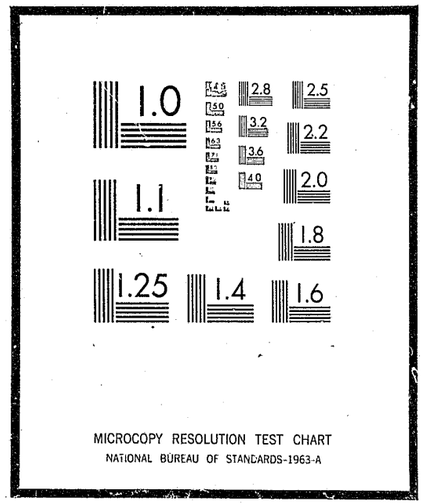

PRIVACY AND SECURITY IN PERSONAL INFORMATION DATABANK SYSTEMS

PREPARED FOR THE NATIONAL SCIENCE FOUNDATION

REIN TURN

R-1044-NSF

MARCH 1974

Rand
SANTA MONICA, CA. 90406

The research described herein was supported by the National Science Foundation under Grant GI-29943. Reports of The Rand Corporation do not necessarily reflect the opinions or policies of the sponsors of Rand research.

Published by The Rand Corporation

PRIVACY AND SECURITY IN PERSONAL INFORMATION DATABANK SYSTEMS

PREPARED FOR THE NATIONAL SCIENCE FOUNDATION

REIN TURN

R-1044-NSF
MARCH 1974

Rand
SANTA MONICA, CA. 90406

C/28/75

PREFACE

This report is one of a set of reports and papers that present the results of an exploratory research project, "Protection of Privacy of Personal Information in Databanks: Theoretical and Technical Aspects," supported by National Science Foundation Grant GI-29943.* The project was motivated by a widespread national concern that computerized personal information databanks in governmental and civilian sectors, however indispensable for the operation of a modern society, have the potential to infringe upon individual privacy and other civil rights of the citizens.

One line of investigation of the project, the substance of this report, was focused on the protection requirements of personal information databank systems and on the design of effective protection systems. This investigation led to the establishment of classifications of databank systems and sensitivity scales for personal information, derivation of a protector-intruder interaction model as an aid for protection system design, and examination of the technical implications of implementing protection systems.

The objective of this work was to establish a framework for determining protection requirements of personal information databank systems and to provide insights into protection system design. Correspondingly, the material in this report should be helpful to databank system analysts and designers, researchers in the data protection field, and all those who are seeking solutions to the data privacy and security problem.

---

*Appendix B lists the papers and reports published under this grant.

SUMMARY

The problems of potential violations of citizens' rights through computerized personal information databank systems remain in the focus of political, societal, and technical concerns in the United States and other countries. The solutions to these problems will involve legislative as well as technical means. This report focuses on the latter--the technical aspects of implementing information privacy safeguards and data security mechanisms in resource-sharing computer systems.

The nature of the databank ownership, the purposes for which data are collected and used, and the characteristics of the associated computer facility are factors that strongly affect the protection requirements. They are used in this report to establish a *databank classification system* that can be used to establish a vulnerability scale for databank systems. On this scale, the databank systems that are most vulnerable to both the violations of citizens' rights and data security are those operated by private agencies for maintaining identifiable records on individuals, where such databanks are serviced by computer facilities that permit open programming by their users, are accessible on-line from remote terminals, and are shared by other applications and users that are not related to the databank in question.

The type, sensitivity, and potential economic value of personal information are equally important in determining the protection requirements. That is, certain categories of personal information are sensitive, their indiscriminate use or disclosure may cause harm to the individuals concerned, and they ought to be protected against unauthorized access. Also important are such characteristics of personal information as accuracy, precision, completeness, currency, age, and relevance. The different types of personal information and their characteristics are discussed in detail, and an illustrative sensitivity scale is established. Rated highest on this scale is information that, if used indiscriminately, may lead to a threat to the physical safety of the individual involved.

The *sensitivity scale* is then used to propose a protection-oriented classification system for personal information that can be used as a basis for selecting an appropriate protection level. An important consideration here is whether or not protection must be provided on the basis of a statutory requirement. Nine classification levels are defined, starting with "public by statute" as a level requiring the least degree of protection. The highest classification level is "secret by statute," where the existence of the record may need to be kept secret even from the data subject.

In certain databank systems where the so-called "rational" protection policy is applied, it is important to estimate the *economic value* of the protected information in order to choose a commensurate level of protection. To the data subject, the value of protection is related to the losses that he may suffer if the protection fails, such as loss of actual or potential earnings, and victimization by extortion or fraudulent schemes. More indirectly, loss of self-respect or social acceptance are equally undesirable. A detailed discussion and some statistics are presented in Sec. II.

Value of personal information to the databank custodian and users is related to the costs of reproducing the original data in the case of unauthorized modification or destruction, penalties and liability for unauthorized disclosure, and inability to effectively perform the databank functions. To a profit-minded intruder, the value of personal information is its price in the "marketplace" of sensitive mailing lists, frauds, or blackmail.

Economic profit is but one of the motivations for threats against personal information databank systems. Other motivations are related to strengthening individual power, gaining a political advantage, overzealousness in performing duties, informal cooperation between databank agencies, and the like. Sources of threats are also varied: profit-seeking threats are likely to be launched by external intruders and the operating personnel of the databank or user agencies; other types are more apt to involve the databank controller or management directly, or may be perpetrated with their tacit approval.

It is useful in the design of protection systems to examine the general nature of the interaction of a databank protector with a profit-minded intruder. The protector is implementing a rational protection policy--one where the amount of protection provided is balanced against the value of protected entities and the protection cost. The intruder, likewise, examines an intrusion plan from the point of view of the potential gains, expenditures, and risks. This interaction can be expressed in terms of simple mathematical relationships which identify the variables involved and illustrate the situation from a theoretical point of view. Given the ability to express in closed functional forms the potential gains of the intruder, losses of the protector, and the protection costs for both, it is possible to use this model to determine the optimal expenditure policies for both the intruder and the protector. However, at present such functional expressions, as well as the associated measures of security-effectiveness, are still to be derived. Hence the protector-intruder interaction model discussed in Sec. III is an illustration of the potential utility of this approach, rather than a practical tool.

The development of measures of security-effectiveness is an important goal in data security research. Although much more work is still to be done in this area, several candidates have been identified: logical measures that reflect the compliance of a system with security-oriented design-principles; "work-factor" measures that are related to the security system's resistance to tampering, manipulation, and circumvention; and probabilistic measures that reflect the security system's reliability as well as the active security system's ability to detect and discriminate threats. These and the protection and intrusion costs are discussed in detail in Sec. III.

The *protection costs* involve a variety of initial costs for performing security requirements analyses, and designing and implementing the security-oriented hardware and software. The operational costs include the processing time and storage space that must be allocated for the operation of access control mechanisms, auditing, and real-time threat-monitoring devices and procedures. Quantitative information on operating costs is scarce, but it has been estimated that the cost of

a relatively sophisticated access control system may increase the computational overhead 5 to 10 percent, the operating system code by 10 percent, and the main memory requirements of the operating system by 10 to 20 percent.

The *intrusion costs* are much harder to determine, as they tend to be random variables--a lucky intruder may stumble on an exploitable vulnerability by accident while another intruder may spend months in fruitless analyses. As suggested in this report, the protector's costs in penetration testing of a security system may provide an initial estimate of the expected intrusion cost.

Section IV examines in detail the elements of *total protection systems* and the associated design principles, procedures, and technical implications for databank systems. The elements involved are subjects' rights safeguards, procedures for maintaining data confidentiality, data security techniques, integrity management, and methods of auditing, validation, and testing. The objective is to establish a frame of reference for specifying and designing protection systems that provide different levels of security as required by different databank structures and sensitivity levels of stored information.

Important among *subjects' rights safeguards* are procedures that inform the subjects of the existence of records about them in a databank and permit inspection and amendment of incorrect or incomplete records. For a databank this implies establishment of access logs and additional data fields in records for noting the inspection date, actions taken, and linkage to comments or rebuttals that the subjects may have submitted.

The principles for *maintaining confidentiality* include (1) reduction of exposure by collecting only necessary personal information; (2) increasing anonymity by separating identifying information from the rest of the data; (3) reduction of sensitivity (in statistical databanks) by random error inoculation and similar techniques; and (4) providing appropriate access control techniques.

The basic principles of *data security* are (1) the defensive design of the system's hardware and software; (2) establishment of complete control over all users' actions and their processes within the system;

(3) use of concealment techniques (cryptography) in data files and communication links; (4) establishment of effective physical protection techniques; and (5) implementing appropriate integrity management techniques.

The role of *integrity management* in a protection system is to assure quality of the stored data, as well as the correct operation of the protection system. Techniques for data integrity include computational techniques for error detection, such as check-sums, and various codes. Correct performance of the protection system is dependent on the reliability of the associated hardware and integrity of the system's personnel. The original correctness of the protection system is established by *validation and testing*. An important part of these is program proving--a technique for formally verifying that a software routine under examination is performing precisely as planned and in no other way. Testing is necessary to maintain confidence in the integrity of the protection system.

The discussion of the elements of a protection system identifies numerous procedures for their implementation. In general, not all of these are required in every databank system. To illustrate protection systems that provide different protection levels, Sec. IV concludes with the specification of three hypothetical, so-called "model" protection systems that implement "high," "medium," and "low" level protection systems.

The design and implementation of cost-effective protection systems for personal information databank systems is still an unresolved problem. Further research is required in software and hardware techniques for access control, auditing, validation, and testing, as well as for cost-effective implementation of subjects' rights safeguards. Tools and techniques are needed for databank system's analysis for determining protection requirements. Methodologies must be developed for protection system synthesis, implementation, and optimization. The objective of this report is to contribute to the clarification of these problems and, hopefully, to the formulation of the solutions.

ACKNOWLEDGMENTS

The author acknowledges substantial contributions to this report by Norman Z. Shapiro, his colleague at The Rand Corporation who originally formulated the protector-intruder interaction model in Sec. III, provided material for the discussion of the economic value of personal information, and made contributions to several other parts of the report. Discussions with I. S. Reed and M. L. Juncosa, who participated in this research, provided important improvements and insights. Valuable suggestions also came from Helene R. Mills, D. Hollingworth, and J. R. Marlatt, all of The Rand Corporation. The responsibility for the conclusions reached in the report, and for errors which may have evaded careful scrutiny, rests, of course, with the author.

CONTENTS

PREFACE iii

SUMMARY v

ACKNOWLEDGMENTS xi

FIGURES xv

TABLES xvii

Section
I. INTRODUCTION 1

II. DATABANK SYSTEMS 9
A Structural Model 9
Classification 11
Public and Private Databanks 11
Dossier and Statistical Databanks 14
Centralized and Decentralized Databanks 15
Dedicated and Shared Databanks 16
Closed and Open Databanks 16
Off-line and On-line Databanks 16
Specific and Integrated Databank Systems 17
Personal Information 17
Information Types 18
Characteristics 19
Collection 24
Sensitivity Scales 25
Classification 28
Economic Value 32
Threats to Databanks 41

III. A MODEL OF PROTECTOR-INTRUDER STRATEGIES 47
Protection Policies 47
An Interaction Model 48
Measures of Effectiveness 55
Security Systems 56
Access Control Barriers 58
Effectiveness 60
Protection and Intrusion Costs 63
Protection Costs 63
Intrusion Costs 70

IV. PROTECTIVE SYSTEMS FOR DATABANKS 72
Elements of Total Protection 72
Subjects' Rights Safeguards 73
Maintenance of Confidentiality 76

Data Security 76
Integrity Management 80
Audit, Validation, and Testing 82
Model Protection Systems 85

V. CONCLUDING REMARKS 90

Appendix
A. SELECTED AMERICAN VALUES 93
B. RAND PUBLICATIONS UNDER NSF GRANT GI-29943 95

REFERENCES 97

FIGURES

1. The Databank System 12
2. Intruder's Gain 52
3. Protector's Loss 53
4. Optimal Choices of X and Y 54
5. Access Control Barrier 59

TABLES

1. Databank Classification Dimensions 13
2. Public Views on Information Sensitivity 23
3. A Scale for Data Sensitivity (British) 27
4. Illustrative Sensitivity Scale for Personal Information 29
5. Classification of Identified Personal Information Items and Records 33
6. Combined Statistics on "Pigeon Drop" and "Bank Examiner" Bunko Frauds in California 36
7. Threats to Public Databanks 45
8. Threats to Private Databanks 46
9. Examples of Processing Time Requirements for Access Control Procedures 67
10. Illustrative Examples of Encryption/Decryption Processing-Time Requirements 68
11. Implications of Subjects' Rights Requirements 74
12. Principles for Maintaining Confidentiality 77
13. Principles for Data Security 78
14. User Capabilities and Security Risks 81
15. Illustrative Security Vulnerability Rank-Ordering of Databank Classes 82
16. Integrity Management Techniques 83
17. Techniques for Auditing, Validation, and Testing 86
18. Definition of Protection Levels 86
19. A "Low" Level Protection System 87
20. A "Medium" Level Protection System 88
21. A "High" Level Protection System 89

I. INTRODUCTION

Personal information databank systems--computerized collections of information on individual citizens--have become an indispensable component in our way of life. They are used to process financial transactions, to maintain records on the interactions of citizens with their government, and to respond to a myriad of requests for information which is then used to make decisions and take actions about individual citizens.

Under these circumstances, the integrity of the information in a databank system becomes an important consideration, and it is necessary to implement techniques of *data security*--the protection of the data, as well as programs and computer equipment, against unauthorized, accidental or deliberate disclosure, modification or destruction.

Equally important to the individual is the nature of information that is stored about him, and the purposes for which it is being used. Indeed, it is pointed out in the Congressional Hearings of 1966-1967 [1,2], and more recently in 1972 [3], in numerous books, reports, and articles (for example [4-6]), and in recent reports by HEW Secretary's Advisory Committee on Automated Personal Data Systems [7] and by a study for the National Academy of Sciences by a group headed by Alan F. Westin [8], that the establishment of personal information databank systems raises serious concerns about potential violations of *privacy* and other *civil rights* of the involved individuals.

Although in this report the focus is on data security aspects of personal information databank systems, it is useful to briefly discuss their privacy and civil rights protection aspects, since these match the importance of data integrity requirements in providing a rationale for data security.

The concern over potential violations of individual civil rights by computerization of personal information record-keeping systems is a manifestation of the classical conflict between, on the one hand, government, business, and industry, which desire the increased efficiency and economy that computerization can bring, and, on the other

hand, the citizens who refuse to accept the infringement on their rights and freedoms that computerization of personal information record-keeping systems may entail.

In general, two issues are involved: concern for the preservation of democratic institutions and concern for protecting the citizen from individual injury. Implicit in both is a concern for human rights, but in the first case it is a collective concern while in the second case it is an individual concern.*

This conflict is not new. Records on individuals have been kept from the days of antiquity. Although to date the automation of record keeping--establishment of computerized record-keeping systems and databanks--has not significantly altered the practices and policies employed in manual record-keeping systems [8], it has increased the *degree* of potential violation of individual privacy and other freedoms.

For example, in manual data files several considerations tend to limit the ability of an intruder to utilize personal data to generate threats to individuals [9]. Among these are his ability to:

- Bring together data that have been available, but have been uncollected and uncollated;
- Record new data with the precision and variety required to gain deeper insight into the private person;
- Keep track of a particular person in a large and highly mobile society;
- Gain access to already filed data about a person;
- Detect and interpret potentially revealing private information among the data to which he has access.

Computerization of personal information data files tends to make the above considerably easier and more efficient. On the other hand, however, computer technology also offers opportunities for setting up more effective access control features than possible in manual record-keeping systems. Hence, computerization has brought the privacy versus

*The author is indebted for this formulation of the databank problem to Jerry Marlatt of The Rand Corporation.

efficiency conflict in the use of personal data into a much sharper focus, and has accelerated the search for ways and means for reducing the potential infringement on individual rights.

Several concepts are involved. First, there are the philosophical, political, and legal questions dealing with individual rights of *privacy* and *due process*, as they relate to the collection and use of personal information. These are defined in the HEW Secretary's Advisory Committee report as follows [7]:

> An individual's privacy is directly affected by the kind of disclosure and use made of identifiable information about him in a record. A record containing information about an individual in identifiable form must, therefore, be governed by procedures that afford the individual a right to participate in deciding what the content of the record will be, and what disclosure and use will be made of the identifiable information in it. Any recording, disclosure, and use of identifiable personal information not governed by such procedures must be proscribed as an unfair information practice unless such recording, disclosure or use is specifically authorized by law.

Included in this statement are the basic elements of an individual's right to due process--to have rules of conduct specified in advance, a fair hearing to defend himself against punitive actions, and an appeal to higher authority for review. The privacy and due process considerations are summarized in the following requirements formulated for personal information by the HEW Committee [7]:

- 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 in a record and how it is used.
- There must be a way for an individual to prevent information about him obtained for one purpose from being used or made available for other purposes without his consent.
- There must be a way for an individual to correct or amend a record of identifiable information about him.

- Any organization creating, maintaining, using, or disseminating records of identifiable personal data must assure the reliability of the data for their intended use and must take reasonable precautions to prevent misuse of the data.

A basic consideration in the potential violation of individual rights through personal information databank systems is the *initial* collection of personal data, and the following questions are but an example of those that should be answered by any databank system: Are the collected personal data necessary for the stated purposes? Is it necessary to store the data in computer-accessible form? Is the length of time proposed for retaining the data warranted by their anticipated uses?

While the answers depend on the specifics of the activity involving the use of personal information, a "principle of least privilege" should be applied in all cases--only the necessary data, and no more, should be collected and kept in computer-accessible form, and retained for a minimal time period. In accord with due process considerations, subjects of the data or their representatives should participate in answering these questions.

Important in the collection and use of personal information is the concept of *informed consent*: when a person agrees to share information about himself with the society, there is no invasion of privacy *unless* the conditions of consent are violated by the databank or its users. However, an informed consent requires a clear explanation to the subject of the purpose of data collection, the extent of dissemination, the degree of protection provided for the data, and their ultimate disposition. For example, while a person may agree to participate in a political sciences study of voting patterns, the conditions of his informed consent and, hence, his privacy would be violated if the data were given to a political group for mailing-list purposes.

Another concept associated with personal information is *confidentiality*--a special status given in the record-keeping system to personal data whose use and dissemination are restricted to only specified purposes and users. Data confidentiality may be established by statutory

means (such as in the case of the census information), by administrative procedures (such as salary information), or as part of the consent conditions. Confidentiality requirements are implemented by operating procedures and data security techniques.

In addition to privacy, due process, and confidentiality, other important concepts in our society and form of government affect the operation of databank systems and, at times, may lead to conflicting requirements and goals. Among these are:

- *Accountability* of a governmental agency to the society and its representatives; a reflection of the American anxiety that the government should be controlled by, and be responsive to, the people;
- *Openness* (freedom of information) within the government as specified by the Freedom of Information Act [10] and similar acts in individual states.

Added to these is the traditional American desire for efficiency. In the context of databanks, this manifests itself in terms of *utility* of the databank to its present and (potential) future users, *flexibility* in information processing, and *economy* in the databank operation.

Each of the above represents demands that may be rightfully placed on the design and operation of a personal information databank system. Among the conflicts in goals that this may generate are:

- Overzealous efforts by the databank agency and users to satisfy the accountability requirement (i.e., to do as good a "job" as possible) may lead to efforts to gather more information on the subjects than necessary and, thereby, increase the potential for violation of the subjects' rights.
- Satisfying the accountability demands of the society by allowing inspection of the databank activities implies increased openness and compliance with the freedom-of-information requirements. This, in turn,

implies less confidentiality and, consequently, a potential violation of privacy of the subjects.

- Increasing confidentiality afforded to the databank in the name of protecting privacy also reduces openness and accountability.
- Society demands economy, efficiency, and a high degree of utility from public agencies. This is in conflict with the subjects' demand for confidentiality, and restrictions in the use, dissemination, and sharing of data.

In general, custodians and users of databank systems tend to regard demands for privacy protection by databank subjects as counterproductive to their traditional goals of economy, utility, and completeness of the data files. Further, in governmental databank systems the accountability and confidentiality requirements are usually established by legislation or, as authorized in legislation, by heads of executive agencies. However, in the name of economy, utility, or interagency cooperation, a databank agency may find ways to permit data sharing and exchanges, thus setting the stage for potential violations of the subjects' rights.

