CGTM No. 49

(see SLAC PUB 479 for final version)

COMPUTERS AND PRIVACY: THE PRESENT AND THE FUTURE

Lance J. Hoffman

Stanford University
Stanford Linear Accelerator Center
Computation Group
Stanford, California 94305

ABSTRACT

The computer privacy problem is stated in terms of existing systems and current proposals. A review of suggested legal and administrative safeguards is given. The bulk of the paper discusses the current technology, its limitations, and some additional safeguards which have been proposed but not implemented. Finally, a few promising computer science research problems in the field are outlined. A partially annotated bibliography of literature in the field is given.

CONTENTS

I. Introduction

This paper deals with the problem of privacy in large, computerized data banks. Section II states the problem in terms of existing systems and current proposals. A review of suggested legal and administrative safeguards is given in Section III. The major section, Section IV, is given over to a discussion of the current technology, its limitations, and some additional safeguards which have been proposed but not implemented. Finally, a few promising computer science research problems in the field are outlined in Section V.

II. The Privacy Problem

In the last several years, computer systems used as public utilities have moved from dream to reality. There are now a large number of multi-terminal, on-line, time-sharing systems in both commercial and academic environments.[1,2,3,4,5] Many people fully expect a public "data bank grid" to come into existence in the very near future; they point out that "it is as inevitable as the rail, telephone, telegraph, and electrical power grids that have preceded it, and for the same reasons. It is much less expensive and more efficient to share information than to reproduce it."[6]

Unfortunately, current information networks do not have adequate safeguards for protection of sensitive information. However, since the benefits derivable from computerization of large data banks are so great, pressure in some circles[7,8,9,10] is building up to "computerize now!". Computerization offers benefits in both economy and performance over many current systems.

Social scientists and statisticians, for example, have suggested the creation and maintenance of a National Data Bank.[9] Its use would remedy many defects of current files and procedures which result in information unresponsive to the needs of vital policy decisions. Some of these defects, as pointed out by Dunn, are:

"1) Important historical records are sometimes lost because of the absence of a consistent policy and procedure for establishing and maintaining archives.

2) The absence of appropriate standards and procedures for file maintenance and documentation lead to low quality files that contain many technical limitations in statistical usage.

3) Many useful records are produced as a by-product of administrative or regulatory procedures by agencies that are not equipped to perform a general purpose statistical service function.

4) No adequate reference exists that would allow users to determine easily whether or not records have the characteristics of quality and compatibility that are appropriate to their analytical requirements.

5) Procedures for collecting, coding and tabulating data that were appropriate when developed now lead to some incompatibilities in record association and usage required by current policy problems and made possible by computer techniques.

6) There are serious gaps in existing data records that stand in the way of bringing together records of greatest relevance for today's problems.

7) The need to by-pass problems of record incompatibility in developing statistics appropriate for policy analysis, places severe strains upon regulations restricting the disclosure of information about individuals. Technical possibilities for using the computer to satisfy these statistical requirements without in any way violating personal privacy have not generally been developed and made available by the agencies."[11]

To take advantage of the economies and capabilities of the computer, governmental agencies and private organizations such as credit bureaus are making use of computerized personal dossier systems. The New York State Identification and Intelligence System (NYSIIS) provides rapid access to criminal histories, stolen property files, intelligence information, etc., for use by "qualified agencies".[12] Santa Clara (California) County's LOGIC system includes a person's name, alias, social security number, address, birth record, driver and vehicle data, and other data if the person has been involved with the welfare or health departments, the district attorney, adult or juvenile probation, sheriff, court, etc.[10] Other municipalities have created similar systems.

These large data banks will make it easy for the citizen in a new environment to establish who he is and, thereby, to acquire quickly those conveniences which follow from a reliable credit rating and an acceptable social character. At the same time, commercial or governmental interests will know much more about the person they are dealing with. We can expect a great deal of information about the social, personal, and economic characteristics to be supplied voluntarily --often eagerly-- in order to acquire the benefits of the economy and the government.[13]

On the other hand, systems designed with insufficient consideration given to access control could be illicitly search for derogatory information.

Systems with insufficient input checking might be given false and slanderous data about a person which, when printed out on computer output sheets as the result of an inquiry, looks quite "official" and hence is taken as true. "On the horizon in technology is a laser scanning process that would enable a twenty-page dossier to be compiled on each of the United States' 200 million citizens. Such information could be stored on a single plastic tape reel. Under such conditions it might be cheaper to retain data than to discard it.[14] Clearly, we must decide what information to keep and when to keep it. As Paul Baran points out[15], we face a balance problem. How do we obtain the greatest benefit from computerized data banks with the least danger?

