Feistel, H., "Cryptography and Computer Privacy," Scientific American, Vol. 228, No. 5, May 1973, pp. 15-23.

37. Saltzer, J. H., "Protection and the Control of Information Sharing in Multics," Communications of the ACM, Vol. 17, No. 7, July 1974, pp. 388-402.

38. Linden, T. A., "A Summary of Progress Toward Proving Program Correctness," AFIPS Conference Proceedings, Vol. 41, 1972 Fall Joint Computer Conference, AFIPS Press, Montvale, N. J., 1972, pp. 201-211.

39. Weissman, C., System Security Analysis/Certification Methodology and Results, SP-3728, System Development Corporation, Santa Monica, California, 8 October, 1973.

40. Chastain, D. R., "Security vs. Performance," Datamation, November 1973, pp. 110-111, 116.