In private databank systems, pressures for public accountability, as well as accountability to the databank subjects, are virtually nonexistent. Confidentiality measures may be incorporated, but it is likely that the motivation is to protect the databank operations from competitors rather than to protect the privacy of the data subjects. In some private databank systems, even the data accuracy considerations may yield to economy and efficiency.

Finally, maintenance of a proper balance of political power between the government and the governed, as well as between the different branches or levels of the government, is a problem. In general, computer-based information systems in government are "power enhancing" as they tend to reinforce existing governmental power structures [11] and to increase the difficulty for citizen groups to gain access to the same information and perform similar analysis as the agencies they attempt

to examine. Recognizing this, the West German state of Hessen, in one of the few existing statutes for controlling personal information databanks [12], has given a Data Protection Commissioner the power to examine proposed databanks, also from the point of view of their effects on existing balance of power.

Firm accountability, openness, and data confidentiality requirements must be established by law to reduce the pressures of these conflicting design goals on databank systems and to assure proper protection of the subjects' privacy. The Fair Credit Reporting Act [13], the Code of Fair Information Practices [7], and the various provisions enacted or proposed for enactment in Western European countries [12, 14,15] and in Canada [16] contain a number of such requirements.

Within the described framework of societal and political pressures that bear on the designers and operators of personal information databank systems, this report examines the technical aspects of providing data security in different types of databank systems. The equally important technical problems dealing with privacy and other civil rights protection requirements are discussed in less detail, since they are not entirely within the scope of the research being reported. The overall objective is to clarify the data protection problem and organize the material that must be used in the design of protective systems.

More specifically, Sec. II presents a structural classification of databank systems that focuses attention on protection problems; identifies and examines the relevant characteristics of personal information, the "commodity" being protected; proposes a sensitivity scale and classification system for personal information; and examines the general nature and sources of potential threats against databank systems. In Sec. III, the focus is on a game-theoretic framework for the design of protection systems that is based on a mathematical formulation of the interaction of a databank protector with an intruder who is striving to gain economic profit. The interaction model is defined and the necessary variables are identified and discussed. Among such variables are measures of security-effectiveness of various protection mechanisms, the value of protected entities to the parties involved, and the costs of both the protection and intrusion.

Section IV examines in detail the elements of a system that provides protection to subjects' rights, maintains data confidentiality, provides data security, includes elements of data integrity management, and implements the necessary features for auditing the effectiveness of protective devices and procedures. Prototype protective systems that implement three levels of protection (low, medium, and high) are formulated in terms of protection features that should be included. Concluding remarks and recommendations are presented in Sec. V.

II. DATABANK SYSTEMS

The term *databank* implies a systematically organized collection of data to which a number of users have access. An often used, synonymous term is *data base*. A computerized personal information *databank system* consists of the data files, the associated computer facility (processors, storage devices, terminals, communication links, programs, and operating personnel), a management structure, and other "interested parties."

A STRUCTURAL MODEL

Implicit in the definition of the databank system are several agencies, groups of persons, or individuals that have distinct roles in the functioning of the databank system. The interactions of these groups have a distinct bearing on the privacy protection and data security required and provided. The following can be identified [17]:

- *Subject*--an individual or an organization about whom data are stored in the databank system.
- *Controller*--a person, agency, or institution (public or private) with authority over the databank system and its operations. For example, the controller may be a legislative body or the director of an agency. The controller authorizes the establishment of the databank system, specifies the population of subjects and the type of data collected, and establishes the policies for data collection, use, dissemination, and protection.
- *Custodian*--the agency and its personnel in physical possession of the data files. The custodian is charged with the proper operation of the databank and is responsible for implementing and abiding by the policies established by the controller.

- *Collector*--the agency and its personnel who collect data from subjects and transmit these to the custodian. Included may be various preprocessing steps, such as conversion of the data into computer-readable form, if these are not performed by the custodian agency. In some instances, the collector agency's staff may be very large, but relatively poorly trained in questions of privacy protection.
- *Users*--person or agency authorized by the controller or the custodian to utilize specified subsets of the personal data for purposes specified by the controller, subject to the disclosure and dissemination policies of the databank system.
- *Databank*--the personal information files in computer-readable form, and the associated storage devices.
- *Computer facility*--the computer equipment for information processing and interacting with the databank and its users.

Other parties who may be involved and who may be interested in the databank and its uses are:

- *Society*--the population within which the subjects have rights and obligations, and whose welfare also affects the welfare of the subjects. The society's claim for freedom of information and openness of databank operations, as well as the accountability of all other elements of the databank systems to the society, may lead to conflicts with the right of privacy of its individual members.
- *Intruders*--persons or agencies which either deliberately or accidentally gain unauthorized access to the databank or make unauthorized use of the data that are normally available to them as authorized users.

- *Other users*--agencies and their personnel that share the computer facility with a databank, but are not authorized users of the databank.

Figure 1 illustrates the structure of a generalized databank system, displays the more prominent lines of communication between its components, and shows where the privacy, confidentiality, security, and other dimensions manifest themselves. It should be observed that these components need not be unique; multiple roles and overlapping functions are common. For example, the controller, custodian, and collector may be the same agency, and under the various circumstances discussed later, any one may become an intruder.

CLASSIFICATION

The nature of the databank ownership, the type of data collected and their principal use, and the characteristics of the associated computer facility strongly affect the threats to individual privacy and the complexity of the data security problem. It is useful, therefore, to establish a classification system for databanks which adequately reflects the privacy and security requirements. Table 1 depicts the selected classification dimensions and the corresponding classifications.

Public and Private Databanks

This classification dimension refers to the nature of the databank controller. *Public* databank systems are owned and operated by government agencies at all levels of government under the authority of the corresponding legislative bodies. They may belong to a single governmental agency, cooperatively maintained by several agencies in a single or several levels of government, or serve as an independent service organization. Examples of these types include a databank system operated by a State Department of Motor Vehicles; a regional law enforcement system (e.g., the Cincinnati-Lane County CLEAR system [18]); the National Crime Information Center (NCIC) which links federal and state law enforcement systems [19]; and the New York State Identification

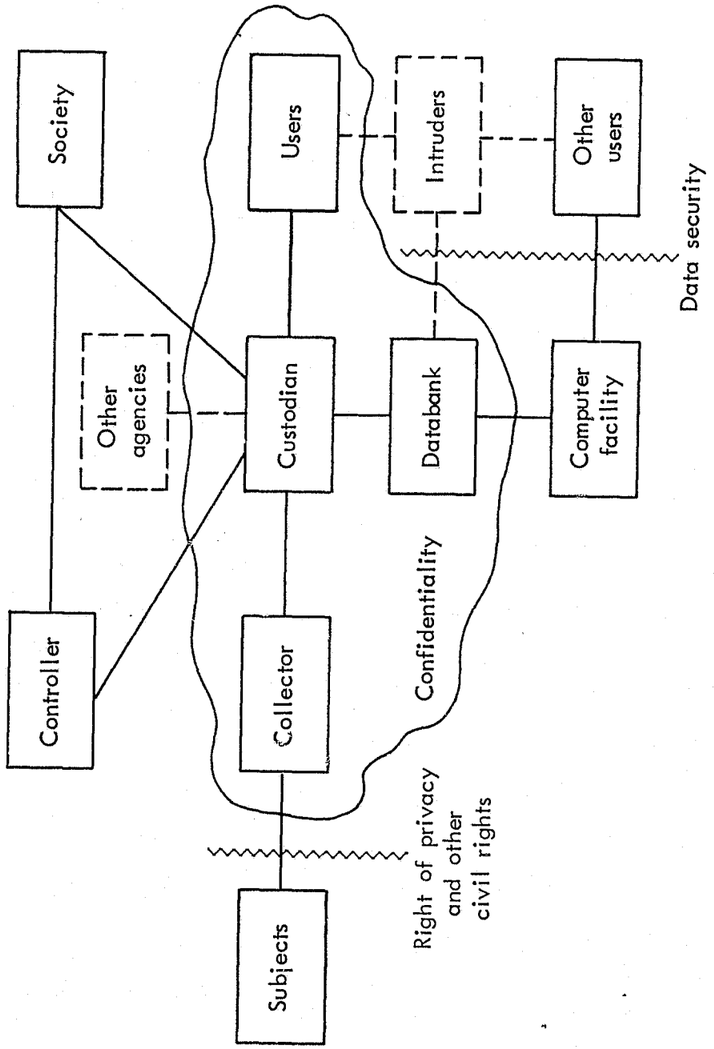

Fig. 1 — The databank system

Table 1

DATABANK CLASSIFICATION DIMENSIONS

Dimension	Classes and Subclasses
Controller and custodian	Public Single agency Multiple agencies Independent service organization --- Private
Function and uses	Dossier Administrative Intelligence --- Statistical
Scope and user community	Specific --- Integrated
Databank organization	Centralized Geographically Functionally --- Decentralized Geographically Functionally
Computer facility usage	Dedicated (to the databank) --- Shared (with others)
Access (to databank and to the computer facility)	Off-line --- On-line Local terminals Remote terminals Direct access to computer Indirect access
User services (in the databank and in the computer facility)	Closed (transaction-oriented) --- Open (user programs allowed)

and Intelligence System (NYSIIS), which is a statewide repository of criminal history information [20].

All nongovernmental databank systems can be classified as *private*. In this category are databanks controlled by a single industrial or business enterprise as well as those operated for a large segment of some industry. The Medical Information Bureau (MIB), which stores and handles insurance applicants' and policyholders' medical history exchanges, is an example of the second type of private databank. There also exist composite *public-private* databank systems which, typically, are private institutions that maintain and analyze government-collected personal information for preparing statistical summaries for supporting public programs.

Dossier and Statistical Databanks

The principal classification criterion here is whether or not individuals must be identified in the response to an information request. Databanks where an individual must be precisely identified in the output may be classified as *dossier-databanks*.* If the main purpose of such a databank is to store factual records of transactions involving individuals and to provide information for making decisions about them, the databank has an *administrative* function. Examples of large administrative databanks are the system at the Social Security Administration which contains the earnings and benefit payment records of more than 150 million citizens [8]; the California Department of Motor Vehicles databank on driver licenses and car registrations which has over 42 million records; and the over 10 million financial records of the customers of the Bank of America [8]. Hundreds of other administrative databank systems can be found in industry, business, hospitals, schools and universities, and state, county, and local governments all over the country.

Intelligence databank systems are another type of dossier-databanks. Their purpose is to store information about individuals for making judgments about their intentions, views, trustworthiness, and future

*The term "dossier" is used to mean "a file of information on an individual."

actions. Information in these systems is often collected without the subjects' knowledge. Federal security clearance records, investigative files in law enforcement agencies, and consumer credit reports are representative of records kept in intelligence systems. Examples of large intelligence databank systems are the NCIC and NYSIIS already mentioned above, and the TRW-Credit Data Corporation with its credit information records on more than 25 million individuals [8].

Statistical databanks gather information on individuals for the purpose of gaining knowledge about *classes of individuals* rather than specific ones. Identification of subjects within a statistical databank is required only for updating their records for longitudinal or time-series studies. Accuracy of data is desirable but not essential. Large governmental statistical databanks are maintained by the U.S. Bureau of the Census and the American Council of Education [8].

Depending on the specific information stored and the needs of the information seeker, the dossier and statistical databanks can also be used for other purposes, although on a limited scale. For example, administrative records can be used for certain statistical purposes; statistical records can, when identification is included, be used for intelligence purposes; and so forth. However, there appears to be no present large-scale trend in gathering into administrative databanks information which may have intelligence value but is irrelevant for the administrative databank purposes [8].

Centralized and Decentralized Databanks

This classification dimension refers to the physical organization of the databank system. In a *centralized* databank the files are at one geographical location and are processed at the same computing facility. In a *decentralized* databank, however, some of the files are at different locations, use different computing facilities, and have different custodians and operating personnel. Implicit in the definition of the databank system is that even if some of the data files are at different locations, they are still available to all authorized databank users through some communication channel (i.e., the data are *functionally* centralized in both cases).

Most of the contemporary databank systems are centralized in the above sense. An example of a large *decentralized* databank system is that operated by the Internal Revenue Service [21], which has "branch" databanks at several locations in the United States.

Dedicated and Shared Databanks

This dimension applies separately to the databank files and to the associated computing facility. A *dedicated* databank is used only by the agency that is also the databank custodian, while a *shared* databank is used by other agencies in addition to the custodian agency. This classification dimension can be also applied to the computer facility. Thus, a dedicated computer facility is used only by the databank system. A shared computer facility has other users in addition to the databank system. NYSIIS [20] and the Santa Clara County LOGIC [22] systems are examples of shared databanks. NYSIIS is operated on a dedicated and LOGIC on a shared computer facility. Many smaller databanks in industry and business are dedicated, but operated on a shared facility. Typical of these is the processing of personnel records on a computer service bureau's facility.

Closed and Open Databanks

The criterion for this classification is whether users can write and execute their own programs for operating on the databank files--the *open* databank--or whether they are restricted to only those programs provided by the databank system--the *closed* databank. The latter is also called a *transaction-oriented* databank system. Many administrative databank systems are transaction-oriented, while open databank systems are more frequent among statistical record-keeping systems.

Off-line and On-line Databanks

In an *off-line* databank system (or computer facility) information and processing requests are placed into a queue and processed at a time determined by the operating system's scheduling procedures. A user typically does not know when his request is executed and has no control over the processing.

An *on-line* computer facility is equipped with input-output terminals, located at the computer facility or at remote sites. A user can interact with his programs or information requests when these are executed in the computer. An *on-line databank* can be similarly characterized.

The two types of on-line databanks are (1) *directly* on-line, where the user enters the request directly to the computer from his terminal, and (2) *indirectly* on-line where a databank employee acts as an intermediary. In the latter case, the user places a telephone request to an employee of the databank, who operates a terminal and conveys the output back to the user over the telephone. The NYSIIS operation is an example of an indirectly on-line system.

Specific and Integrated Databank Systems

A *specific* databank stores information on a set of individuals who have a certain common characteristic, for example, the National Driver Register which contains information on persons whose driver licenses have been denied, terminated, or suspended [23]; the Organized Crime Information System maintained by the Justice Department [24]; and the Narcotic Addicts System of the Bureau of Narcotics and Dangerous Drugs [25].

An *integrated* databank contains a variety of information on an individual, used for a variety of purposes. Most administrative databanks maintained by local and state governments and private industry on taxes, vital records, employment, education, social programs, and the like are in this category.

PERSONAL INFORMATION

Personal information (personal data) has been defined as follows:*

> The term "personal data" includes all data that (1) describe anything about an individual, such as identifying characteristics, measurements, test scores; (2) indicate things done

*Senate Bill S.2810, 93d Cong., 1st Sess. Introduced by Senator Goldwater on December 13, 1973, U.S. Government Printing Office, Washington, D.C., 1973.

> by or to an individual, including, but not limited to, records of financial transactions, medical treatment, or other services; or (3) afford a clear basis for inferring personal characteristics or things done by or to an individual, including, but not limited to, the mere record of his presence on a place, attendance at a meeting, or admission to some type of service institution.

Depending on the circumstances, context, existing statutes, and the subject's and the society's value systems, certain records or information items in the records may or may not require protection. That is, certain categories of information about individuals are sensitive and ought to be protected, while other categories are not. Various classes of personal information items and their characteristics that affect privacy and security questions, and guidelines for their classification are discussed below.

Information Types

Although personal information in databank systems appears in many forms, two general types can be identified: factual statements and evaluative statements. Both refer to an individual's person, and his past and present activities. Evaluative statements often also contain predictions about his future behavior.

- *Factual statements*--can be proven to be true on the basis of formal documentation or other irrefutable evidence. Personal information items of this type include [26]:

  --*Identifiers* (assigned or formally assumed) such as names, social security number, numbers of licenses, certificates, permits, and so forth.
  --*Permanent physical characteristics*, including sex, race, blood type, fingerprint classification, height, eye color, handicaps, and scars.
  --*Genealogical information*, such as birthdate and place, parents' names, nationality.

  --*Current status information*, including a person's time-varying physical characteristics, address, occupation, education level, place of employment, military service status, family, participation in social programs, judicial status, finances, property, health, licenses and privileges, political affiliation, and membership in associations.
  --*Transaction history* in the areas listed above as current status information.

- *Evaluative statements* are opinions, judgments, and allegations regarding the individual by others and, in some instances, the individual's self-evaluation. These deal with the individual's character, views, values, habits, behavior, and activities.

A general problem associated with computerized information systems is the tendency of their users to regard all data in the system as factual. It is important that the "computer knows best" syndrome be exposed and replaced with the understanding that computerized databanks are just as fallible as the manual record-keeping systems they have replaced.

Characteristics

Associated with files, records, and specific information items is a set of characteristics that are important from the privacy and security point of view. In particular, among those characteristics that reflect the *subjects'* concern with privacy and due process are:

- *Accuracy*. The information item stored in the databank may be incorrect, for whatever reason. However, the databank subjects expect and the users assume that at least the factual data are correct.
- *Precision*. This refers to the amount of information in the data item. For example, if coding is used to categorize individuals into representative groups,

the number of different choices available for assigning an individual determines how closely the selected group corresponds with the actual case. Low precision necessarily results in miscategorizations and, hence, may lead to incorrect decisions and actions.

- *Completeness*. An information item may provide some, but not all applicable information. A well-known example is the storing of records of arrest without information on disposition.
- *Currency*. This characteristic refers to the time-accuracy of the present status and the more recent historical information: Has this information been updated?
- *Age*. At the other end of the time spectrum, certain historical information may become too old to remain included in the active records and used for decisions and actions. It should be purged from the record.
- *Relevance*. This characteristic refers to the bearing that an information item has on the specific purposes and uses of the databank. Certain information that is not related to the specific decision or action may, nevertheless, emotionally affect the decisionmaker. Other information items, such as an individual's religion, race, and sex, have constitutional or statutory restrictions on their use for certain decisions, and their inclusion into a record must be justifiable.

These characteristics reflect the expectations of the databank subjects as well as the constraints that must be placed on the users. Increasingly, decisions about individuals are made by anonymous officials or, as may be the case, even by computer programs. The least that can be done for the subjects is to assure data integrity--that information used is accurate, complete, and up-to-date. Precision must be maintained in data coding to assure that each case is given appropriate individual consideration. The age characteristic reflects

a basic American expectation--the forgiveness by the society of misdeeds of the distant past which have long been atoned. Relevance of information used in decisionmaking is also a part of an American expectation for fairness. It is not fair to the individual concerned when decisions about him are not based on the facts of the situation, but also on available extraneous information which may adversely bias the decisionmaker.

Other characteristics of personal information are more directly related to data security requirements and implementation of data security systems. Principal among these are:

- *Availability* of a given information item outside the given databank system. There are two components to availability of information regarding an individual: (1) available to others, and (2) available to individual himself. Regarding the latter, it is a general custom to make certain evaluatory statements unavailable to the data subject.
- *Sensitivity*. The potential of an information item or a set of information items to damage the individual when made available to the public, government, or specific individuals and agencies or when surreptitiously modified. The damage may be economic, psychological, or physical. In some instances, the interests of the society might also be adversely affected if the information were released to the public or to the subject individuals.
- *Economic value*. The "convertibility into cash" of personal information by intruders, the cost of the information loss to the databank custodian and users, as well as the potential economic damage to the data subject.
- *Identifiability*. The degree to which an information item, or a set of items, allows unique identification of the corresponding individual. This is determined,

in part, by the uniqueness of the information and the context of the databank. For example, the occupation "governor" uniquely identifies an individual in a databank of state officers and employees.

The degree of *availability* of certain information items outside the given databank may greatly influence the necessity for providing protection. Certain information is publicly available to anyone, such as names, addresses, and telephone numbers in telephone books or city directories, or property tax information at the assessor's office. However, in the context of a specialized databank system, the same public information may become very sensitive. Other information items, while individually available as public information, may become sensitive when collected into a single record. More often than not, the information content of a record is greater than the sum of the information in its parts.

Information *sensitivity*, likewise, depends on the context of data collection and individual circumstances. To illustrate this, Table 2 summarizes the results of two recent surveys on public views of information sensitivity [27,28].