III. Legal and Administrative Safeguards

The problem of controlling access to computerized files --how to safeguard the processes of inputting to and retrieving from computerized data banks-- has recently gained more and more attention from concerned citizens. We will examine some of this new interest in this section, deferring mention of the technical solutions to Section IV.

Bauer has given a brief but sound discussion of policy decisions facing the designers of a computerized data bank, and has pointed out that we now have the "special but fleeting opportunity ... to explore the issue of privacy with objectivity and in some leisure. ... the public's fears of dossier-type police information systems have been thoroughly aroused; left unchecked they may become so strong as to in fact prevent the creation of any publicly supported information systems. The reactions to proposals for a Federal data center are a case in point. Were such blanket prohibitions to be imposed the development of socially useful information sharing would be enormously impeded. Furthermore, without public trust, information systems could well be fed so much false, misleading or incomplete information as to make them useless. Thus it becomes imperative not only to devise proper safeguards to data privacy, but also to convince the public and agencies which might contribute to a system that these safeguards are indeed being planned, and that they will work."[16]

Fortunately, the federal government is aware of the computer privacy problem and has quite effectively shot down proposals which did not adequately consider the effect of a centralized data bank on privacy.[17,18] Most of the states, however, lag seriously in awareness of contemporary data processing capabilities and techniques. Some of the more highly computerized areas are, however, trying to approach the idea of regional data banks in a rational manner. At least one state (California) has an intergovernmental board on automatic data processing which has solicited and received comments on confidentiality and the invasion of privacy from concerned members of the

technical community.

As Senator Sam J. Ervin, Jr. has pointed out,[19] the threat to privacy comes from men, not machines; it comes from the motives of political executives, the ingenuity of managers, and the carelessness of technicians. Too often, he says, an organization may seize upon a device or technique with the best intentions in the world of achieving some laudible goal, but in the process may deny the dignity of the individual, the sense of fair play, or the right of the citizen in a free society to privacy of his thoughts and activities.

"The computer industry, the data processing experts, the programmers, the executives-all need to set their collective minds to work to deal with the impact of their electronic systems on the rights and dignity of individuals.

"While there is still time to cope with the problems, they must give thought to the contents of professional ethical codes for the computer industry and for those who arrange and operate the computer's processes.

"If self-regulation and self-restraint are not exercised by all concerned with automatic data processing, public concern will soon reach the stage where strict legislative controls will be enacted, government appropriations for research and development will be denied. And the computer will become the villain of our society. It is potentially one of the greatest resources of our civilization, and the tragedy of slowing its development is unthinkable."[19]

Though Senator Ervin gave that speech on 1 May 1967, so far only Chariman Watson of IBM, of all the computer manufacturers, has commented publicly on the subject.[20] The Washington, D.C. chapter of the Association for Computing Machinery (ACM) has gone on record as opposing the creation of a national data bank until the proposers can show that "such a system is still economically attractive under the legal and technical constraints necessary to protect individual liberties in the American society".[21] (It has been alleged, however, that this vote reflects the views of a minority of that chapter's members and cannot necessarily be taken to represent the view of the chapter.)

We often forget that no "right to privacy", similar to the "freedom of speech" or the "right to vote", exists in the Constitution. Thus, the amount of privacy an individual is entitled to and when that privacy is

violated varies according to the whim of a particular court or legislative body.[19,22,23] Prosser, of the University of California School of Law at Berkeley, has compiled an excellent review of this subject.[24]

Recently, significant efforts have been made to create a more satisfactory situation. In 1966, John McCarthy suggested a "computer bill of rights". Some of the rights he proposed were these:

"No organization, governmental or private, is allowed to maintain files that cover large numbers of people outside of the general system.

"The rules governing access to the files are definite and well publicized, and the programs that will enforce these rules are open to any interested party, including, for example, the American Civil Liberties Union.

"An individual has the right to read his own file, to challenge certain kinds of entries in his file and to impose certain restrictions on access to his file.

"Every time someone consults an individual's file this event is recorded, together with the authorization for the access.