Identifiability, likewise, affects the need for protective techniques. If a data item or a set of them cannot be related to a unique individual or a sufficiently small set of candidates, the data may not need privacy protection. The output from statistical databanks, in particular, falls in this category.

There are several degrees of identifiability:

- *Explicit* identification on the basis of some identifying information items which may be either directly associated with the record or linked to the record through traceable codes, aliases, and the like.
- *Inferential* identification through some combination of data items unique to an individual even though no explicitly identifying data items are included.

Table 2

PUBLIC VIEWS ON INFORMATION SENSITIVITY

American Survey

Information Type	Objected to Inclusion in a Databank[a] (%)
Police records	14
Medical records	17
School records	20
Tax records	22
Credit ratings	22
Employment records	24
Salary records	43
Political activity records	45

British Survey

Information Type	Objected to Open Publication[a] (%)
Address and telephone number	33
Occupation	12
Education	17
Political views	42
Religious views	28
Income	78
Medical history	51
Details of sex life	87

[a]The questionnaire referred to current status information, not necessarily historic information.

- *Anonymity* of the record (except possibly for temporary identifiability in the initial data collection, conversion, and merging process).

The prevention of inadvertent identification of individuals by correlating sets of statistical summaries--the residual disclosure problem--remains among the more difficult ones in statistical databank systems [29,30].

Collection

Personal information is collected into databanks (1) with or without the individual's *consent*, and (2) with or without the individual's *knowledge*. Considering the four possible combinations of these, an individual contributes his personal information for the following reasons:

- *Mandatory*. Information must be provided under the penalty of law, without the individual's consent but with his knowledge. Examples are the U.S. Census, the governmental taxation agencies, and law enforcement agencies.
- *Quasi-mandatory*. Information must be provided in order to qualify for certain privileges or benefits which, in principle, are optional. Information is given with consent and knowledge. Examples are the automobile drivers license, public welfare privileges, and the like.
- *Voluntary*. An individual participates in some survey or research program. He provides information with full consent and knowledge.
- *Surveillance*. Information is gathered without the individual's consent or knowledge, such as for various investigations of intelligence gathering. A variation here is the situation where, for the purpose of obtaining certain benefits, such as employment or a security clearance, an individual may consent to an agency's collecting information about him even if the substance of the information or its sources remain unknown to him.

Gathering information without an individual's knowledge, and keeping the existence and content of such a record secret from the individual is, if the information is used for making decisions and taking actions regarding the individual, contrary to the principles of due process. Indeed, it has been recommended that no personal information databank whose very existence is secret should be permitted [7].

Information that is gathered from all citizens in a *mandatory* manner requires and is often given special protection. For example, the information gathered for the U.S. Census is given by law a guarantee of confidentiality from the scrutiny of other government agencies as well as assurances that the information is to be used for statistical purposes only.

Maintenance of privacy and confidentiality requirements during the data collection process is often rather complex. It may involve various intermediary groups, administrators, transcribers into computer-readable form, and couriers. Special procedures and strategies may need to be implemented [31-33].

Sensitivity Scales

Since it is reasonable to expect that the databank subjects will demand more protection for information they consider sensitive, it would be useful to establish a *sensitivity scale* for assigning sensitivity levels to different information items, records, and files. The sensitivity levels could be used as a part of the input information for establishing data security requirements.

Several suggestions for sensitivity classification have already been made for specific record-keeping systems. For example, a guideline document for collection and dissemination of pupil records in schools [34] defines the following levels (in increasing order of sensitivity):

- *Category A*. Official administrative records that constitute the minimum personal data on students necessary for the operation of the school (identifying data, academic work completed, level of achievement, and attendance).
- *Category B*. Verified information of clear importance but not absolutely necessary (intelligence and aptitude test scores, health and family background data, teacher and counselor ratings, and verified reports of serious or recurrent behavior patterns).

- *Category C.* Potentially useful information, but not verified or clearly necessary beyond immediate use (legal or clinical findings, personality test results, unevaluated reports by teachers or counselors).

Another set of sensitivity levels has been proposed for criminal justice information systems [35]:

- *Highly sensitive*, such as arrest records without conviction, criminal history records accessed on a class basis, and intelligence files.
- *Confidential*, such as criminal justice information on individuals disseminated to criminal justice agencies, and research reports derived from criminal justice information on individuals.
- *Restricted*, the lowest sensitivity level.

A third sensitivity scale proposed by a Committee of the British Computer Society [36] is listed in Table 3.

While it is not likely that an all-encompassing sensitivity scale can be established, a general approach can be based on the *values* that individuals aspire for, and that are expected of him by members of his immediate social sphere (family, friends, associates, employers, community), authorities, and society at large. A partial list of contemporary American values [37] is given in Appendix A. It must be remembered, however, that in present times such values are changing and that tolerance of nonconformance with the traditional values is increasing.

In general, personal information becomes sensitive when its dissemination may have potentially adverse effects on the individual and his interactions within his social sphere, or when it can reveal that the individual does not satisfy the values expected by some group in his social sphere or those strived for by the individual himself. There are two types of situations: (1) the individual knows the information but wants to limit its circulation, (2) some group knows the information

Table 3

A SCALE FOR DATA SENSITIVITY (BRITISH)

Value Scale	Examples
10	Secret information (diplomatic secrets, defense secrets)
9	Police records relating to convictions
8	Confidential police records (e.g., records used by inquiry agents)
7	Commercial secure information (e.g., trade secrets)
6	More sensitive financial information (e.g., company finances)
5	Financial information (e.g., bank records, medical records)
4	Vehicle licensing systems
3	Public information in schools
2	Public utilities account inquiry systems
1	Selected general information (e.g., titles such as Miss or Mrs. which indicate marital status)
0	Information collected and available, such as telephone books, professional listings

but wants to keep it from the individual for "his own good" or for society's good. Examples of the first case are an individual's transgressions, views, or associations in the distant past. Examples of the second case include (1) the results of medical or psychiatric examinations, IQ scores, and other information that may adversely affect the individual's psychological well-being; or (2) information on ongoing criminal investigations of the individual.

There may also exist a third situation--a group that would attempt or threaten to deliberately disseminate the adverse information or use it to cause losses to the individual. Such a group or their agents are among the potential intruders to a databank system. Further, intrusions are not necessarily planned against specific individuals but may also be perpetrated on a "class action" basis. Indeed, the capability of a computerized system to rapidly process large amounts of

data makes such a class action intrusion more attractive than an "individualized" one. For example, using psychiatric files, a "mailing list" can be compiled on individuals who are very susceptible to a certain sales approach.

In general, an individual's desire to withhold information from a specific group depends on the value he places on the group's view of him, the potential adverse effects on his well-being that may result, and on his assessment of the likelihood of intergroup dissemination of this information. For example, he may reveal an embarrassing item to a group of loyal friends but not to his family. Clearly, the composition and ranking of these groups is highly dependent on the information involved and the social sphere of the individual.

As an example of using the value-based approach, Table 4 presents an illustrative sensitivity scale for personal information. It is also possible, of course, for the adverse effects that result from a disclosure to escalate. For example, release of information that leads to loss of self-respect may further lead to antisocial behavior, loss of employment, and serious mental conditions.

Classification

The establishment of a standard, accepted classification system for controlling the dissemination and use of personal information is a necessary step toward systematic specification of data security requirements and design of data security systems.

Given such a classification system, standard protection requirements, handling and accountability procedures, personnel clearance criteria, retention periods, criteria and authority for initial classification and reclassification, disposition procedures and penalties for willful compromise could be established for each classification level. Such a system has existed for many years for defense classified information [38] and its implementation in computerized information systems is under study [39,40].

A standard classification system established and enforced under federal and state statutes has the obvious benefits of clarifying for everyone involved the level of protection that can be expected and

Table 4

ILLUSTRATIVE SENSITIVITY SCALE FOR PERSONAL INFORMATION

Sensitivity Level	Potentially Adverse Effect on Subject	Revealed To	Examples of Information Revealed
5	Physical safety and well-being	Community Associates	Subject is an undercover agent of an investigative agency
	(Not applicable)	Subject	Subject is under investigation for a criminal offense
4	Physical liberty, right to refrain from self-incrimination	Authorities	Self-reported information anti-social or illegal activities
	Mental and physical health and well-being	Subject	Psychiatric evaluations
3	Economic security and opportunities, employment, self-advancement	Employer Agencies involved	Lapses of self-control, medical and psychiatric records, criminal history
	Family life	Family	Extra-marital affairs, sexual deviations
	Devotion to family, domestic virtues	Subject	Evaluative statements by family members
2	Reputation, respectability, recognition, acceptance	Friends Associates Community	Information about political views, anti-social behavior, criminal history, evaluative statements by subject, finances
	Self-respect, strength of character, competence, loyalty	Subject	Evaluative statements by others
1	Solitude, privacy, friendship, tolerance	Community Associates Subject	Remarks made in private, publicly available information not widely disseminated, information on preferences, property, leisure activities
0	No applicable adverse effects, possibly annoyance	Anyone	Widely published and available factual statements

must be provided, and the consequences of not doing so. There are also some shortcomings. The confidentiality-openness-accountability conflicts between the databank subjects, custodians, and the society are continuing and, just as in the case of classification of defense information [41], there will be charges that databanks are using classification as a means for escaping their accountability to the public.

Another consideration in the establishment of a classification system is the number of classification levels that are defined. Too many levels may make the administration and enforcement of the system excessively cumbersome and costly. On the other hand, too few classification levels may result in overclassification and excessive protection requirements. However, based on the belief that consolidation of classification levels is easier than their expansion, the illustrative classification system presented below establishes a rather detailed structure.

The following groups can be identified for the purpose of controlling access to and dissemination of *identified* personal information items and/or records:

1. The *subject* of the information, and those formally representing his interests (physician, psychiatrist, lawyer, accountant, guardian).
2. *Personnel, management, and users* of the databank system in an agency (public or private). There are two subclassifications here:
   a. *All* personnel and users;
   b. A set of *authorized* personnel and users (only this group or some subgroup of it will have authority to actually enter, modify or delete information in the records).
3. *Users in other agencies* and databank systems who have an established need-to-know.
4. *Agencies with subpoena power*, such as courts, grand juries, governmental investigative committees.

5. *General public.* Anyone who requests to see the information (with the possible exception of minors and citizens of foreign countries).

For each of the above, access to the *content* of the information item or record is the principal consideration. However, for the *subject* (but not necessarily his representatives) and the general public there is another consideration--his *knowledge* of the existence of a record in a *specific* databank (data file). For the others it is assumed that it is not necessary to distinguish between knowledge of the record and access to its content. It is also assumed that, as suggested by the HEW committee [7], there will be no databanks whose very existence is secret.

A few remarks about the agencies with subpoena power. Using this power, such agencies can demand access to any record or set of records which they consider important for their investigations, and which are not provided privileged status by federal or state law [42]. Among the latter are the U.S. Census data and certain medical and psychiatric records [7]. In general, patient-physician and client-lawyer communications are held immune from subpoena. Other information that is granted statutory protection in various states includes drug abuse, alcoholism, and venereal disease records; information on victims of sex crimes; adoption proceedings; and illegitimacy records [6].

However, personal information gathered for statistical and research purposes in social, political, and behavioral sciences, and in education and psychology, has no statutory protection and often the promises of confidentiality have no substance. Several recent cases have illustrated this danger [43,44]. Hence the question of whether or not statutory protection against subpoena is provided is an important classification dimension.

Other important considerations in classification include:

- The subject's right to access the content of his record; the proposed Code of Fair Information Practices [7] holds this an important right which must be upheld.

- The statutory requirements of the Freedom of Information Act that, except for certain specific cases, every agency shall "on request for identifiable records" make these "promptly available to any person" [10,45]; exempted are "personnel and medical files and similar files the disclosure of which would constitute a clearly unwarranted invasion of personal privacy," and investigative files for law-enforcement purposes.
- Other personal data that have statutory limitations on their dissemination, but are not protected against subpoena.

An illustrative classification system based on this approach includes nine categories. Table 5 defines the categories and provides examples. Actual assignment of classification levels to specific information items depends, however, on the circumstances and is difficult to treat in general terms in this report.

Economic Value

A rational data security policy in a computerized information system would require that expenditures for data security be related to the *value* of protected information as expressed in some economic terms such as dollars, and to the expected *threats* against this value. Details of these relationships are discussed in Sec. III.

Value to the Subject. The economic value to an individual of his personal information stored in a databank system depends on the circumstances. There are three distinct situations:

- The value is in the *existence* of appropriate personal information in a particular databank and there are no special concerns about restricting its dissemination (e.g., in a databank used for disbursing or accruing economic benefits, or for handling the individual's assets).

Table 5

CLASSIFICATION OF IDENTIFIED PERSONAL INFORMATION ITEMS AND RECORDS

Classification	Subject[a]		Databank Users		Other Users[b]	Agency With Subpoena Power	General Public	Examples of Information That May be so Classified
	K	A	Authorized	All				
Category AS (public by statute)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Property tax rosters, marriage licenses, automobile registration.
Category A (public)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Membership lists in organizations, telephone books.
Category B (limited official)	Yes	Yes	Yes	Yes	Yes	Yes	No	Military personnel records.
Category C (restricted)	Yes	Yes	Yes	Yes	No	Yes	No	Criminal history records.
Category D (confidential)	Yes	Yes	Yes	No	No	Yes	No	Salary information.
Category DS (confidential by statute)	Yes	Yes	Yes	No	No	Yes	No	Adoption records. Juvenile crime records.
Category ES (privileged by statute)	Yes	Yes	Yes	No	No	No	No	Identified social sciences research data.
Category F (sensitive)	Yes	No	Yes	No	No	Yes	No	Psychiatric examination records.
Category G (secret by statute)	No	No	Yes	No	No	No	No	Organized crime investigation records.

[a] K = knows about the existence of the record; A = has access to the content of the record.

[b] With appropriate need to know.

- The value is in the existence of the personal information in a databank *and* in limiting its dissemination (e.g., salary information, medical and psychiatric records).
- The value is in the *nonexistence* of the information in a databank (e.g., criminal history records, and other information which limits the benefits an individual might otherwise receive, or limits his activities).

In all three cases it is in the individual's interest that the information not be altered in a manner that is unfavorable to him. In the first two cases this includes a total removal of his record from the databank. The direct economic losses that may result from such lapses of data integrity procedures may be removal from some benefit program or loss of assets. The indirect economic loss includes expenses required to straighten out the situation, or lost economic opportunities.

Failures to restrict the dissemination of sensitive information about an individual may also lead to consequences that are unfavorable to the individual. Indirect economic losses may result if the disclosed information could lead to:

- Loss of earnings or potential earnings.
- Victimization by extortion, threatening disclosure of sensitive information.
- Victimization by fraudulent schemes which are based on obtaining sensitive information.

A quantitative assessment of the economic losses due to unauthorized disclosure of personal information may require collection of the following data:

1. The subject's earnings at risk (both the dollar amount and relative magnitude as compared with the subject's total assets).

2. For each sensitive personal information item:
   a. Probability that its disclosure will lead to loss of earnings.
   b. The expected time duration of the loss and the fraction of earnings lost.
   c. The amount the subject is willing to pay to prevent its disclosure (i.e., victimization through extortion).
   d. Probability that its disclosure will lead to subject's victimization through fraud.
   e. The subject's assets at risk to victimization by fraud.
   f. Subject's evaluation of the economic losses that may result from unauthorized disclosure.
3. For sensitive information that may affect the individual's reputation, family relations, or self-respect, collection of statistical data on damages awarded by courts.

Sample surveys of the databank subjects and employers may be the best strategy for obtaining the data listed above. Other sources of information include crime statistics on extortions and "confidence game" frauds, and hiring and dismissal criteria of a sample population of employers. As an example of the latter, the following are among the items considered in appointments to federal positions [46]: "Any criminal, infamous, dishonest, immoral, or notoriously disgraceful conduct, habitual use of intoxicants to excess, drug addiction, or sexual perversion," and "Any facts which furnish reason to believe that the individual may be subjected to coercion, influence, or pressure that may cause him to act contrary to the interest of national security." Similar considerations are likely to apply in nongovernmental employment, except that "interest of national security" is replaced by "interest of the company." For example, one list of personal information that "indicates instability" of employees includes information on unpaid bills, drinking, marital problems, uncontrolled credit buying, habitual absenteeism, frequent job switching, and too many moves [47].

Some information on the value of information in victimization cases through extortion or fraud (confidence schemes) or damages awarded in character defamation cases is available from criminal statistics, reports, or literature. For example, descriptions of so-called bunko schemes show that several thousand dollars are involved in an average case [48]. Statistics in Table 6 present California statewide statistics on reported bunko cases of the "pigeon drop" and "bank examiner" variety.* Typically, these frauds are perpetrated on victims that the criminal meets by chance. However, since prior availability of personal information on prospective victim (as well as in the selection process itself) cannot but enhance the success of the fraud, statistics on these types should represent lower bounds for economic losses that may ensue when sensitive personal information falls into the hands of bunko artists.

Table 6

COMBINED STATISTICS ON "PIGEON DROP" AND "BANK EXAMINER" BUNKO FRAUDS IN CALIFORNIA

Year	Number of Cases Reported[a]	Total Loss ($)	Average Loss ($)	Range ($)
1968	162	418,000	2,580	31-15,000
1969	194	508,000	2,620	65-32,600
1970	219	518,000	2,360	12-18,000
1971	249	532,000	2,120	5-10,000
1972	266[b]	621,000	2,330	25-17,000

[a]It is estimated that the reported cases represent no more than 10 percent of all cases.

[b]Excludes one case involving more than $1 million in securities.

Damages awarded in character defamation cases have ranged from 6 cents to over a million dollars [49]. More meaningful statistics should be obtainable from criminal statistics reports.

*Personal communication by K. Kashiwaba, Organized Crime and Criminal Intelligence Branch, California Department of Justice, Sacramento, California.

A somewhat different view of the value of privacy can be obtained by examining the case of unlisted telephone numbers. For example, a recent survey shows that in large cities as many as 31 percent of residential telephone numbers are unlisted [50]: Los Angeles area, 31 percent; New York City, 21; and Washington, D.C., 18.7.

While these percentages are indicative of the desire for privacy, the expenditures involved are too small to be meaningful in assessing the economic value people place on privacy. For example, in New York the additional monthly charge for an unlisted number is 85 cents a month; in Los Angeles, 15 cents.

Value to the Custodian and Users. The databank custodian is responsible to the controller, directly or through the agency's management structure, for maintaining data integrity and security, and for implementing the various features for privacy enhancement that may be required by statutes. Hence, a substantial loss in data integrity, or unauthorized disclosure of data, is likely to result in penalties or sanctions against the custodian, the agency, and the users in addition to costs incurred in restoration of the data or the protective system. Among these may be the following:

- The legal liability of the custodian for damages incurred by the subjects. Legislation now pending in the Congress [50] and in various state legislatures, for example in California [51], asserts the liability of the databank custodian "even if the error of the damage has not arisen through any act or omission of his own." Class action suits against databanks are to be permitted, but the custodian's liability is to be no more than $50,000 or 2 percent of the assets, whichever is greater.*
- Penalties for deliberate or negligent violation of integrity or nondisclosure requirements are also specified in the pending legislation. For example,

*Proposed in House Bill, H.R. 10610, by Representative Koch et al., October 1, 1973.

fines up to $3000 or imprisonment of not more than one year are proposed.

- Restrictions may be placed on the databank operation in the form of reduced budgets, operational constraints, and stricter control.
- Insurance costs may increase. Insurance is available against libel and defamation of character suits as well as against loss of equipment and files [53].

Among other costs that may result from the loss of data integrity are the following:

- Cost of recreating data file(s). If proper precautions have been taken, backup files have been established and are kept current. Even then, considerable cost may be involved in reconstructing the correct current data files from the last backup files and transactions that have occurred since then [54].
- Cost of failure to properly operate the databank system until data files are reconstructed. In certain administrative databanks it is imperative that benefit or payroll checks be on time and correct. In industry, a delay in paychecks may cause a costly work stoppage. Interruption of the development of other data-processing applications when personnel are diverted to databank reconstruction may be costly.

The value of personal information to the authorized users is, likewise, essentially indirect. Users are also liable for penalties if they cause unauthorized disclosures, and would be constrained in their operations when data integrity is lost.

The trend toward setting specific penalties in the pending data protection bills provides a framework for estimating the (negative) value of personal information to the databank custodian and users. For example, given an estimated or empirically derived threat environment, the expected

number of attempts to violate data security can be calculated and used to estimate the corresponding penalties. This amount, then, would represent the economic "value" of personal information.

Value to Intruders. A *rational intruder* was previously defined as one who tries to gain access to a databank system in order to realize *economic profit*. There has always existed a flourishing market for trade secrets, marketing information, and customer lists that are acquired by intruders in industrial espionage operations. Their value has amounted to millions of dollars annually [35,36].

The value of personal information to a rational intruder is more difficult to assess. Two categories can be distinguished: (1) information on specific individuals, and (2) information on classes of individuals. In both cases, an intruder may act either as the end-user or as a middleman. In the latter case, the end-use may actually be "irrational"--one where gain is emotional or political rather than economic. The leaking of Senator Eagleton's psychiatric treatment and traffic violations data to the press during the 1972 presidential election campaign illustrates the latter.

The value of information on specific individuals may vary from next to nothing to thousands of dollars, depending on the prominence of the individual, the nature of information, and the susceptibility of the individual to extortion, political smear, or litigation, and his willingness to pay. Here the value to the intruder coincides with the value to the individual, as discussed previously. However, this value is slowly eroding as society's attitudes become more liberal.

A different market exists for modification or erasure of information in specific databank systems. For example, complete removal of his files from a police system is the dream of every criminal. Financial obligations could, likewise, be solved by erasure or modification of the appropriate data files, and a person's wealth could be increased instantaneously by manipulating his account balance. Insertion of bogus accounts in benefit distribution systems and even in industrial payrolls is another potentially rewarding intrusion activity. Any quantitative assessment of the value to the intruders from activity of this type requires an empirical "market analysis" and gathering of statistics.

Qualitatively, however, the value should be proportional to the sensitivity level of the information (as shown in Table 4).

Another type of intrusion involves attempts to use the high-speed search and correlation capabilities of the databank's computer (or a computer at the intruder's disposal elsewhere) to compile "mailing lists" of persons with specific characteristics and use these (or sell them for use) for various activities detrimental to the listed persons. Among these are the bunko schemes discussed previously (see Table 6), sale of quack cures, and other not necessarily illicit sales activities which capitalize on some weakness of the target population that is revealed by the sensitive information used for compiling the list.

A brief examination of the mailing-list industry may provide background data useful in estimating the market value of the sensitive mailing lists. The direct-mail advertising business in 1970 had a volume of over $3 billion which generated an estimated sales volume of $30 billion [7]. The Direct Mail Advertising Association, Inc. has some 1600 members [8]. The largest among these is L. R. Polk & Company which had sales of $65 million in 1972. Polk maintains a databank of over 200 million names of organizations and individuals and sells over 10,000 different lists [57]. Among these is a Household Census list on 24 million families which contains over a dozen information items.

The prices of the lists are in the range of $25 to $50 per thousand names, where the price depends on the length of the list more than the selectivity. Custom-made lists go as high as $70 per thousand names. Federal and state governments also sell a large variety of lists, at nominal prices, presumably as part of the Freedom of Information Act obligation to make public records available [58].

No data are available on a possible market in nonstandard lists of sensitive information (i.e., those prepared from confidential files). Lists of divorced persons, unwed mothers, felons, etc., are probably available, but not from the standard mailing-list firms. Rather, they are likely to be compiled by persons at agencies where the information is available. The price depends solely on who is contacted to produce the list, his cost and his estimate of "fair return" in view of his need, involved risks, and the buyer's willingness to pay. No statistics are available on the prices of such lists.

THREATS TO DATABANKS

Threats to data privacy, confidentiality, and security in a personal information databank system may arise from *all* components of a databank system. For example, without the consent of the subjects, the controller may change disclosure rules; the custodian, collector, or users may disregard confidentiality procedures or use data for purposes not originally specified; or the databank personnel, users, or even the subjects themselves may attempt to gain unauthorized access.

Although such treats have been extensively discussed in the literature [1-9,59], it is useful to establish a threat taxonomy, discuss the threat motives, and identify the relevant threat characteristics. The following classes of threats to personal databank systems can be identified:

- *"Legislative" threat.* Modification by the controller of the existing disclosure regulations (e.g., to permit inter-databank linkages previously prohibited). Seizure of the databank by some agency through the use of its legal powers (e.g., by subpoena).
- *"Executive" threat.* Modification of the disclosure regulations by the custodian, collector, or a user agency, independently or in some combination, but with the tacit approval of the controller.
- *Subversive threat.* Deliberate, unauthorized violation of disclosure regulations by the personnel of the custodian, collector, or user agencies who are normally authorized access to the sensitive data files.
- *Intrusion.* Unauthorized, surreptitious penetration into databank system and protected data files by unauthorized persons through technical means (e.g., computer terminals, communication links, electromagnetic emanation, computer and computer software [59]).

- *Physical invasion.* Overt seizure and/or destruction of the databank facility and data files.
- *Accident or negligence.* Disclosure through malfunctioning of the equipment, software, or lack of appropriate data security provisions.
- *Lack of data integrity.* Storage, use, and dissemination of erroneous, incomplete, or out of context personal information.

While the last two are not deliberately planned, they are still threats to data privacy and security. The *motivations* for launching or causing these threats fall into two categories: the "principal" motivations of those whose actions instigate the threats, and the "middleman" motivations of their actual perpetrators. Of course, the two may coincide. The following are among the principal motivations:

- *Aggrandizement of power and control* over specific individuals, groups of individuals, or some part of the society. This could be regarded as one of the likely motives for a legislative threat, or an executive threat designed to increase the government control over individuals or groups considered unpatriotic or disruptive.
- *Quid pro quo*--informal cooperation between agencies that exists in any bureaucratic structure: an agency performs favors to other agencies in order to build up "credit" for future favors by the recipient agency [60,61]. Such cooperation has also been called the "information buddy" system [8].
- *Economic gain*, as discussed in the previous section, is a principal reason for rational intrusion and subversive threats.
- *"Purging" of records*--surreptitious, selective erasing of records or data items in the records of the individuals or organizations to avoid unfavorable decisions. This is a likely motivation for subversive threats and intrusion.

- *Disruption* of the normal operation of the databank system to sabotage the operation of the user agencies as parts of campaigns against the "establishment" or revenge for personal grievances.
- *Coercion* by superiors, or by outsiders. The leverage may be political, financial, or psychological (as in any extortion case). A reason for subversive and physical threats, and for clandestine intrusions.
- *Curiosity* is a possible motive, but not very likely nor malevolent if it is not associated with any of the other motives listed above.
- *Antisocial sentiments.* Desire to perpetrate revenge on or persecution of certain individuals or classes of individuals who can be located by data obtained from the databank. While wholesale persecution of a group implies that the instigator must be a psychopath and the motive is unlikely, nevertheless precedents exist--to wit, the Nazis were able to locate the Jewish population in Germany through the records of the German Census office [62].

The main motive of the "middleman" intruders who attempt to perpetrate the threats for their clients is economic gain.

Associated with each type of threat are, in addition to the motivation, the following considerations:

1. *Potential payoff*--an estimate of the gain to be expected if the threat can be successfully carried out. For rational intrusions the payoff is economic gain.
2. *Technical feasibility*--in view of the databank structure, operation, disclosure regulations, and security techniques, the question of whether perpetration of the threat is within the intruder's resources (funds, equipment, expertise, manpower, and time). Mainly applicable to the intrusion threat.

3. *Cost*--the estimated expenditures of resources required to perpetrate the threat.
4. *Risk*--an estimate of the probability of not succeeding, as well as the probability that additional costs may incur, and an estimate of these costs. Such costs include penalties and other sanctions.

Analysis of these factors is an essential step in the design of data security systems. Development of the necessary criteria for measuring the amount of protection, estimating the costs of their implementation as well as their negation, and the trade-off functions for optimizing countermeasures are discussed in Sec. III. Illustrative qualitative estimates of the threat characteristics for public databanks and private databank systems are given in Tables 7 and 8, respectively.

A body of empirical evidence on the realism of intrusion threats against computerized information systems has been gathered by D. B. Parker of the Stanford Research Institute [63,64]. His statistics show that between 1964 and July 1973, there had been 65 successfully perpetrated cases of "computer crime," with total losses of $90.5 million (excluding the $2 billion estimated loss to insurance companies of the alleged fraud at the Equity Funding Life Insurance Company). The average loss was $1.39 million. Although only twelve of these cases involve personal information on individuals, mostly copying by authorized employees for sale as mailing lists, these statistics nevertheless illustrate the existence of a "threat population."

Table 7

THREATS TO PUBLIC DATABANKS

Characteristics	Threats			
	Legislative	Executive	Subversive	Intrusion
Motivation	Power and Control Quid pro quo	Power and Control Quid pro quo Coercion	Economic gain Coercion Antisocial Curiosity	Economic gain Purging Disruption Antisocial
Payoff	Intangible: power political good will	Intangible: power advancement good will	Economic gain Intangible: satisfaction	Economic gain Intangible: satisfaction immunity
Technical feasibility	Depends on the level of government: increasingly higher at lower levels of government	Feasible in most of the agencies	Generally quite feasible, may require collusion	Depends on the databank system but generally feasible
Cost	Increasingly higher at higher levels of government: requires much persuasion and lobbying	Moderate	Requires subverting personnel; moderate to high	Depends on the security techniques; presently not very high
Risk	Not high but may lead to political difficulties	Censure by controller Dismissal Penalties	Censure Dismissal Penalties	Penalties

Table 8

THREATS TO PRIVATE DATABANKS

Characteristics	Threats			
	Legislative[a]	Executive	Subversive	Intrusion
Motivation	Power and Control Quid pro quo Economic gain	Quid pro quo Economic gain	Economic gain Coercion Antisocial Curiosity	Economic gain Antisocial Curiosity Disruption
Payoff	Good will Economic gain	Good will Economic gain	Economic gain Satisfaction	Economic gain Satisfaction
Technical feasibility	No problem	High	Requires subverting personnel, collusion	Depends on databank system; generally quite feasible
Cost	Low	Low	Moderate	Depends on security techniques; presently moderate
Risk	Almost none	Censure Dismissal	Censure Dismissal Penalties	Penalties

[a]The controllers of private databanks are private agencies.

III. A MODEL OF PROTECTOR-INTRUDER STRATEGIES

PROTECTION POLICIES

Given a databank system that stores and processes sensitive personal information, a data security system must be designed and implemented by the databank protector--an organization within the databank system--to reflect the data protection policies of the controller and custodian agencies. Such policies may be classified as either dogmatic or rational.

A *dogmatic* data protection policy requires that absolute protection of the privacy of the subjects and security of the data should be provided "at any cost." Proponents of this policy can be found among social policy advocates and in national security circles. A *rational* data protection policy, on the other hand, specifies that protection should be provided only up to a certain level that is dictated by economic considerations of the situation. As is often the case, however, an *optimal* protection policy lies somewhere between the two extremes and takes into account the conflicting goals of the databank subjects, users, and custodians (as discussed in Sec. I), reflects the economics of the situation, and also contains elements of "compassion, fairness, and forgiveness" as sought by the subjects [65].

But even a pure dogmatic protection policy cannot escape the economic nature of the problem--neither an individual intruder nor a large agency possesses unlimited resources. On the other hand, if certain information is considered valuable enough, large resources may be committed for its acquisition legally, illegally, or unethically. Likewise, if the information is equally valuable to the protector, large investments may be made to provide adequate protection. It is clear, however, that the proper level of protection of given information should depend not only on the value of the information to the subjects and the protector, but also on its value to the potential intruders. Thus, the prudent investment decisions by the protector when implementing a rational protection policy would be:

- Not to commit large resources to protect information of little value to the potential intruders.
- Not to expend large resources to protect information whose release would not disturb the subjects, even if the information would be valuable to the potential intruders.
- To commit most of the resources to protect information that is valuable to the intruders, and whose acquisition by the intruders would be detrimental to the subjects, the custodian, or the users of the databank system.

In order to make the right decisions, the protector must be able to assess in some quantitative terms the value of protected information, its value to the intruders, the costs and effectiveness of various security options, and the resources available to the would-be intruders. These involve difficult problems that require further research [66,67]. However, one objective of the research reported here was to establish a framework for further research in the area and to clarify the involved relationships and variables. Toward this end, a protector-intruder interaction model was formulated and is described below.

AN INTERACTION MODEL*

In general, a large class of intrusions can be regarded as attempts to compile one or more "mailing lists," L, each containing N records. Each record in L can be assumed to have a market value v that is a function of a number of "market variables" and the sensitivity level s of the records. The total market value V of the list L is, then,

$$V = vN. \tag{1}$$

To compile a list L an intruder invests X units of resources for penetrating the system security features, gaining access to the desired data

*An abbreviated form of this model has been published previously [17].

files, and doing the required data-processing operations. The various components of X are examined in a later section.

If the intruder requires a minimum profit, rX, r > 0, then his maximum allowable expenditure of resources for compiling a list L of N records, each with value v, is

$$X = vN/(1 + r), \tag{2}$$

where, for simplicity, it is assumed that this intrusion is an isolated event that does not significantly benefit from previous intrusions, as may be the case when the security system is changed frequently. Further gains from selling copies of the list could be easily handled in the model. If the intrusion is not independent of previous ones, appropriately prorated resource expenditures could be used. Further, both the number of records obtained, N, and the expenditure of resources, X, are probabilistic quantities that must take into account the probability of failure to obtain the list, and the probability that the databank's security system may be able to cause additional costs over and above X.

Based on these considerations and available empirical information, estimates can be made of the resources that an intruder might expend for obtaining various lists of sensitive information at different rates of return.

To counter this and other intrusion threats, the databank protector expends Y units of resources for data security measures. For simplicity the various components of Y may be ignored for the present. Let I(X,Y) represent the intruder's information transfer function, i.e., the amount of information (number of records in list L) obtained by the intruder when he expends X amount of resources to overcome the Y amount invested by the protector. It is clear from the previous discussions of X and Y that I(X,Y) is not a simple function. However, some of its elementary properties are:

$I(0,Y) = I(X,\infty) = 0$, for $X,Y > 0$;

$I(X,0) = I(\infty,Y) = N_T$, for all records in the databank;

I(X,Y) is monotone nondecreasing in X and monotone nonincreasing in Y.

Let h(N) be the value function to the intruder of a list of N records of personal information, and c(N) be the cost function to the protector, users, and subjects of the loss of the same N records of information, occurring as a result of the intruder's acquisition of the information. Then, for given X and Y, the expected *net gain* of the intruder, g(X,Y), is

$$g(X,Y) = h[I(X,Y)] - X, \tag{3}$$

while the *net loss* to the protector and the subjects, f(X,Y), is

$$f(X,Y) = c[I(X,Y)] + Y, \tag{4}$$

assuming that for simplicity the entire expenditure Y can be charged to the loss of this intrusion.

Given the choice of compiling some combination of lists $L_1,\ldots,L_M$ whose records have unit market values $v_1,\ldots,v_M$, respectively, and sufficient information regarding the nature of the security system and the protector's expenditure, Y, an intruder may attempt to vary his investment, X, to maximize the expression (3). A rational protector would use his estimates of the value of protected information, the technical feasibility of threats, and the likely resources X available to the potential intruders to vary his protection expenditure, Y, to minimize the expression (4).

It follows that if the functions h, c, and I are suitably differentiable in a region containing X and Y, the selected values of X and Y will satisfy

$$\frac{h'[I(X,Y)]\ \partial I(X,Y)}{\partial X = 1}, \tag{5}$$

$$\frac{c'[I(X,Y)]\ \partial I(X,Y)}{\partial Y = -1}, \tag{6}$$

where the prime denotes differentiation. If one or more of the functions h, c, or I are not differentiable in the region containing (X,Y), the expressions (5) and (6) must be replaced by more complex conditions.

To illustrate the potential utility of this model, assume a hypothetical intruder's information transfer function $I(X,Y) = X^{1/2}/Y$, which in fact is rather unfavorable to the intruder, as $X = Y^2$. Then, if v is unit value of information to the intruder, and u is unit loss to the protector, then the corresponding intruder's gain function, g, and protector's loss function, f, are

$$g(X,Y) = (v\ X^{1/2}/Y) - X, \tag{7}$$

$$f(X,Y) = (u\ X^{1/2}/Y) + Y. \tag{8}$$

Taking the derivatives of these expressions as shown in Eqs. (5) and (6) yields X_Y and Y_X, the optimal values for X given Y, and for Y given X:

$$X_Y = \frac{v^2}{4Y^2}, \tag{9}$$

$$Y_X = u^{1/2}\ X^{1/4}. \tag{10}$$

For this function there also exist stable equilibrium points (X^*,Y^*). This represents the "optimal" situation in a sense that, given the described interaction of the intruder and the protector and no other costs or constraints, neither could further improve his payoff. The values of X and Y at the equilibrium point, as functions of v and u, are

$$X^* = \left(\frac{v^2}{4u}\right)^{2/3}, \tag{11}$$

$$Y^* = \left(\frac{vu}{2}\right)^{1/3}. \tag{12}$$

Figure 2 depicts the intruder's gain g as a function of his expenditures X for several values of the protector's expenditures Y, and the value parameters u and V. Figure 3 shows the protector's loss, and Fig. 4 the optimal choices of X and Y as functions of each other and the parameters u and v.