"If an organization or an individual obtains access to certain information in a file by deceit, this is a crime and a civil wrong. The injured individual may sue for invasion of privacy and be awarded damages."[25]

Additional suggestions have been made concerning legislative methods of safeguarding privacy. In 1967, the United States government proposed a Rights to Privacy Act banning wiretapping and electronic eavsdropping. (In 1968, however, the pendulum swung the other way and the Senate passed a "safe streets" and crime-control bill which granted broad authority for wiretapping and eavesdropping, even without a court order for a limited period of time.)

Even if a statute controlling access to sensitive information in files of the federal government were passed, the computer privacy problem will still be a long way from solved. A threat which is possibly even more serious is the misuse of data in the files of private organizations or in the files of state or local governments. Medical records in the files of hospitals, schools, and industrial organizations contain privileged information. When these records are kept in a computerized system, there must be control

over access to them. Some disconcerting examples of what has has happened when controls are lax are mentioned in a paper by Baran[15].

The California Assembly has before it currently (June 1968) a bill (AB 1381 - 1968 Regular Session) which if passed would (1) recognize an individual's right of privacy, and (2) recognize computerized data in state files as "public records". This bill, if passed, would be a landmark in the fight to establish a "right to privacy" and would seem to guarantee the right of an individual to read his own file.

The licensing or "professionalization" of (at least some) computer scientists, programmers, and operators seems to be the most frequent suggestion in the papers on computer privacy which are not written solely for computer scientists. In addition to Ervin (see above), advocates of this measure include Michael[26], Brictson[14], and Ramey[6]. Parker has been the main supporter of the ACM guidelines for Professional Conduct in Information Processing[27], but Brictson makes the best argument the author has seen for these to date[14]. With such current and potential outside interest in professional conduct of computer people, there has been very little published discussion about these matters. In view of Senator Ervin's unsettling predictions above, perhaps the computer community should give these problems more attention than it has to date.

This concludes the discussion of legal and administrative safeguards for the protection of sensitive information. We can now turn our attention to the technical solutions that have been proposed.

IV. Technical Methods Propsed to Date

A. Access Control in Conventional Time-Sharing Systems

Various technical methods for controlling access to the contents of computer memories have been suggested. In this discussion, these methods will be broken up into two categories -- those which are necessary for proper operation of a time-sharing system, and those which enhance the privacy of data in a shared system.

1. Methods Necessary for a Properly Operating Time-Sharing System

First let us consider the controls required in any time-sharing system. A means must be provided to lock out each user from the program and data of all other (unauthorized) users. In addition, a user must not be allowed to interfere with the time-sharing monitor by improper use of input/output commands, halt commands, etc. The latter capability is generally obtained by denying to the user of certain "privileged" instructions, which may be executed only be "privileged" programs such as the operating system.

The former is generally provided by memory protection schemes such as relocation and bounds registers[28], segmentation[29,30], paging[31], memory keys which allow limited (e.g., read-only) access[32], etc.

These access control methods all protect contiguous portions of (real or virtual) computer memory from alteration by an errant program. They do not, however, provide protection of a user file from unauthorized access. Towards this end, software schemes have augmented the hardware schemes described above.

2. Methods Which Enhance Data Privacy

With respect to the methods which enhance the privacy of data in a shared system, Paul Baran observed in 1966 that "It is a very poorly studied problem ... There is practically nothing to be found in the computer literature on the subject."[33] Since then, awareness has grown, largely as a result of congressional interest.[17,18] An entire session of the 1967

Spring Joint Computer Conference was devoted to this issue. But only very recently has there been developed a working system with more than password protection at the file level.[38]

In nearly all systems to date, a user's password will get him into his file directory and into any file referenced in that directory. The most elaborate scheme so far is that of Daley and Neumann[34] which features directories nested to any level used in conjunction with passwords. Each directory has access control information associated with itself. So, unless one has the "key" to each directory which appears on the chain to the desired file, one cannot get at the information in that file. Password schemes permit a small finite number of specific types of access to files, although Daley and Neumann[34] effectively provide more flexible control via a type which allows a user-written program to decide whether each requested access to a file is allowed.

3. Limitations of These Models

The methods of Section IV.A.1 perform their task acceptably -- they guarantee the system integrity. However, the password methods of Section IV.A.2 fall short of providing adequate software protection for sensitive files. Password schemes can be compromised by wiretapping or electromagnetic pickup, to say nothing of examining a console typewriter ribbon. Moreover, in some systems the work factor, or cost, associated with trying different passwords until the right one is found is so small that it is worth it to the "enemy" to do just that. Centralized systems tend to have relatively low work factors, since breaking a code in a centralized system generally allows access to more information than in a decentralized system. Some methods used to raise the work factor back to at least the level of a decentralized system are given later in this paper.