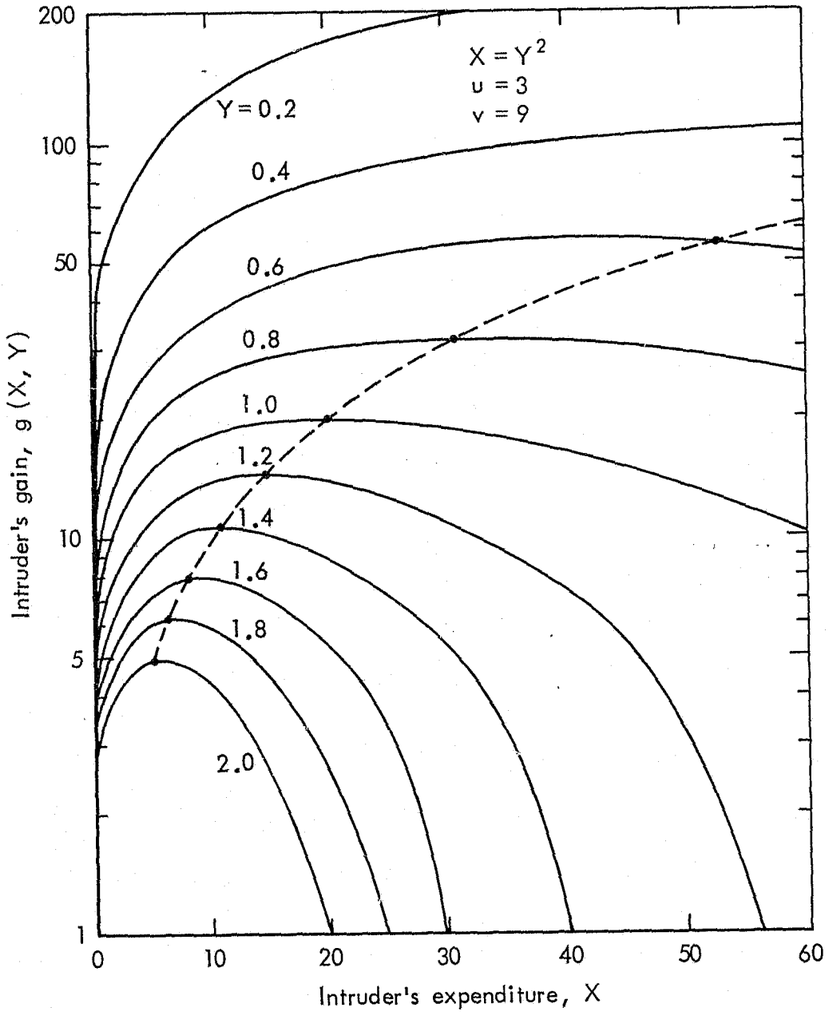

Fig. 2— Intruder's gain

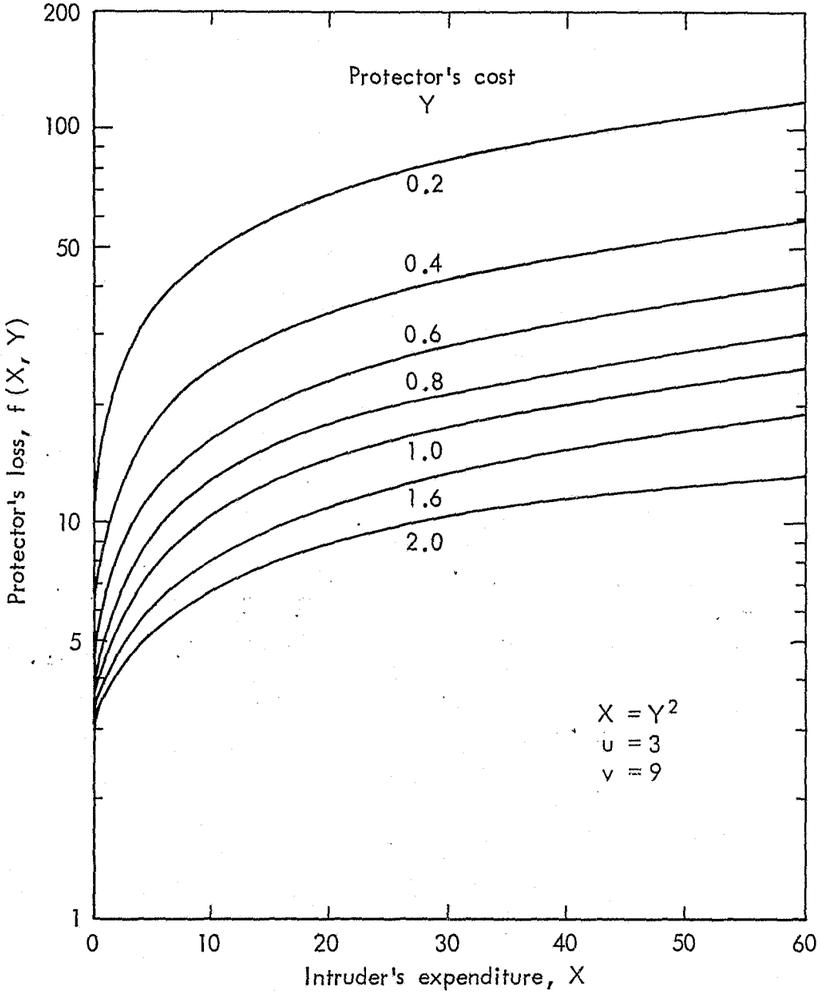

Fig. 3— Protector's loss

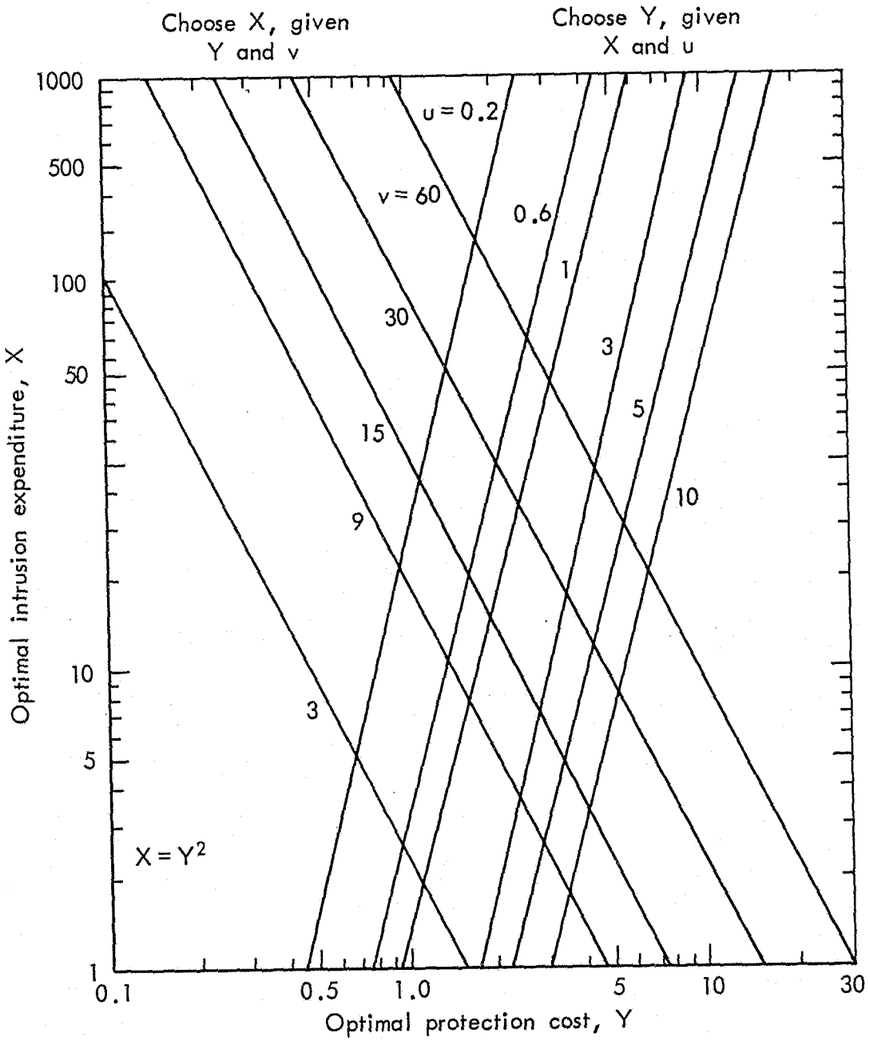

Fig. 4— Optimal choices of X and Y

Information such as shown in Figs. 2 and 3 allows the protector (and the intruder) to evaluate the marginal returns of incremental increases in expenditures. Figure 4 provides guidance in the choice of the optimal response to the other party's move. It must be emphasized again that these figures represent only a hypothetical situation presented to illustrate the potential utility of the model.

The model also underscores the need for obtaining analytical or empirical quantitative expressions for the following:

- *Value* of information to the intruder, the function h[I(X,Y)], and the *cost* of losing the same information to the protector, and function c[I(X,Y)].
- *Effectiveness* of the security system (i.e., the amount of security provided) as represented by the intruder's information transfer function, I(X,Y).
- *Costs* to intruder and protector, X and Y, as functions of security system design parameters.

The value of personal information to the intruder, subject, and protector was discussed qualitatively in Sec. II. Quantitative methods are still to be developed. The *effectiveness* and *cost* of the data security system and its components are discussed, also rather qualitatively, in Secs. IV and V.

MEASURES OF EFFECTIVENESS

The information transfer function I(X,Y) in the above model relates the expenditures of the intruder to obtain a certain quantity of information, X, with the protector's expenditure, Y. In theory, X may be in the form of a polynomial in Y, or a system of such polynomials along with various constraints. For example, the polynomial

$$X = a_n Y^n + \ldots + a_2 Y^2 + a_1 Y + a_0, \qquad (13)$$

where the coefficients, a_i, may represent the composite security effectiveness of the data security system, that is, the a_i may be functions

of the marginal effectiveness contributions, Δa_{ij}, of the individual, separable access control barriers, B_j, as well as their own security-boosting (or weakening) effects on each other.

Security Systems

In reality, the situation is not as simple as that. It is not clear which characteristics of the access control mechanisms or the data-processing system itself best reflect security effectiveness, what units of measurement should be used, and how they can be related to the "dollar cost," and how such measurements should be carried out in practice. It is very unlikely that these can be expressed in a simple, continuous functional form as in (13) above.

Any security system can be regarded as consisting of the following components [68-70]:

- A *passive subsystem* of access control mechanisms and security barriers that have certain levels of intrinsic resistance to intrusion and, thus, act to delay the intrusion process (e.g., a password system for restricting access, a lock or a door, or the use of a cryptographic technique).
- An *active subsystem* of surveillance, detection, and threat discrimination devices, to reinforce the passive subsystem and to assure that no intruder would have an indefinite amount of time for the penetration of the passive subsystem. Examples of devices in the active subsystem include physical intrusion detectors and alarms [71] and, in computer systems, software procedures for detecting attempts to discover a password through iterative trial and error techniques.
- *Operational security procedures* for system personnel and users. These specify the safekeeping procedures for the protected items, establish accountability, and establish penalties and other deterrence-enhancing provisions.

If the security system design is approached from this point of view, *time* required in the actual on-line penetration process is a crucial variable for both the protector and the intruder, and both can expend resources to constrain this variable for the other. The protector can expend resources to increase intrinsic delay of the passive subsystems as well as shorten the detection and threat discrimination time; the intruder can use more sophisticated penetration approaches and tools or do more off-line preparation for penetrating the passive subsystem and for avoiding detection, all of which require more resources. To illustrate the latter, during a recent penetration test of the operating system software of a large resource-sharing information system in an "adversary atmosphere" where the detection of the test was considered equivalent to the test's failure, the test team employed an approach reminiscent of a miniature military campaign [72]: reconnaissance, camouflage, diversions, saturation of defenses and an element of surprise. The programs used for penetration were encrypted, activities similar to penetration were launched at the same time, and these involved complex but otherwise purposeless data-processing tasks.

Time is presently used as a measure of effectiveness in rating protective containers, safes, and vaults for providing physical protection. For example, the federal General Services Agency (GSA) has established a security level classification scheme for file cabinets which includes the following [71]:

Class 1.	30	man-minutes against surreptitious entry
	10	man-minutes against forced entry
	1200	man-minutes against lock manipulation
Class 2.	20	man-minutes against surreptitious entry
	5	man-minutes against forced entry
	1200	man-hours against lock manipulation
Class 3.	20	man-minutes against surreptitious entry
	0	man-minutes against forced entry
	200	man-minutes against lock manipulation

As can be seen even in this relatively simple situation, the effectiveness measure is a vector of several components corresponding to the different threats that can be launched against the system. In a data security system, the threat domain of possible penetration tactics is considerably larger and the effectiveness measures more complex.

Access Control Barriers

As a first-order approximation, a data security system may be regarded as a network of access control barriers and associated controls. Figure 5 illustrates a portion of such a network containing a single barrier and its controls. The barrier governs the access from a user capability A to a more privileged capability B. Shown as dashed lines are the possible vulnerabilities of the barrier to penetration, circumvention, and attacks against its controls.

There are several classes of access control barriers. Within the computer system itself there are the following:

- *Technical barriers*--passive access control techniques implemented in hardware, such as identification systems, and surveillance, detection, and recording devices associated with the active subsystem.
- *Logic barriers*--access control techniques implemented in software, algorithms for encryption, and password generation and authentication.

Other security barriers in a databank system include:

- *Physical barriers*--doors, locks, protective housings, and shields on communication cables.
- *Procedural barriers*--requirements placed on the use of the system and data, such as strict controls over modification of the software, and taking special precautions when outside maintenance engineers are testing the system.

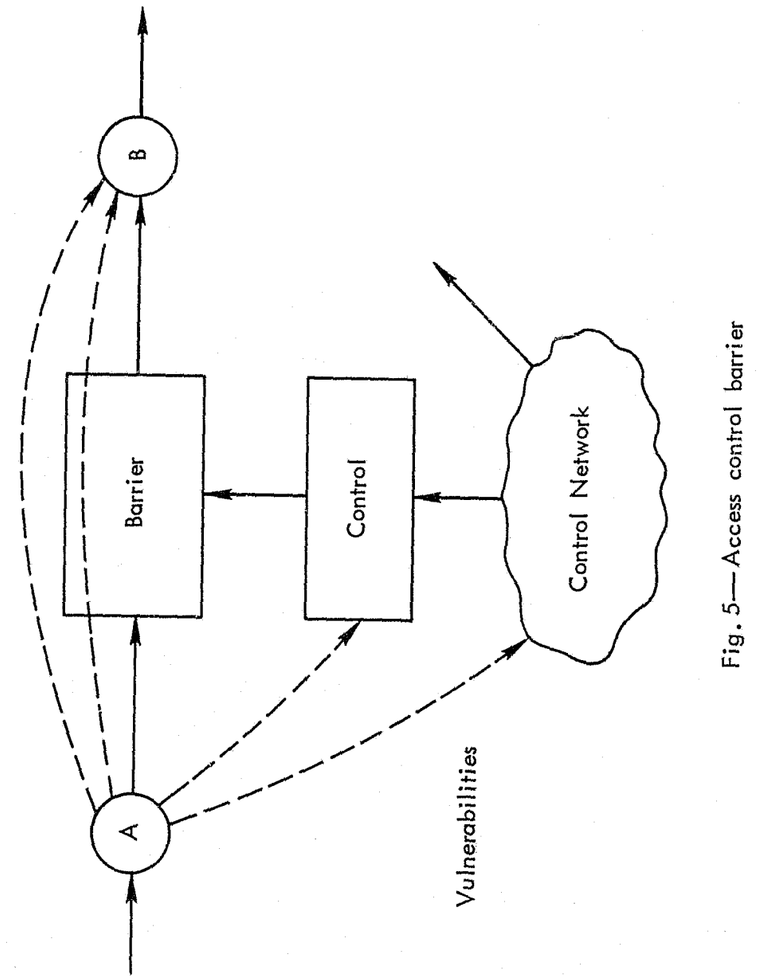

Fig. 5—Access control barrier

- *Human barriers*--security force personnel, and all others charged with improving the data security environment.
- *Psychological barriers*--disciplinary procedures and penalties for security violations or for attempted intrusions.

Each of the above contributes to security system effectiveness as well as to costs. At the same time, each of these, especially the human-oriented ones, may fail and thereby contribute to the system vulnerabilities. Hence each barrier can be characterized by a set of effectiveness measures, the associated costs, and reliability estimates.

Effectiveness

Since an access control barrier must allow passing of authorized users (or access requests by authorized programs) but stop unauthorized ones, its functioning depends on one or more "keys" that will open the corresponding "locks" in the barrier. Physical and technical barriers may actually use physical keys or cardkeys. Logic barriers use various codes. In both cases the functioning of the barrier depends on the fact that only authorized users are in possession of, or know, the key. Consequently, a measure of effectiveness of a barrier, B_i, has the following postulated components:

- The *probability of deception*, p_{Di}; the probability that the barrier accepts a forgery or a facsimile as the bona fide key. While p_{Di} can be made zero in logic barriers, it is likely to be nonzero in other types of barriers. p_D can be evaluated by experimental or analytical techniques.
- *Resistance to manipulation*, R_{Mi}; can be measured in terms of the average number of logical and arithmetic operations required for trial-and-error or analytic solution of the key [73]. R_{Mi}

can be converted into units of time when an assumption is made about the computational resources employed by the intruder.

- *Resistance to circumvention* or nullification, R_{Ci}; a measure of the inability of an intruder to capture or subvert the control of the barrier, or disable the barrier into an "open" position. The specifics of this component and its measurement require further research.
- *Detection and discrimination probability*, p_{Ai}, of the active subsystem, against attempts to manipulate or circumvent a barrier B_i. An effective detection and discrimination subsystem requires availability of information that characterizes these attempts.
- *Expected detection and discrimination time*, t_{Di}; required by the active subsystem for barrier B_i.

In addition to the above components of the effectiveness measurement vector, other components are specific to different types of barriers. For example, human guards exhibit a different kind of resistance to "manipulation" than hardware locks. As an illustration of the resistance to manipulation, consider an access control barrier operated by passwords that are comprised of five alphabetic characters. There are $26^5 \approx 1.2 \times 10^7$ possible passwords if arbitrary character combinations are used. The expected number of trials for finding the correct one is 6×10^6. The expected time depends on the time per trial. However, if the passwords are selected to have mnemonic capability (i.e., similarity to English), only 150,480 five-character combinations are available (those that contain at least two vowels in an appropriate arrangement [74]). To test all of these at a rate of ten per second would require slightly more than four hours. However, an intruder may be able to obtain the desired password with less effort by wiretapping the communication link to the appropriate terminal and recording the sign-in protocol. Other examples of security barriers' penetration resistance, especially for cryptographic techniques, are given in the literature [17,75,76].

Different classes of databanks, likewise, may have different security effectiveness measurement problems. Statistical databanks, for example, have the residual disclosure problem--how to avoid selecting aggregation cell sizes and individual characteristics that are aggregated in such a way as to prevent unique or near-unique identification of involved individuals by correlating various aggregations and applying available information.

Despite the described conceptual advances, practical measures of effectiveness of software-implemented access control techniques are still to be derived. It is also necessary to derive a calculus for combining the effectiveness, cost, and reliability contributions of the individual barriers. A number of caveats and considerations are also applicable to security effectiveness measurement. Among these are the following:

- Not all aspects of access control and data security are both quantifiable and practically measurable. Hence those aspects that can be measured do not necessarily characterize the entire security system (e.g., the number of locks on the front door does not measure the security of the entire house or even of the front door--it only measures the security against attempts to come through the door in the normal way).
- It is important to clearly establish the context and the scope in which measurements are made and used (e.g., there may be a tendency to concentrate on external intrusion, but as far as the management is concerned, incompetent operators, fire, flood, etc. are also damaging and may be more credible).
- Measurements may have time-dependent values (e.g., most of the operating systems are in continuous updating process; a barrier's resistance to intrusion may have a high threshold value, but reduce to zero after an intrusion has been accomplished; a system's personnel security consciousness and vigilance decreases as threats fail to materialize).
- It may not be possible or even desirable to aggregate the various types of measurements into a single numerical value. A single numerical value tends to hide the relative contributions of its components and thus may be misleading.
- Care must be taken in using statistics and empirically derived probabilistic measures as the sample spaces are likely to be extremely small. Indeed, a threat of a particular type may materialize and be attempted only once.

PROTECTION AND INTRUSION COSTS

In any protector-intruder interaction the protector is on the defensive, since the intruder has the initiative in the choice of the time and tactics. In addition, the protector cannot be certain whether the would-be intruder will be an outsider, a user, or someone among the database system's personnel. Consequently, the security mechanisms must be in operation at all times, access control tests must be applied to all information processing requests, and some of the system's resources--processing time and storage space--must be allocated to the security function whenever the databank system is operational.

Due to the basic asymmetry in operating modes, there are considerable differences in the nature of costs involved for protection and for intrusion.

Protection Costs

Costs for data protection can be categorized as the initial nonrecurring costs of establishing a data security system, and the recurring costs of operating the security system. Among the *initial costs* are the following:

- Security requirements analysis, specification, and design of the total security system.

- Capital expenditures for physical security included are improvements and purchase and installation of security hardware. Improvements and equipment may range from installation of card-key systems and closed-circuit TV to shielded transmission cables.
- Operating system vulnerability analysis and implementation of security-oriented features.
- Design and coding of security software modules for access control management, data encryption, password generation and distribution, audit trail recording, and so forth.
- Validation and verification testing of security-oriented software: operating system, system utilities, security routines, and application programs.
- Construction or modification of data files to include security-oriented information, such as sensitivity-level indication.
- Design and implementation of accountability procedures for collection and dissemination of sensitive information; control of demountable storage media; control of system and application programs, and design of backup systems.
- Generation of security-oriented personnel policies, security manuals, and security education programs.

From the point of view of the protector-intruder interaction model, these costs are a major part of the protector's security expenditure Y. The initial *hardware* expenditures represent the quality of the security system in controlling physical access to the computer facility, terminals, transmission links, demountable data storage units, and hard copy. These are particularly important against unaided external threats. Expenditures on the quality of *software* design, implementation, validation and testing are, however, the key to the security system's effectiveness against sophisticated intrusion launched internally with the help of subverted users or system personnel.

Operational costs of a data security system include the fixed direct costs--such as salaries and equipment rental--and the indirect costs that depend on the databank activity. Among the operational costs are:

- Reverification of the system's software and hardware after modifications or repairs.
- Processing-time requirements of the various data security software modules--identification, authentication, application of privacy transformations, recording of audit trails, security-oriented housekeeping operations, and so forth.
- Contribution to the main and secondary storage requirements by the data security software modules.
- Contributions to both the processing time and storage requirements of features by improved data integrity (such as check-sums or hash totals) and "tightened" design of all system's and application software (such as complete parameter testing).
- Contributions of continual security checking and testing.

Processing time requirements of data security software modules depend on their complexity and where in the system they are applied. Access control procedures may be (1) data independent, and (2) data dependent [77]. *Data independent* procedures are applied at the initial log-in time or at the initial file opening time and may involve a single password, a set of passwords, or an interactive dialogue with the user. *Data dependent* access control procedures are applied every time such a field is accessed in a record.

In general, an access control test consists of the following steps [78]:

1. Searching a list of access control records of N_u users, each having access to N_v different files, for a particular user's

request to access a particular file. The expected search time, $t_s = \frac{1}{2}N_u N_v t_{sr}$, where t_{sr} is the average time to testing an access control record against the submitted identification (also included in t_{sr} may be the prorated time of fetching the entire access control list from a secondary storage).

2. Authenticating the user's claimed identity and checking his access and processing request against his authorization. The time, t_a, depends on the number of fields that are checked, N_f, and the average testing time per field, t_f, $t_a = Nt_f$.

The time measures, t_{sr} and t_f, can be expressed as weighted sums of individual instruction execution times, where the weights are the numbers of instructions of each type that are used in the access control procedure. The overall time measures, t_s and t_a, can be converted into economic units (e.g., dollars) by applying appropriate cost factors. The storage requirement includes the programs and the access control list. Many different implementations of the list are possible. One type of relatively unsophisticated implementation of the access control list (where every field is a single word) requires approximately $N_u N_v (N_f + N_h)$ data words, where N_h is the number of fields in the access record used for linkages to the data files. Table 9 illustrates the processing time and storage requirements for a particular access control system [78]. The storage requirements for this example are:

Programs	17k words
Access control list	4k
Security modules	5k
Linkage modules	1.5k

Use of *privacy transformations* (encryption) to protect sensitive information in transit and in secondary storage requires encryption of every dataword when stored or transmitted, and decryption upon reception or fetching into the main store for processing. The transformation involved may be (1) a simple substitution; (2) a bit-by-bit modulo-2 addition of a key word, k_j, to a dataword, m_i, to produce its ciphertext

Table 9

EXAMPLES OF PROCESSING TIME REQUIREMENTS FOR ACCESS CONTROL PROCEDURES [78][a]

Access Control Type	Number of Fields Checked, N_f	Check Result	CPU Time[b] Per I/O (ms)		Percent Increase Over Benchmark	
			Read	Write	Read	Write
No test (benchmark)	--	--	3.22	2.40	--	--
Data independent	--	--	3.92	3.16	22	32
Data dependent	1	Passes	4.24	3.70	32	54
Data dependent	2	All pass	4.72	3.70	47	77
Data dependent	6	All pass	6.74	6.24	109	160
Data dependent	1	Fails[c]	4.38	3.76	36	57
Data dependent	2	All fail	5.08	4.40	58	83
Data dependent	6	All fail	7.72	6.82	140	184

[a] Executed on CDC 6400.

[b] Normal I/O time and access control check, $t_a + t_{sr}$.

[c] Includes time to erase failed field in record being delivered to user.

equivalent, e_{ij}: $e_{ij} = m_i + k_j$ (modulo 2); (3) a transposition f_r of the bit sequence of m_i, $e_{ik} = f_r(m_i)$; (4) a combination of these; or (5) a more sophisticated transformation [79,80].

For substitution transformations the processing time for encryption or decryption of one word depends on the time to apply the transformation, t_t, and the time to fetch or produce the key word, t_k. If the key period, P_k, is small, the keywords may be precomputed and stored in a protected area. If P_k is large, it is more economical to compute k_j in sets as the transformation process proceeds. In certain "infinite-key" transformations [81], a pseudo-random number generator is used. Table 10 provides information on encryption-processing time and storage space requirements for several types of transformations. It must be pointed out, however, that simple transformations can be readily "cracked" by using modern computational techniques [75].

Housekeeping operations for data security and integrity include the following:

- Erasing a storage area before allocating it to a new user. The processing time, t_z, depends on the size of the storage area to be erased, N_z, and on the operations involved. Typically this is a simple operation of writing a random number into the storage locations involved.
- Performing a complete parameter testing in subroutine calls and applications programs. The number of conditional transfers is likely to be increased over those that would be used in a nonsecurity environment. The added processing time and space requirements depend on the specifics of the routines involved.
- Testing for data integrity by computing check-sums for each individual record involves N_w modulo-p additions in each record of N_w words. The check-sums may be stored with the corresponding records and require one or more words of storage per record.

Table 10

ILLUSTRATIVE EXAMPLES OF ENCRYPTION/DECRYPTION PROCESSING-TIME REQUIREMENTS [82]

Encryption Algorithm	Time Increase Per Word Due to Encryption	
	Assembly Language	FORTRAN
Constant key-word: $e_i = m_i + K$	0	2.68
Long key period: $e_i = m_i + k_j$	1.73	4.03
Double key: $e_i = m_i + k^a_{j_1} + k^b_{j_2}$	2.64	6.60
Infinite-key: $e_i = m_i + k_i$	4.21	9.96

Recording of audit-trails and system transaction data is another security-oriented operation in a databank system. The objective is to provide information for ex post facto detection of security violations, establishing accountability for sensitive data. Maintenance of transaction data is also likely to become a legal requirement in personal information databank systems. For example, a recently introduced bill in the U.S. Senate [51] requires that a personal information databank system "maintain a complete and accurate record of every access to and use made of any data in the system, including the identity of all persons and organizations to which access has been given."

A typical audit-trail record generated at the time of user's log-in may include the following information [83]: User identification, terminal identification, account number, user's access control profile, time, and system status. Audit-trail records generated when files are opened, closed, or changed may include the following: User identification, user's access, control profile, file name, file access control profile, file form, and user's actions in the file.

Depending on the particulars of the databank involved, such as the size and activity level, considerable storage requirements can be made by recording audit-trail and transaction data. The processing time requirements would not be as great--essentially one additional I/O operation per transaction.

In general, it is estimated [83] that the access control and other data/security features tend to increase the overall processing time by 5 to 10 percent, the operating system memory code by 10 percent, and the main memory requirements of the operating system by 10 to 20 percent. Correspondingly, the resources available for productive computation are diminished and, if a certain peak processing load must be accommodated, a more capable processor may have to be acquired.

Finally, although the above discussion has illuminated the individual components and considerations of protection cost, a cohesive cost model for use in the protector-intruder interaction equations is still a topic for further research.

Intrusion Costs

Consider now an external intruder's expenditure of resources, X, in penetrating a particular databank system to obtain a set of N records that have a total value vN. As discussed previously, a rational intruder would require a profit rX. Such an intruder may be an agency with very large resources available or an individual with only limited capital available.

The intruder has a variety of options available for executing the penetration. These range from subverting a databank user or employee to a purely technological effort carried out from outside the databank system. Although it is likely that a technological penetration will be chosen only if the subversion costs or risks are too high, the following discussion is focused on this option.

A technological penetration has a variety of tactics available, from attempting to obtain valid passwords through wiretapping to cryptanalytic solution of transformed data files [59]. However, for successfully executing any of these the intruder must

- Obtain sufficient *information* about the databank system to determine the structure of the data files, the nature of the security system, the standard operating procedures, and the likely vulnerabilities.
- Formulate an *intrusion plan* that exploits the vulnerabilities within the cost and risk constraints.
- Make the necessary *preparations*--acquire equipment, prepare necessary programs, test, and practice.
- Gain *physical access* to some part of the databank system--terminal, communication links, demountable storage units, or any other part of the system that can be exploited.
- *Penetrate* into the computer system by deceiving, circumventing, or nullifying the access control system; acquire the desired information; and escape detection by the active security system sufficiently long to complete the action.

It is highly likely that most of the intruder's resources are expended in the first three of the above activities, especially in the preparation task. Indeed, his activities here are quite similar to those of the protector in the design of the security system--system vulnerability analysis and, if the penetration is attempted through programming, writing error-free security system penetration programs (these must be error-free in order to minimize detection by a threat monitoring system that may have been implemented). Therefore, as a first-order approximation, the protector's costs in this area could also provide a basis for estimating a part of the intrusion costs. One set of data points for these costs is provided by operating systems' vulnerability analysis experience by a security system research and development group [84]:

- Cost per finding a flaw $100-1000,
- Typically a group of 3 or 4 analysts,
- 3-6 months for a thorough analysis,
- 2-5 flaws per man-month,
- 1-2 terminal hours per man-day,
- Comprehensive systematic assessment,
- A great deal of interactive hypothesis generation and assessment.

Since this activity is performed with all possible information available, a real intruder may require larger expenditures. He must also be careful about rashness in attempting to exploit his flaw hypotheses--unsuccessful penetrations are likely to be detected and will alarm the protector. Clearly, one benefit of penetration testing studies is increased insight in the intrusion tactics and costs.

IV. PROTECTIVE SYSTEMS FOR DATABANKS

Previous sections of this report have discussed the structure and classification of databank systems, examined the sensitivity scales for personal information, and analyzed the various factors that affect the design and implementation of data security systems. The results of these investigations are now used to derive a set of "model" protective systems that provide different degrees of protection. The objective is to establish a frame of reference for specifying and designing *total* protection systems for specified levels of information sensitivity. However, the level of detail is necessarily low, since the specification of any working-level "model system" deserves a report of its own, as illustrated in the case of criminal justice databank systems [84,85].

ELEMENTS OF TOTAL PROTECTION

A total protective system for a personal information databank system must include procedural or technical provisions for:

1. *Subjects' rights safeguards*, as required by applicable statutes and regulation.
2. *Maintenance of data confidentiality* in the authorized collection, storage, use and dissemination of the raw or processed, but identifiable personal information.
3. *Data security* against deliberate intrusion.
4. *Integrity management* as it applies to hardware, software, data, and personnel.
5. *Auditing, testing, and evaluation* of the performance and effectiveness of the above four, and their compliance with applicable laws and regulations.

The design of each of the components depends further on the purpose of the databank, the sensitivity levels of stored personal data, the structure of associated computer facility, and other factors.

Subjects' Rights Safeguards

The potential violation of a data subject's rights through the collection, storage, use, and dissemination of information about him is the most important consideration in personal information databank systems. Indeed, one major reason for other components of the total protection systems is to implement the necessary safeguards. The other reason for their implementation is, of course, to assure that society's interests, as represented by the purposes supported by the databank, are also safeguarded. Table 11 depicts the principal elements of the subject's rights safeguards, procedures that may be used for their implementation, and the likely technological and cost implications. The costs are chiefly in the form of additional storage requirements for records, the need for special software or hardware, and additional processing time. In some cases also, special personnel and computer terminal facilities may be needed.

The different structures of databank systems also have impact on the subjects' rights aspects of total protective systems. In *public* databanks, those operated at various levels of government, the pending federal and state legislation (for example, [51,52]) would establish essentially all the subjects' rights listed in Table 11; however, certain law enforcement intelligence databanks and databanks containing certain medical information are exempted. But until such legislation is passed, various subsets of the rights may be implemented and made available at the discretion of databank custodians or controllers. Any information declared public by statute (such as the Freedom of Information Act or Federal Records Act) is available for inspection by the subject or his legal representative.

In *private* databanks, except for those under the Fair Credit Reporting Act, providing for any of the subjects' rights is at the discretion of the custodian of the databank. The Fair Credit Reporting Act applies to all databanks that contain information that "is used or expected to be used, or collected in whole or in part, for the purpose of considering the consumer's eligibility for consumer credit, insurance, employment, or other authorized business purposes" [11]. Pending legislation [51,52] also applies to private personal information databanks.

Table 11

IMPLICATIONS OF SUBJECTS' RIGHTS REQUIREMENTS

Subjects' Rights	Safeguarding Procedures	Technical Implications for Databanks
Right to know (existence of a databank)	• Public notice • Databank index	---
Right to know (existence of a record on him in a databank)	• Notification when the individual a. Contributes data b. Gives permission to collect data	---
	• Automatic notification a. When record is set up b. Periodically, if so required by law or regulation c. Notification as part of normal correspondence (when bills, checks, or forms are mailed)	• Special printout • Data fields in the record indicating date and means of most recent notification • Special programs
	• Notification upon individual's request	• Special staff to handle requests • Special programs and data fields in the record
Right to inspect (his record in a form readily understood by the subject)	• Automatic mailing of a printout of the record along with notification existence • Mailed to the subject on request	• Requirements as for "automatic notification" • Conversion of coded data into descriptions; existence of coding tables, conversion programs
	• On request by the subject at the databank facility	• Special personnel and display facility • Other requirements as above, except on-line conversion of record into understandable form
Right to challenge (the accuracy, relevance and completeness of the record, point out its possible obsolescence, and offer amendments)	• Providing special forms for suggesting the amendments, and a review board • Establishing purging policies and procedures • Implementation of data integrity management systems • Inclusion of rebuttal to all future use of the record	• Special data fields in the record for comments and rebuttals or linkages to these, and dates of time-dependent data • Transaction log for tracing distribution of erroneous information, and tracking data derived on the basis of erroneous entries
Assurance of compliance (the use and disclosure of the data for stated purposes of the databank only)	• Implementation of techniques to maintain confidentiality • Implementation of data security system • Implementation of auditing systems • Procedures for proving that all the rights requirements are being followed	• Special hardware, software, and operational procedures • Special transaction logs showing what use is made of the data, who used, etc. • Software to assure that only specific data be made available for specific decisions or actions

Most of the personal information stored in *administrative* databank systems is provided by the individuals themselves, and as a normal business practice, they may ask periodically to review its accuracy and provide additional pertinent facts. However, individual records may also contain evaluative information which the individual may not know about or have access to. This information is used to make decisions about eligibility, advancement, administrative actions, and the like, in many cases without notifying the individual of the action. For example, governmental benefit programs are involved in a great deal of investigative information-gathering as a part of their charter of assuring that only eligible individuals receive the benefits. Essentially none of the subject's rights are provided for in these situations.

Intelligence databanks may have a compelling reason for denying all subjects' rights if they are overridden by the interests of the society, for example in investigations of organized crime. In all cases, however, the existence of the databank should not be secret [7], although this has turned out to be the case in recent situations of dissident investigations [3].

The situation is somewhat different in *statistical* databanks where information is not used for making decisions on individuals. Except for the U.S. Census, information is provided by the subject voluntarily, or is derived from information in administrative databank systems. In general, the only serious subjects'-right interest here is knowledge of the existence of his records and maintenance of confidentiality and data security to avoid disclosure of any sensitive, identifiable information. If there is a threat to violation of confidentiality through some compulsory legal process, he should be notified of this [7] and be given an opportunity to take legal action.

In *centralized*, *integrated* databank systems that maintain personal records for multiple administrative purposes, there is a possibility that the Principle of Least Privilege is not enforced, i.e., more information than necessary may be provided for specific decisions by simply making the entire record available. This would be a noncompliance with the purposes of original data collection, and a violation of subjects' rights.

The other databank classification dimensions (see Table 1) have no significant impacts on the subject's rights component of the protective system.

Maintenance of Confidentiality

As pointed out previously, a variety of restrictions may be placed on the authorized use and dissemination of personal information. The present practice is to distinguish between (1) data that are given confidential or privileged standing by a statute, (2) data that are promised confidential treatment by the databank custodian, but could be obtained by others through some compulsory legal process, (3) data that are publicly available, and (4) data that *must* be made publicly available by statute. A more detailed classification has been suggested in Tables 4 and 5.

Several principles that can be applied for maintaining confidentiality are listed in Table 12. Most of these are intended for databanks that do not have statutory confidentiality but, nevertheless, collect sensitive personal information. Typical among these are statistical databanks maintained for social, behavioral, and political sciences research [86].

The different databank structures have implications on maintaining confidentiality in a way similar to their effects on subjects' rights. In general, those databanks that have *integrated* records, or that are *shared* by agencies, have more stringent requirements for implementing and enforcing confidentiality procedures.

Data Security

The components of data security systems were discussed in Sec. III, and details of these systems have been discussed extensively in the literature [53,58,82,83,88-91]. Table 13 presents a list of security principles, procedures for their realization, and associated technical implications in databank systems.

The structure of the databank system's computer facility, the number of people that gain access to it, the data-processing and information retrieval capabilities that are provided to the users, and the control

Table 12

PRINCIPLES FOR MAINTAINING CONFIDENTIALITY

Confidentiality Principle	Procedures	Technical Implications for Databanks
Reduce exposure	• Collection of only those personal data that are absolutely essential for databank purposes: application of the "principle of least privilege."	• Reduction of processing time and storage requirements. • Collection of additional data from the subjects if a set of items not previously collected becomes necessary.
Increase anonymity	• Removal of identifying information totally, or replacing with code numbers (that are provided special protection).	• Creation of linkage indices between identifying information and assigned codes [9]. • Special processing for information updating.
Control access	• Application of the Principle of Least Privilege to authorizing access to the data--only those users with definite need-to-know requirement. • Application of data security techniques.	• Implementation of access control techniques--special passwords and software for their implementation, hardware, software, and procedures for data security.
Establish accountability	• Implementation of regulations where those authorized access are responsible for the confidentiality of data they retrieve from the system.	• Marking of sensitivity levels on hard copy form of data. • Including sensitivity level codes with data records. • Keeping accountability logs of all accesses, processing, retrievals, output.
Reduce sensitivity	• Encoding of sensitive items in ways such that the encoded data still "makes sense." • Application of transforms to "mix" data in controlled but irreversible fashion [79] (statistical databanks). • "Inoculate" sensitive data with random errors in an irreversible but controlled way [28,88] (statistical databanks).	• Special processing and program. • Possible need to gather the data again if no originals are kept, and different statistics are needed.
Demand compliance	• Establishment and enforcement of codes of ethnics for the users, and administrative regulations which include penalties for violations of the confidentiality safeguards. For statutory confidentiality such penalties are established in the statutes. • Implement auditing systems.	• Storage requirements, programs, and processing for monitoring all accesses and uses of confidential data; transaction logs as in the accountability system. Additional logs. • Special personnel for monitoring enforcement and investigating cases of noncompliance. Special auditors.

Table 13

PRINCIPLES FOR DATA SECURITY

Security Principle	Procedures	Technical Implications for Databanks
Defensive design	• Concentrate security functions into protectable units. • Limit intruder's capabilities by compartmented design. • Limit operating system modules' and security mechanisms' capabilities to only those absolutely necessary (apply Principle of Least Privilege). • Complete design testing and verification. • Logical completeness of all subsystems--designed to perform precisely as specified, and in no other way, under any possible conditions, including system errors, application program errors, and operator errors.	• Special security software "kernel." • Design of security system and a series of independent barriers; special software and hardware for each. • Additional software or hardware modules for providing services, each highly specialized. • Software or hardware aids for testing and verification, expert personnel. • Additional instructions for complete parameter check-in, handling of asynchronous interrupts, and the like. Additional storage space. Much more effort for program design and implementation, and in proving logical completeness.
Complete control	• Exercise of total control over all user's actions in the system. • Identification of the users' terminals and computers. Authentication of the identity at all security barriers. • Monitoring the operation of the security system, detection of possible intrusion, discrimination of activity; use of entrapment [92]. • Implementation of auditing procedures.	• Software and hardware for identification and authentication routines, storage space for identifying keywords and passwords. • Instrumentation hardware and software for the active security system, and for entrapment modules. Storage for intrusion "signatures." Development of instrumentation techniques. • Auditing software and storage requirements.
Concealment	• Adoption of a "minimum visibility" posture for the entire databank system, especially its computer facility. • Use of cryptographic transformations in data files and in communication systems [76]. • If feasible, encrypted processing. • Maintaining information about the security system design confidential (but not depending on confidentiality in the security system design). • Erasing all storage areas before allocation to other users.	• Programs, computing-time overhead and storage space requirements for encryptions (as discussed previously). • Processing-time requirements for erasing storage areas.
Physical protection	• Providing protection to the computer system against overt attacks, as well as clandestine activities: wiretapping, eavesdropping, electromagnetic pickup [93,94].	• Special hardware and electronic systems.
Integrity	• Application of integrity management techniques.	• Special hardware, software, and storage requirements.

that can be exercised over the facility and personnel have an important effect on the databank's vulnerability to intrusions, and the requirements for security.

The *shared-dedicated* and *specific-integrated* databank classification dimensions are mainly addressing the exposure of the computer facility to user populations. In a dedicated computer facility, all users are databank agency personnel, under agency control. In a shared system, the computer facility, but not the databank, is used by the personnel of different, independent agencies. In the specific-integrated dimension, security requirements are higher. In a specific databank system, data are stored for only a limited class of applications. Both the computer and the databank used by the personnel of a single or several agencies have the same operating policies (such as in a regional criminal justice information system), while in an integrated databank system several agencies pool their data into the same databank, and the personnel of all cooperating agencies have access to the databank. Hence it may be necessary to exercise access controls over every field, or subsets of fields, in a record.