There is an even more serious problem with password systems. In all current systems, information is protected at the file level only -- it has been tacitly assumed that all data within a file was of the same sensitivity. The real world does not conform to these assumptions. Information from various sources is constantly coming into common data pools, where it can be used by all persons with access to that pool. The problem of what to

do when certain information in a file should be available to some but not all legal users of tha file is not well-studied. At Project MAC for example[1], it is currently the case that if a user has a file which in part contains sensitive data, he just cannot merge all his data with that of his colleagues. He must separate the sensitive data and save that in a separate file; the common pool of data does not contain this sensitive and possibly highly valuable data. Moreover, he and those he permits access to this sensitive data must, if they also want to make use of the nonsensitive data, create a distinct merged file, thus duplicating information kept in the system; if some of this duplicated data must be changed, it must be changed in all files, instead of only one. If there were a method to place data with varying degrees of sensitivity into common files and be guaranteed suitable access control over each piece of data, all the data could be aggregated and processed much more easily. Indeed, many social scientists are in favor of a National Data Bank for this very reason.[7,35] On the other hand, precisely because the problem has not been solved satisfactorily, lawyers[36,55], computer scientists[33,37,56] and the general public have become concerned about such a system.

In a recent thesis, Hsiao[38] has suggested and implemented files which contain "authority items"; these authority items control access to records in files. This is the first working system which controls access at a lower level than the file level. The implementation depends on a multi-list[39] file structure, but the idea of an authority item associated with each user is independent of the structure of the file. The accessability of a record depends on whether the file owner has allowed access to the requestor. This information is carried in the authority item. Capabilities[40] (such as read only, read and write, write only, etc.) appear to reside with the file rather than with each record.

A problem with Hsiao's scheme is the duplication in each authority item of entries for protected fields of one file. If there are J users of the system and each has K private fields in each of L files, then $(J-1)xKxL$ entries must be made in each authority item for user protection. Since there are J users, $T = Jx(J-1)xKxL$ entries must be maintained in the authority

items by the system. For the not unlikely case J = 200, K = 3, L = 2, we calculate T = 238,000. This price in storage and maintenance may well prove too much to pay in many instances.

Some other methods for access control have been proposed. Graham[41] has suggested a technique involving concentric "rings" of protection which may prove a reasonalbe way to provide flexible but controlled access by a number of different users to shared data and procedures. Dennis and van Horn[40] have proposed that higher-level programs grant access privileges to lower-level programs by passing them "capability lists".

Graham's scheme has several disadvantages. It assumes a computer with hardware paging and/or segmentation; since no large computer systems (of the type that would be necessary for a public utility) with these hardware facilities are as yet serving a large user community in an acceptable manner, this assumption may be premature, particularly in light of the alternatives such as extended core storage bulk memories.[42,57] The Graham scheme rules out the use of one-level memories such as associative memories,[54] Lesser memories,[43] etc. If the data bank has many different data fields with many different levels of access, the swap times necessary to access each datum in its own (two-word or so) segment will rapidly become prohibitive. In addition, the Graham scheme imposes a hierarchy on all information in the data base; this brings on quite a few problems in the passing of control from one procedure to another, as Graham points out in his paper.

The scheme of Dennis and van Horn suffers from all the drawbacks of the Graham scheme except the last. Compensating for this relative simplicity in the control structure however, a very large number of their meta-instructions must be executed for each attempt to access data which is not in a file open to every user.

B. Some Proposed Safeguards to the Privacy of Information in Files

In this section, we discuss countermeasures that have been proposed to more adequately insure against unauthorized access to information in files. Petersen and Turn have published an excellent paper[44] on the threats to information privacy, and much of the material of this section has been drawn from that paper.

The most important threats to information privacy are shown in Figure 1.

Accidental

- User error
- System error

Deliberate, passive

- Electromagnetic pick-up
- Wiretapping

Deliberate, active

- Browsing
- Masquerading as another user
- "Between lines" entry while user is inactive but on channel
- "Piggy back" entry by interception and transmitting an "error" message to the user
- Core dumping to get residual information

Figure 1. Some Threats to Information Privacy (extracted from [44])

We can encounter these threats by a number of techniques and procedures. Petersen and Turn have organized the various countermeasures into several classes: access management, privacy transformations, threat monitoring, and processing restrictions.