In an *off-line* computer facility, all processing is executed in a batch-processing mode: the necessary programs are entered or requested through the computer operator. Requests are queued by the system according to some priority rule and, in modern computers, processed in a multiprogramming fashion. Several programs may be in execution at the same time and share various resources--core storage, input-output processors, operating system modules. The principal protection requirement is prevention of inadvertent infringement of one program on the resources of another one. Programs may also be written with intrusion intent [95], and it may be necessary to use a full complement of security techniques.

In an *on-line* system the user can interact with his processing requests while they are in execution from terminals, within the computer facility, or at remote locations. The communication system used introduces new vulnerabilities--wiretapping and related intrusion techniques [59,96]. To prevent these, either physical protection must be provided to communications lines, or cryptographic techniques must be used.

Depending on the capability provided to the intruder at the terminal, he may be able to enhance his penetration chances by observing the interactions of his activities and the system defenses. Table 14, adapted from Ref. 97, illustrates the relationships between users' capabilities at a terminal and vulnerabilities of a system.

The *closed-open* classification dimension also relates to users' capabilities. In a closed, transaction-oriented system, the user can interact with the system only in a special language provided by the system, and cannot enter his own programs (from the terminal or in batch processing mode). This limits the possibilities for sophisticated penetration. An open system, however, allows the user to submit his own programs. Several degrees of capability are involved, as depicted in Table 14.

While it is intuitively clear that a dedicated, specific, off-line closed databank system has the least vulnerability, and a shared, general, on-line, open databank system has the most, it is difficult to rank databank systems in order of increasing vulnerability. However, Table 15 presents a ranking system that, also intuitively, seems reasonable.

Integrity Management

Integrity management in databank systems has two components--*data* integrity and integrity of the *protective system*. The basic concerns of data integrity are the assurance of data quality (relevance, completeness, accuracy) in the collection phase, and the subsequent maintenance of data quality in the databank system. Integrity management for the protective system is concerned with hardware, software, and personnel reliability, and prevention of tampering or subversion. Table 16 lists the basic considerations in integrity management, the procedures for their implementation, and the associated technical implications for databank systems.

The structure of the databank's computer facility has considerable implications on the integrity management of the protecting system:

- Computer networks and remote, on-line terminals introduce the communications system reliability problem.

Table 14

USER CAPABILITIES AND SECURITY RISKS[a]

Capability	Shared Resource	Exploitable or Created Vulnerability
A. Just watch	Display surface	Malfunction Bug terminal Introduce jamming device
B. Initiate program (as in A and limited controls) Manual probes	Operating system Application programs Data	Insufficient legality check Illegal sequencing Crash the system
C. Transaction only (as in B and capability to enter parameters)	Time Increased I/O bandwidth	Logic path complexity Data aggregation
D. Interpretive programming (as in C and capability to enter programming statements)	Limited pseudo-machine (the interpreter)	Higher-order complexity Ability to overload the processor or storage to deny these to others
E. Higher-order language programs (to be compiled)	Limited real machine (the compiler)	Break into the machine language code
F. Assembly or machine language program statements	Near-total system control Real addresses Real operation codes	Change operating system code Exploit design incompleteness
G. Machine language (in monitor state)	Total system control	No security features Modify operating system
H. Access to hardware	Total system control	Modify hardware or operating system software

[a]Adapted from Ref. 97. Presented in increasing order of risk.

Table 15

ILLUSTRATIVE SECURITY VULNERABILITY RANK-ORDERING OF DATABANK CLASSES

Classification Dimension	Weight	Type of Databank System[a]															
		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
Open	8									x	x	x	x	x	x	x	x
Closed	0	x	x	x	x	x	x	x	x								
On-line	4					x	x	x	x					x	x	x	x
Off-line	0	x	x	x	x					x	x	x	x				
Specific	2			x	x			x	x			x	x			x	x
Integrated	0	x	x			x	x			x	x			x	x		
Shared	1		x		x		x		x		x		x		x		x
Dedicated	0	x		x		x		x		x		x		x		x	
Vulnerability score (rank)		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

[a]Each different system configuration is identified by a capital letter for ease of reference.

- Shared and open databanks imply complex software and less control over personnel.
- Integrated databanks imply many users of the data, thus complicating the tasks of maintaining data integrity.

Audit, Validation, and Testing

Auditing has been defined as an independent, objective examination of the databank system, its use and, in particular, the adequacy, effectiveness, and compliance with the protective system. For example, an audit can detect vulnerabilities, risks, and erroneous data. A formal audit is usually performed by an agency outside the databank system, but internal auditing is also an important part of assuring effective protection. The basic "raw material" for an audit is system documentation and records of system transactions--the audit trail--which permit

Table 16

INTEGRITY MANAGEMENT TECHNIQUES

Principles	Procedures	Technical Implications in Databank Systems
Accuracy control	Use of self-checking numbers, check-sums, hash totals, and error-detecting codes [89].	Special programs; storage space and processing time requirements.
Quality control	Double-checking of initial correctness, completeness, and dependability; use of multiple sources; correctness verification by the data subject. Correcting data derived from incorrect or unreliable information.	Special personnel for these activities. Maintenance of transaction logs that show what data were derived.
Reliability	Providing hardware reliability through high quality design, redundancy techniques, and other failure-tolerant design techniques. • Fail-safe procedures for handling malfunctions. • Control of all software and hardware modifications, testing for inadvertent or deliberate unauthorized changes.	Special hardware and software for failure-tolerant operation, use of error detection and correction codes, automatic diagnostic routines, and the like.
Personnel integrity	• Personnel policies to obtain trustworthy personnel, policies to maintain security consciousness [89,98]. • Division of programming tasks and knowledge of the protective systems; control of console operators.	• Transactions logs, logs of all console operations. • Special personnel for educational programs, and for security-oriented management of programmers.
Backup and recovery	• Maintenance of periodically updated duplicates of data files at secure locations. • Maintenance of appropriate logs for updating backup files to current status.	• Special storage facilities, preparation of backup files, transaction logs. • Documentation of system software and data structures.

the auditor to determine accountability for sensitive data and changes to the protective system (e.g., access criteria, authorizations, passwords, etc.), the particular access control criteria in effect at a given time, and the agreement of the criteria and actual accesses allowed [99].

The transaction logs may need to record a substantial part or even all of the transactions that take place and the access control criteria that were applied. Among the information that may be required are the following: file actions, file corrections, updates of specified data items, console operator actions, changes in protective systems, and security violations.

Validation and testing are activities aimed at proving the effectiveness of the protective system. In the case of hardware, this involves testing the behavior of various hardware subsystems under unusual situations (such as the use of unassigned operation codes), the ease of access to communication links, the electromagnetic radiation from the system, and the like. Validation and testing *software* is, however, a much more difficult task. The major approaches here are:

- Program proving through formal methods--demonstration that a program performs exactly as desired and in no other way [100].
- Penetration testing of system software to discover flaws and vulnerabilities [72,83].

Currently both have shortcomings: at present, only small programs (about 100 lines of code) can be handled by program proving methods. The techniques are mostly manual and are subject to errors. Penetration tests, in turn, can show only that a particular test team, using a particular approach, can or cannot defeat the security system. However, methods of vulnerability analyses developed for penetration testing can also be used for more general analyses of operating system programs and can be applied to identify at least the simpler vulnerabilities.

The effects of different databank structures on auditing, and on validation and testing, are clear: the more complex the computer

facility (e.g., in on-line systems and computer networks), the greater the user's capabilities (e.g., in open systems); and the greater the user population (e.g., in integrated, shared systems), the more complex the associated programs are and the more difficult it is to keep track of activities within the system.

Table 17 presents a list of techniques associated with auditing, validation, and testing, and indicates their technical effects on databank systems.

"MODEL" PROTECTION SYSTEMS

This subsection outlines a set of representative protection systems for personal information databanks. The protection level in a databank system depends on the sensitivity of the stored data (see Tables 4 and 5), the structure of the databank system (as discussed in Sec. II and summarized in Table 1), and the threat environments (Tables 7 and 8). However, the number of different types of databank systems that can be constructed on the basis of these considerations is too large to permit individual analysis of each. Therefore, representative protection systems will be discussed only for three levels of protection, "low," "medium," and "high," as defined in Table 18.

Based on the discussion in this section and on Tables 11 through 17, a set of protection provisions is specified for each representative protection system. The specifications are illustrative rather than firm recommendations, in that they are based on the author's views of what is reasonable, rather than on thorough analyses. Indeed, at present such analyses are difficult to perform, and some are not even feasible. The purpose of describing such representative, "model" protective systems is to illustrate different protection levels that may be desired, and to provide a framework for real-world exercises of this sort. Tables 19 through 21 describe the "low," "medium," and "high" levels of protection, respectively.

Table 17

TECHNIQUES FOR AUDITING, VALIDATION, AND TESTING

Principle	Procedures	Technical Implication in Databank Systems
Traceability	• Audit trail logs. • Modular programs and procedures.	• Special software. • Special storage devices, commuting time for composing and writing records.
Balances	• Computation of balances of quantitative data. • Quantification of qualitative data for balancing.	• Computing time for accumulation of balances, storage space for results. • Special software routines.
Independence	• Establish separate internal auditing section. • Use separate auditing software.	• Special personnel; special software. • Periodic use of computer time for system examination, running of test cases.
Proven correctness	• Program design and coding using software engineering techniques. • Application of techniques of graph and logic theories; automatic program proving.	• Special personnel, software, and use of application programs.
Testing	• Analysis of software in machine language form, analysis of core dumps; formation of flaw hypothesis, testing, and iteration.	• Special personnel, software aids to analysis, simulation programs; "hand-on" test of the computer.

Table 18

DEFINITION OF PROTECTION LEVELS

Data Sensitivity Level (Table 4)	Data Classification (Table 5)	Protection Level
0, 1	A, AS	Low
2, 3	B, C, D, DS	Medium
4, 5	ES, F, G	High

Table 19

A "LOW" LEVEL PROTECTION SYSTEM

Protection Element	Protection Provisions
Subjects' rights	• All statutory provisions. • Use of data only for purpose that was stated when data was collected except with subject's consent which may be obtained when collected. • Notification of existence when subject contributes information. • Right to inspect records upon request; right to request correction of factual information. • Right to request destruction of information no longer used.
Confidentiality provisions	• All statutory provisions. • Need-to-know proof for disclosing information not public by statute.
Data security	• Databank may utilize a computing facility of any type. • Restricted access to computer facility. • Simple password for access to data from terminals or in off-line mode. • Normal procedures against physical damage to data files.
Integrity management	• Normal quality control for initial correctness of data. • Accuracy control on more important information, such as identification and information representing status in the system--use of hash totals or check-sums. • Normal hardware and software reliability provisions. • Backup of data files and recovery procedures.
Auditing, validation, and testing	• Auditing procedures as required by good business practices. • Normal validation and testing of hardware and software.

Table 20

A "MEDIUM" LEVEL PROTECTION SYSTEM

Protection Element	Protection Provisions
Subjects' rights	• All statutory provisions. • If not provided by statute, then: --Use of data only for purpose for which it was collected, unless other uses consented to by subject. --Notification of existence of record when interacting with subject. --Right to inspect record (except parts that are not accessible to subject by statute or subject's consent). --Right to correct factual information and challenge evaluative information. --Automatic notification of increased threat environment (e.g., subpoena) or compromise of data. --Right to withdraw voluntarily contributed information. --Right to request destruction of information no longer used.
Maintaining confidentiality	• All statutory provisions. • Establishment of regulations and procedures for handling data with different sensitivity levels; labeling of hard copy and file items. • Accountability provisions for sensitive data. • Reduction of identifiability by removal of identification from more sensitive data and use of special identification numbers. • Reduction of sensitivity by coding or using grosser descriptions; use of error inoculation technique in statistical databanks for special cases. • Reduction of exposure by careful examination of need for sensitive information.
Data security	• Use of "open" operating mode and "shared" computer facilities should be avoided when more sensitive data are involved; use of transaction-oriented operation as much as possible. • Full use of defensive design principles. • Physical access control to computer facility and software terminals that allow on-line programming. • Access control system for files, processing restrictions; authentication of identity; use of effective password and personal identification procedures and devices, especially in on-line, shared, or integrated systems. • Cryptographic transformation of more sensitive information in removable files. • Cryptographic protection for communication system for remote, on-line terminals that handle especially sensitive information. • Real-time threat monitoring when especially sensitive data are involved, and in shared or open databanks.
Integrity management	• Full use of data accuracy control for sensitive factual information--hash totals, check-sums, and error detecting codes (except in certain statistical databank systems). • Full data quality control (except in statistical databanks). • High-reliability hardware; full control over all modifications. • Clearances for personnel handling security system components, such as passwords, cryptographic keys, and authorization tables. • Controls on programming the operating system, security system software modules, and documentation. • Full provisions for backup and recovery. Frequent generation of backup files. Frequent reloading of operating system software from backup files to prevent modification. Appropriate documentation.
Auditing, validation, and testing	• Full traceability of transactions for more sensitive informations, especially in open, shared, and integrated databanks. Normal auditing provisions for others. • Transaction balancing for sensitive information in shared, open, or integrated systems. • Internal auditing of system security and accountability procedures and compliance with these. • Use of program-proving techniques, if feasible, for operating system and security modules of open systems; thorough testing for others. • Penetration testing (either actual or analytic) for open, shared, and integrated systems handling more sensitive data. • Full enforcement of all protection provisions with internal disciplinary action as legal procedures against violators.

Table 21

A "HIGH" LEVEL PROTECTION SYSTEM

Protection Element	Protection Provisions
Subjects' rights	• All statutory provisions. • If not provided by statute, then: --Automatic notification of existence of record when it is established (except when existence is to be withheld on basis of appropriate authority). --Right to inspect parts of record that are accessible to subject, right to require correction of factual information, and rebuttal of evaluative information. --Automatic notification of increased threat or compromise of data. --Right to withdraw voluntarily provided information (as in certain statistical databanks), or demand anonymity. --Automatic destruction of information no longer used, and notification of subjects thereof.
Maintaining confidentiality	• All statutory provisions. • Stringent regulations and procedures for handling information; labeling all hardcopy and records in files. • Stringent accountability procedures. • Separate storage of identifying information. • Fullest possible reduction of sensitivity in statistical databanks, use of error inoculation, aggregation, and irreversible transformations [78]. • Reduction of exposure in administrative databanks by applying stringent need-to-know test before collecting sensitive data.
Data security	• If feasible, use of manual systems for the most sensitive data; otherwise use of dedicated and closed compiler facilities only. Generalized databanks should be avoided. • Full use of defensive design principles. • Physical access control to all parts of the computer system, including remote terminals in on-line databank systems, and hardware cabinets. • Use of equipment against electromagnetic pickup and eavesdropping. • A plan of action in case a penetration is detected. • Access control software for files; processing restrictions; reliable identification and authentication system; use of highly effective passwords and codes (e.g., once-only passwords). • Use of cryptographic techniques for all removable files, and on-line files; decryption of the information only immediately before using. • Use of highly effective cryptographic systems in communication links in remotely accessible systems; if appears necessary, also in systems where terminals are in buildings other than the computer facility. • Full-scale real-time threat monitoring; use of entrapment principles. • Fullest use of hardware implementation of data security mechanisms.
Integrity management	• Full use of data accuracy and quality control for factual information in administrative databank systems (except in certain statistical systems). • Highly reliable hardware with fault-tolerant design for critical subsystems. • Fully cleared personnel and users and modification. • Tight controls on programming of operating system, security modules, and application programs. Tight controls over hardware modifications, repairs, maintenance and documentation. • Full provisions of backup and recovery. Equally stringent security requirements for handling backup system. Frequent loading of system software from backup files. Documentation. • Frequent testing of hardware and software for unauthorized modifications.
Auditing, validation, and testing	• Full traceability of transactions in administrative and intelligence systems. • Use of transaction balancing techniques in testing for modifications. • Frequent internal and external auditing by independent auditors. • Application of program proving techniques, if technically feasible, to all software. • Frequent, unannounced penetration tests against all components of the security system. • Full enforcement of all protection requirements and procedures with stringent internal disciplinary actions or legal procedures against violators.

V. CONCLUDING REMARKS

The problems of potential violations of citizens' rights through computerized personal information databank systems, and of finding technical solutions to protection of information in modern resource-sharing computer systems and networks continue to remain in the focus of political, societal, and technical concerns in the United States and other countries.

Formulation of the Code of Fair Information Practices by the HEW Committee [7], enactment of privacy protection legislation in Europe [12,14], and publication of several reports on privacy and databanks [8,16,28] have resulted in numerous legislative proposals in the U.S. Congress and in various state legislatures. The seriousness of privacy protection was further underscored by President Nixon when he stated:

> One part of the current problem is that as technology has increased the ability of government and private organizations to gather and disseminate information about individuals, the safeguards needed to protect the privacy of individuals and communications have not kept pace. Another part of the problem is that clear definitions and standards concerning the right of privacy have not been developed and agreed upon.
>
> I have therefore ordered an extensive Cabinet-level review--which will be undertaken this year--of both government and industry practices as they relate to the balances that must be struck between legitimate needs for information and the right of privacy, and of those measures--including appropriate legislation--that can be taken to ensure that these balances are properly struck.*

As recently suggested in the news media, the year 1974 may become known as the Year of Privacy.†

It is important, however, that the proposed legislative measures for safeguarding the rights of databank subjects be firmly based on technological realities, such that the enacted protection requirements

*State of the Union Message, January 30, 1974.

†"Capital Craze: Companies Full of Personal Data," *Los Angeles Times*, February 20, 1974.

can be implemented and complied with in practice. If this is not the case, then the adage, "the cure is worse than the disease," may indeed become valid: statutory privacy safeguards that cannot be implemented create only an illusion of protection.

Practical data privacy and security safeguards require a practical approach to protection needs. It is clear that not all personal information databank systems require the same level of protection. As discussed here, different items of personal information have different sensitivity levels and value, and different databank structures have different vulnerabilities. The protector-intruder interaction model described in Sec. III underscores design requirements for practical protection systems and identifies the variables involved: estimates of the value of protected personal information to the subjects, protector, and potential intruders; protection and intrusion costs; and probabilities of threats and their detection. If the relationships between these variables can be expressed in well-defined functional form, such relationships can be used to derive general guidelines for formulation of protection policies and making protection investments. However, specific measures of security effectiveness, and procedures for estimating costs must be derived before the protector-intruder interaction model, or any other similar approach, can be used as a practical tool for designing protection.

A total protective system for a personal information databank can be regarded as consisting of techniques and procedures for providing subjects' rights safeguards, confidentiality, data security, data and system integrity, and assurance that such procedures and techniques are being complied with. In Sec. IV each of these categories was examined in detail from the point of view of the basic principles involved and approaches to their realization. The use of this catalog of principles and techniques is illustrated by specifying the components for three types of protection systems. Similar analysis must be undertaken when planning protection systems for real databanks. The material in this report should provide a useful framework for such analyses.

The design and implementation of cost-effective protection systems for personal information databanks and resource-sharing computers in

general is still an unsolved problem. Further research and development is required in software and hardware techniques for access control, validation, and testing. Tools and techniques are needed for databank system's analysis for determining protection requirements. Methodologies must be developed for protection system synthesis and optimization. The focus of this report has been on the last two of these research areas--databank system analysis and protection system synthesis. It is hoped that the results obtained have contributed both to the clarification of the problem and to the formulation of the solutions.