1. Access Management

These techniques attempt to prevent unauthorized users from gaining access to files. Historically, passwords have almost been synonymous with access management. Passwords alone, however, are not enough, as shown in Scetion IV.A.3 . The real issue in access management is authentication of a user's identification. Peters[45] has suggested using one-time passwords: lists of randomly selected passwords would be stored in the computer and maintained at the terminal or kept by the user. "After signing in, the user takes the next work (sic) on the list, transmits it to the processor and then crosses it off. The processor compares the received password with the next word in its own list and permits access only when the two agree. Such password lists could be stored in the terminal on punched paper tape, generated internally by special circuits, or printed on a strip of paper.

The latter could be kept in a secure housing with only a single password visible. A special key lock would be used to advance the list."[44] Another method based on random-number generation has been suggested by Baran.[46]

A novel idea based on the same principle -- the high work factor[46] associated with breaking encoded messages appearing as pseudo-random or random number strings[47] -- has been suggested by Les Earnest.[48] He proposes that the user login and identify himself, whereupon the computer supplies a pseudo-random number to the user. The user performs some (simple) mental transformation T on the number and sends the result of that transformation to the computer. The computer then performs the (presumably) same transformation, using an algorithm previously stored in (effective) execute-only memory at file creation time. In this way, while the user has performed T on x to yield y = T(x), any "enemy" tapping a line, even if the information is sent in the clear, sees only x and y. Even simple T's

(e.g., $T(x) = \left[\left(\sum_{i \text{ odd}} \text{digit } i \text{ of } x\right)^{3/2}\right] + (\text{hour of the day})$) are well-nigh

impossible to figure out, and the "cost per unit dirt"[49] is, hopefully, much too high for the enemy. Petersen and Turn point out that one-time passwords are not adequate against more sophisticated "between lines" entries by infiltrators who attach a terminal to the legitimate user's line. "Here the infiltrator can use his terminal to enter the system between communications from the legitimate user."[44] As a solution, they suggest one-time passwords applied to messages (as opposed to sessions), implemented by hardware in the terminal and possibly in the central processor. I conjecture that this solution will be too costly for most applications. I further conjecture that placing access control at the datum level, rather than at the file level, would eliminate many (though not all) problems associated with this type of infiltration.

Babcock[50] mentions a "dial-up and call-back" system for very sensitive files. When a sensitive file is opened by the program of a user who is connected to the computer via telephone line A, a message is sent to the user asking him to telephone the password of that file to the operator on a different telephone line B. The legal user can alter the password at will

by informing the data center.

2. Privacy Transformations

Privacy transformations are reversible encodings of data used to conceal information. They are useful for protecting against wiretapping, electromagnetic radiation from terminals, "piggyback" infiltration (See Fig. 1), and unauthorized access to data in removable files. Substitution (of one character string for another), transposition (rearrangement of the ordering of characters in a message), and addition (algebraically combining message characters with "key" characters to form encoded messages) are three major types of privacy transformations, which can be (and are) combined to increase the work factor necessary to break a code. This work factor depends (among others) on the following criteria:

" - Length of the key Keys require storage space, must be protected, have to be communicated to remote locations and entered into the system, and may even require memorization. Though generally a short key length seems desirable, better protection can be obtained by using a key as long as the message itself.

" - Size of the key space The number of different privacy transformations available should be as large as possible to discourage trial-and-error approaches, and to permit assignment of unique keys to large numbers of users and changing of keys at frequent intervals.

" - Complexity Affects the cost of implementation of the privacy system by requiring more hardware or processing time, but may also improve the work factor.

" - Error sensitivity The effect of transmission errors or processor malfunctioning may make decoding impossible.

Other criteria are, of course, the cost of implementation and processing time requirements which depend, in part, on whether the communication channel or the files of the system are involved."[47]

More detailed information on uses of privacy transformations is given

in Petersen and Turn[44]. A good unclassified discussion of encrypting and cryptanalysis methods, with particular attention paid to "distributed" communication networks (many terminals, many message switching centers, etc.) has been written by Baran.[46] He also has suggested[49] that we should always make use of minimal privacy transformations in the storage and transmission of sensitive data.