Appendix A

SELECTED AMERICAN VALUES [37]

I. Self-oriented Values
   1. Personal "material" welfare (the right of life and the pursuit of happiness)
      a. Health (physical and mental well-being)
      b. Economic security and well-being
      c. Personal security
   2. Self-respect (the right to be treated as a person and as a member of good standing of the community; honor)
   3. Personal liberty (the right to endeavor to "shape one's own life")
      a. Freedom (from interference)
      b. Privacy
      c. Property rights
   4. Self-advancement and self-fulfillment (success, ambition; the "pursuit of happiness")
   5. Skill and prowess
      a. The intellectual virtues (intelligence, education, know-how)
      b. The physical virtues (strength, dexterity, endurance)
      c. The virtues of the will (strength of character, industriousness, fortitude, initiative, self-control, perseverance)
      d. Competence (pride in workmanship)
      e. Faith (believing in something, having a "sense of values")

II. Group-oriented Values
   1. Respectability (good repute, group acceptance, conformity)
   2. Rectitude and personal morality (honesty, fairness, trustworthiness, reliability)
   3. Reasonableness and rationality (objectiveness)
   4. The domestic virtues (love, pride in family, thrift, prudence)

5. The civic virtues (involvement, good citizenship, law-abidance)
6. Conscientiousness
   a. Devotion to family, duty
   b. Personal responsibility and accountability
   c. Devotion to principle
7. Friendship and friendliness
   a. Friendship proper
   b. Loyalty (to friends, associates)
   c. Friendliness, helpfulness, courteousness, fellowship
   d. Personal tolerance, patience
8. Service (devotion to well-being of others)
9. Generosity, charity
10. Idealism (hopefulness of human solution to human problems)
11. Recognition (getting due public credit for good points, success, status)
12. Forthrightness (frankness, sincerity, genuineness)
13. Fair play (being a "good sport")

III. Society-oriented Values

1. Social welfare, social "consciousness"
2. Equality (tolerance, fairness, civil rights)
3. Justice (legality, proper procedure, recourse)
4. Liberty (the "open society," various "freedoms")
5. Order (public order, "law and order")
6. Opportunity (the square deal for all)
7. Charity (help for the "underdog")

Appendix B

RAND PUBLICATIONS UNDER NSF GRANT GI-29943

1. Hunt, K., and R. Turn, *Privacy and Security in Databank Systems: An Annotated Bibliography, 1969-1973*, The Rand Corporation, R-1361-NSF (in process).

2. Johnson, S., *Certain Number-Theoretic Aspects of Access Control Passwords*, The Rand Corporation, R-1494-NSF (in process).

3. Reed, I. S., *The Application of Information Theory to Privacy in Databanks*, The Rand Corporation, R-1282, May 1973.

4. Reed, I. S., "Information Theory and Privacy in Databanks," *AFIPS Conference Proceedings*, Vol. 42, 1973 National Computer Conference, AFIPS Press, Montvale, N.J., 1973, pp. 581-587.

5. Shapiro, N. Z., and M. Davis, *Uncrackable Databanks*, The Rand Corporation, R-1382-NSF, December 1973.

6. Turn, R., "Privacy Transformations for Databank Systems," *AFIPS Conference Proceedings*, Vol. 42, 1973 National Computer Conference, AFIPS Press, Montvale, N.J., 1973, pp. 589-601. (Also to be published in W. W. Chu (ed.), *Advances in Computer Communications*, ARTECH House Publishing Company.)

7. Turn, R., *Toward Data Security Engineering*, The Rand Corporation, P-5142, January 1974. (To be published in *Proceedings, Workshop on Data Protection*, Computing Center of the Deutsche Forschungs- und Versuchanstalt fur Luft- und Raumfahrt (DFVLR), Oberpfaffenhofen, Germany.)

8. Turn, R., *Remarks on the Instrumentation of Databank Systems for Data Security*, The Rand Corporation, P-5151, January 1974. (To be published in *Proceedings, Workshop on Data Protection*, Computing Center of the Deutsche Forschungs- und Versuchanstalt fur Luft- und Raumfahrt (DFVLR), Oberpfaffenhofen, Germany.)

9. Turn, R., *Privacy and Security in Personal Information Databank Systems*, The Rand Corporation, R-1044-NSF, March 1973.

10. Turn, R., and N. Z. Shapiro, "Privacy and Security in Databank Systems: Measures of Effectiveness, Costs, and Protector-Intruder Interactions," *AFIPS Conference Proceedings*, Vol. 41, Part 1, 1972 Fall Joint Computer Conference, AFIPS Press, Montvale, N.J., 1972, pp. 435-444. (Also published in L. J. Hoffman (ed.), *Security and Privacy in Computer Systems*, Melville Publishing Company, Los Angeles, 1973, pp. 267-286.)

REFERENCES

1. *The Computer and Invasion of Privacy*, Hearings, U.S. Congress, House Committee on Government Operations, Special Subcommittee on Invasion of Privacy, 89th Cong. 2d Sess., July 26-28, 1966, U.S. Government Printing Office, Washington, D.C., 1966.

2. *Computer Privacy*, Hearings, U.S. Congress, Senate Committee on the Judiciary, Subcommittee on Administrative Practice and Procedure, 90th Cong., 1st Sess., March 14-15, 1967, U.S. Government Printing Office, Washington, D.C., 1967.

3. *Federal Data Banks, Computers and The Bill of Rights*, Hearings, Senate Committee on the Judiciary, Subcommittee on Constitutional Rights, 92d Cong., 1st Sess., February 23-25, March 2-4, 9-11, 15, and 17, 1971, U.S. Government Printing Office, Washington, D.C., 1971.

4. Westin, A. F., *Privacy and Freedom*, Atheneum, N.Y., 1967.

5. Miller, A. R., *The Assault on Privacy*, The University of Michigan Press, Ann Arbor, 1971.

6. "The Computerization of Government Files: What Impact on the Individual?" *UCLA Law Review*, Vol. 15, No. 5, September 1968, University of California, Los Angeles, pp. 1371-1498.

7. *Records, Computers, and the Rights of Citizens*, Report of the Secretary's Advisory Committee on Automated Personal Data Systems, U.S. Department of Health, Education, and Welfare, Washington, D.C., July 1973.

8. Westin, A. F., and M. A. Baker, *Databanks in a Free Society: Computers, Record-Keeping and Privacy*, Quadrangle Books, New York, 1972.

9. Michael, D. N., "Speculations on the Relation of Computers to Individual Freedom and Right of Privacy," *George Washington University Law Review*, Vol. 33, October 1964, pp. 270-275.

10. *The Freedom of Information Act*, Hearings, U.S. Congress, House Committee of Government Operations, Subcommittee on Government Information, 93d Cong., 1st Sess., May 2, 7, 8, 10, and 16, 1973, U.S. Government Printing Office, Washington, D.C., 1973.

11. Westin, A. F., "Computers and Problems of Privacy," *Proceedings, AFIPS/Stanford Conference on Computers, Society and Law: The Role of Legal Education*, AFIPS Press, Montvale, N.J., 1974, p. 37.

12. Gassman, H. P., "Data Banks and Individual Privacy: The Situation in the German Federal Republic," in S. Winkler (ed.), *Computer Communication Impacts and Implications*, Proceedings, First International Conference on Computer Communication, October 24-26, 1972, Washington, D.C., pp. 108-113.

13. *Compliance With the Fair Credit Reporting Act*, Division of Special Projects, Bureau of Consumer Protection, Federal Trade Commission, Washington, D.C., April 25, 1971.

14. "Sweden Enacts Privacy Law," *Electronics*, July 19, 1973, pp. 72-73.

15. Westin, A. F., and D. H. Lufkin, *The Impact of Computer Based Information Systems on Citizen Liberties in the Advanced Industrial Nations*, A Report to the German Marshall Fund, New York, 1973.

16. *Privacy and Computers*, Task Force Report, Departments of Communications and Justice, Information Canada, Ottawa, Canada, 1972.

17. Turn, R., and N. Z. Shapiro, "Privacy and Security in Databank Systems--Measures of Effectiveness, Costs, and Protector-Intruder Interactions," *AFIPS Conference Proceedings*, Vol. 41, 1972, Fall Joint Computer Conference, AFIPS Press, Montvale, N.J., 1972, pp. 435-444.

18. Atkinson, Andrews O., "Project CLEAR--An Integrated Regional Information System Serving Government, Law, and Justice," in G. A. Buck (ed.), *National Symposium on Criminal Justice Information and Statistical Systems*, California Crime Technological Research Foundation, Sacramento, 1971, pp. 79-88.

19. Roderick, D. R., "The National Crime Information Center--A Computerized Information System to Serve All Law Enforcement," in *Law Enforcement Science and Technology*, Proceedings of the 1st National Symposium, Thompson Book Company, Inc., Washington, D.C., 1967, pp. 529-532.

20. Gallati, R.R.J., "State Criminal Justice Information Systems," *AFIPS Conference Proceedings*, Vol. 39, 1971 Fall Joint Computer Conference, AFIPS Press, Montvale, N.J., 1971, pp. 303-308.

21. Hix, C. F., Jr., and J. E. Magsam, "A Large Scale Data Entry System for the IRS," *Datamation*, June 1970, pp. 106-109.

22. "The LOGIC Information System," in A. F. Westin (ed.), *Information Technology in a Democracy*, Harvard University Press, Cambridge, Mass., 1971.

23. *The National Driver Register*, Federal Highway Administration, National Highway Safety Bureau, U.S. Department of Transportation, Washington, D.C., September 1967.

24. Carroll, D., "Development and Test of a Prototype Interstate Organized Crime Index," in G. Cooper (ed.), *Proceedings of the International Symposium on Criminal Justice Information and Statistical Systems*, California Crime Technological Research Foundation, Sacramento, 1972, pp. 31-32.

25. Rehnquist, W. H., "Statement by W. H. Rehnquist," in *Federal Data Banks, Computers and The Bill of Rights*, Hearings, Senate Committee on the Judiciary, Subcommittee on Constitutional Rights, 92d Cong., 1st Sess., February 23-25, and March 2-4, 9-11, 15, and 17, 1971, U.S. Government Printing Office, Washington, D.C., 1971, pp. 597-624, 849-924.

26. Hearle, E.F.R., and R. J. Mason, *A Data Processing System for State and Local Governments*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1963.

27. *A National Survey of the Public's Attitudes Toward Computers*, American Federation of Information Processing Societies, Inc., and Time, Inc., New York, 1971.

28. *Report of the Committee on Privacy*, Secretary of State for the Home Department, Her Majesty's Stationery Office, London, July 1972.

29. Fellegi, I. P., "On the Questions of Statistical Confidentiality," *Journal of the American Statistical Association*, March 1972, pp. 7-18.

30. Gellman, H. S., *Statistical Data Banks and Their Effects on Privacy*, Department of Communications and Department of Justice, Ottawa, Canada, 1972.

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

32. Boruch, R. F., "Security of Information Processing: Implications from Social Research," *AFIPS Conference Proceedings*, Vol. 41, 1972 Fall Joint Computer Conference, AFIPS Press, Montvale, N.J., 1972, pp. 425-434.

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

34. *Guidelines for the Collection, Maintenance & Dissemination of Pupil Records*, Russell Sage Foundation, New York, 1969.

35. *Criminal Justice System*, Report by the National Advisory Commission on Criminal Justice Standards and Goals, Washington, D.C., 1973.

36. Ellis, L., "Privacy and the Computer--Steps to Practicality," British Computing Society, London, July 1972.

37. Baier, K., and N. Rescher, *Values and the Future*, The Free Press, New York, 1969.

38. "National Security Information," *Federal Register*, Vol. 37, No. 150, August 3, 1972.

39. *ADP Security Manual*, DoD 5200.28-M, U.S. Department of Defense, Washington, D.C., January 1973.

40. *Security of Automatic Data Processing Systems*, Handbook, U.S. Atomic Energy Commission, Washington, D.C., June 1972.

41. *U.S. Government Information Policies and Practices--Security Classification Problems and Practices (Part 7)*, Hearings, House Committee on Government Operations, Government Information Subcommittee, 92d Cong., 2d Sess., May 1-3, 5, 8, and 11, 1972, U.S. Government Printing Office, Washington, D.C., 1972.

42. Nejelski, P., and L. M. Lerman, "A Researcher-Subject Testimonial Privilege: What to Do Before the Subpoena Arrives," *Wisconsin Law Review*, Vol. 107, No. 4, pp. 1085-1148.

43. Kershaw, D. N., and J. C. Snell, "Data Confidentiality and Privacy: Lessons from the New Jersey Negative Income Tax Experience," *Public Policy*, Vol. 20, No. 2, Spring 1972, pp. 261-269.

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

45. "Legal Aspects of Computerized Information Systems," *The Honeywell Computer Journal*, Vol. 7, No. 1, 1973.

46. *Hearings Regarding the Administration of the Subversive Activities Control Act of 1950 and the Federal Civilian Employee Loyalty-Security Program (Part 1)*, Hearings, House Committee on Internal Security, 91st Cong., 2d Sess., September 23 and 30, 1970, U.S. Government Printing Office, Washington, D.C., 1970.

47. Hemphill, C. F., Jr., "Developing Security-Effective Employees," *Security World*, September 1972, pp. 16-19.

48. Blum, R. H., *Deceivers and Deceived*, Charles C. Thomas, Publisher, Springfield, Ill., 1972.

49. Gregory, C. O., and H. Kalven, Jr., *Cases and Materials on Torts*, Little, Brown and Company, Boston, 1969.

50. "Secret Phones Alarm Utilities," *Los Angeles Times*, February 23, 1972.

51. Senate Bill S.2810, "Right to Privacy Act of 1973," December 13, 1973; House Bill H.R. 10610, October 1, 1973; Senate Bill S.2963, "Criminal Justice Information Control and Protection Act of 1974," February 5, 1974; and House Bill H.R. 12575, February 5, 1974.

52. California Assembly Bill No. 2656 as amended, "California Fair Information Practices Act of 1973," January 7, 1974.

53. Van Tassel, D., *Computer Security Management*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1972.

54. "Computer Security: Backup and Recovery Methods," *EDP Analyzer*, January 1972.

55. Donovan, R., "Trade Secrets," *Security World*, April 1967, pp. 12-18.

56. Hicksom, P., *Industrial Espionage*, Spectators Publications, Ltd., London, 1968.

57. Flannery, J. P., "Commercial Information Brokers," *Columbia Human Rights Law Review*, Vol. 4, No. 1, Winter 1972, pp. 203-235.

58. *Sale or Distribution of Mailing Lists by Federal Agencies*, Hearings, House Committee on Government Operations, Foreign Operations and Government Information Subcommittee, 92d Cong., 2d Sess., June 13 and 15, 1972, U.S. Government Printing Office, Washington, D.C., 1972.

59. Petersen, H. E., and R. Turn, "System Implications of Information Privacy," *AFIPS Conference Proceedings*, Vol. 30, 1967, Spring Joint Computer Conference, AFIPS Press, Montvale, N.J., 1967, pp. 291-300.

60. Stockfish, J. A., "The Bureaucratic Pathology," in *Federal Statistics, Report of the President's Commission*, Vol. II, U.S. Government Printing Office, Washington, D.C., 1971, pp. 455-475.

61. Downs, A., *Inside Bureaucracy*, Little, Brown and Company, Boston, 1967.

62. Westin, A. F., "Civil Liberties and Computerized Data Systems," in M. Greenberger (ed.), *Computers, Communications and Public Interest*, Johns Hopkins Press, Boston, 1971, pp. 150-187.

63. Parker, D. B., *Threats to Computer Systems*, Final Report USRL-13574, Lawrence Livermore Laboratories, University of California, Livermore, March 14, 1973.

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

65. Carroll, J. M., et al., *Personal Records: Procedures, Practices and Problems*, Department of Communications and Department of Justice, Ottawa, Canada, 1972.

66. Turn, R., *Toward Data Security Engineering*, The Rand Corporation, P-5142, January 1974.

67. Turn, R., *Remarks on the Instrumentation of Databank Systems for Data Security*, The Rand Corporation, P-5151, January 1974.

68. Serang, A. M., "Integrated Security Systems," *Law Enforcement Science and Technology*, I, Thompson Book Company., Washington, D.C., 1967, pp. 811-816.

69. Desi, G. R., "An Approach to the Analysis of System Vulnerability to Clandestine Attack," *Law Enforcement Science and Technology*, I, Thompson Book Company, Washington, D.C., 1967, pp. 805-810.

70. *System Security Engineering*, Manual AFSCM 207-1, U.S. Air Force Systems Command, U.S. Government Printing Office, Washington, D.C., December 28, 1967.

71. Healey, R. J., *Design for Security*, John Wiley & Sons, Inc., New York, 1968.

72. Turn, R., R. Fredrickson, and D. Hollingworth, "Data Security Research at The Rand Corporation: Description and Commentary," *ADP Data Security and Privacy: Proceedings of the Conference on Secure Data Sharing*, Naval Ship Research and Development Center, Bethesda, Md., August 1973.

73. Johnson, S., *Certain Number-Theoretic Aspects of Access Control Passwords*, The Rand Corporation, R-1494-NSF, March 1974.

74. Friedman, W. F., and C. J. Mendelsohn, "Notes on Code Words," *American Mathematical Monthly*, August 1932, pp. 394-409.

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

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

77. Conway, R. W., W. L. Maxwell, and H. L. Morgan, "On the Implementation of Security Measures in Information Systems," *Communications of the ACM*, April 1972, pp. 211-220.

78. Woodward, F. G., and L. J. Hoffman, *On Worst-Case Costs for Dynamic Data Element Security Decisions*, Memorandum No. ERL-M413, Electronics Research Laboratory, College of Engineering, University of California, Berkeley, September 26, 1973.

79. Reed, I. S., *The Application of Information Theory to Privacy in Databanks*, The Rand Corporation, R-1282-NSF, May 1973.

80. Feistel, H., "Cryptography and Privacy," *Scientific American*, May 1973, pp.

81. Carroll, J. M., and P. M. McLellan, "Fast 'Infinite-Key' Privacy Transformations for Resource-Sharing Systems," *AFIPS Conference Proceedings*, Vol. 37, 1970 Fall Joint Computer Conference, AFIPS Press, Montvale, N.J., 1970, pp. 223-230.

82. Friedman, T. D., and L. J. Hoffman, *Execution Time Requirements for Programmed Encryption Methods*, Electronics Research Laboratory, Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, October 1973.

83. Weissman, C., "Security Controls in the ADEPT-50 Time-Sharing System," *AFIPS Conference Proceedings*, Vol. 35, 1969, Fall Joint Computer Conference, AFIPS Press, Montvale, N.J., 1969, pp. 119-133.

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

85. *Model Administrative Regulations for Criminal Offender Record Information*, Technical Memorandum No. 4, Project SEARCH Committee on Security and Privacy, California Crime Technological Research Foundation, Sacramento, March 1972.

86. Bisco, R. L. (ed.), *Data Bases, Computers and the Social Sciences*, Wiley-Interscience, New York, 1970.

87. Hansen, M. H., "Insuring Confidentiality of Individual Records in Data Storage and Retrieval for Statistical Purposes," *AFIPS Conference Proceedings*, Vol. 39, 1971 Fall Joint Computer Conference, AFIPS Press, Montvale, N.J., 1971, pp. 579-585.

88. Krauss, L. I., *SAFE, Security Audit and Field Evaluation for Computer Facilities and Information Systems*, Firebrand, Krauss, & Company, Inc., East Brunswick, N.J., 1972.

89. Martin, J., *Security, Accuracy, and Privacy in Computer Systems*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1973.

90. Anderson, J., "Computer Security in a Multi-User Computer Environment," in *Advances in Computers*, Vol. 12, Academic Press, New York, 1972.

91. Brown, W. F. (ed.), *AMR's Guide to Computer and Software Security*, AMR International, Inc., New York, 1971.

92. Hollingworth, D., *Enhancing Computer System Security*, The Rand Corporation, P-5064, August 1973.

93. *The Considerations of Physical Security in a Computer Environment*, International Business Machines Corporation, White Plains, N.Y., October 1972.

94. Wilk, R. J., "Engineering Considerations in Computer Center Security," *Proceedings, First International Electronic Crime Countermeasures Conference*, Edinburgh, Scotland, July 1973, pp. 234-247.

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

96. Carroll, J. M., and P. Reeves, "Security and Data Communications: A Realization of Piggyback Infiltration," *Infor*, October 1972, pp. 226-231.

97. Anderson, J. P., *Computer Security Planning Study*, ESD-TR-73-51, Vol. I and Vol. II, U.S. Air Force Systems Command, L. G. Hanscom Field, Bedford, Mass., October 1972.

98. Rotch, B., "Personnel Aspects of Data Center Security," *Proceedings, GUIDE 33*, Washington, D.C., November 1971, pp. 338-349.

99. *Report on the Workshop of Controlled Accessibility*, National Bureau of Standards, U.S. Department of Commerce, Washington, D.C. (in process).

100. Linden, T. A., "A Summary of Progress Toward Proving Programs Correct," *AFIPS Conference Proceedings*, Vol. 41, 1972 Fall Joint Computer Conference, AFIPS Press, Montvale, N.J., 1972, pp. 201-212.

END