Privacy transformations can be performed by appropriate software in terminals and central processors. When desirable, hardware can be used instead. One current system, for example, uses basically a transposition method and is handled with preset plastic scrambler wheels; changes of these wheels are accomplished by time coordination.[51]

3. Threat Monitoring

Petersen and Turn give a good description of threat monitoring: "Threat montioring concerns detection of attempted or actual penentrations of the system or files either to provide a real-time response (e.g., invoking job cancellation, or starting tracing procedures) or to permit post facto analysis. Threat monitoring may include recording of all rejected attempts to enter the system or specific files, use of illegal access procedures, unusual activity involving a certain file, attempts to write into protected files, attempts to perform restricted operations such as copying files, execessively long periods of use, etc. Periodic reports to users on file activity may reveal possible misuse or tampering, and prompt stepped-up auditing along with a possible real-time response."[44]

Threat monitoring also will help improve the efficiency of the system, by reporting widespread use of particular system facilities. These system facilities can be "tuned", or, if need be, the facilities can be altered to eliminate bottlenecks. If some security restriction is unduly interfering with system operation, threat monitoring should help pinpoint the offending restriction.

4. Processing Restrictions

In additon to normal memory protection features mentioned in Section

IV.A.1, some processing restrictions may be desirable. Suggestions have included the mounting of removable files of drives with disabled circuits which must be authenticated before access[44], erasure of core memories after swapping a program and its data out to an auxiliary storage device, and built-in hardware codes which peripheral devices would transmit to other system components when necessary.[52]

There is a real question as to what price one wishes to pay for how much privacy.[53] In some instances, one might desire a whole processor to implement the entire file control and privacy system.[44] Most users, however, will probably settle for less privacy at less cost. This has been the experience so far of Allen-Babcock Corp. -- they have not implemented their "dial-up - call-back" privacy technique since none of their customers have demanded it.

Petersen and Turn have summarized their countermeasures to threats against information integrity, and the major part of the table they present is reproduced here:

Figure 2. Summary of Countermeasures to Threat to Information Privacy (extracted from [44])

Countermeasure / Threat	Access Control (passwords, authentication, authorization)	Processing Restrictions (storage, protected privileged operations)	Privacy Transformations	Threat Monitoring (audits, logs)
Accidental: User error	Good protection, unless the error produces correct password	Reduce susceptibility	No protection if depend on password; otherwise, good protection	Identifies the "accident prone" provides post facto knowledge of possible loss
System error	Good protection, unless bypassed due to error	Reduce susceptibility	Good protection in case of communication system switching errors	May help in diagnosis or provide post facto knowledge
Deliberate, passive: Electromagnetic pick-up	No protection	No protection	Reduces susceptability; work factor determines the amount of protection	No protection
Wiretapping	No protection	No protection	Reduces susceptability; work factor determines the amount of protection	No protection

Figure 2. Summary of Countermeasures to Threat to Information Privacy (continued)

Countermeasure / Threat	Access Control (passwords, authentication, authorization)	Processing Restrictions (storage, protected privileged operations)	Privacy Transformations	Threat Monitoring (audits, logs)
Deliberate, active: "Browsing"	Good protection (may make masquerading necessary)	Reduces ease to obtain desired information	Good protection	Identifies unsuccessful attempts; may provide post facto knowledge or operate real-time alarms
"Masquerading"	Must know authenticating passwords (work factor to obtain these)	Reduces ease to obtain desired information	No protection if depends on password; otherwise, sufficient	Identifies unsuccessful attempts; may provide post facto knowledge or operate real-time alarms
"Between lines" entry	No protection unless used for every message	Limits the infiltrator to the same potential as the user whose line he shares	Good protection if privacy transformations changed in less time than required by work factor	Post facto analysis of activity may provide knowledge of possible loss
"Piggy-back" entry	No protection but reverse (processor-to-user) authentication may help	Limits the infiltrator to the same potential as the user whose line he shares	Good protection if privacy transformations changed in less time than required by work factor	Post facto analysis of activity may provide knowledge of possible loss
Entry by system personnel	May have to masquerade	Reduces case of obtaining desired information	Work factor, unless depend on password and masquerading is successful	Post facto analysis of activity may provide knowledge of possible loss

Figure 2. Summary of Countermeasures to Threat to Information Privacy (continued)

Countermeasure / Threat	Access Control (password, authentication, authorization)	Processing Restrictions (storage, protected privileged operations)	Privacy Transformations	Threat Monitoring (audits, logs)
Deliberate, active, cont'd: Entry via "trap doors"	No protection	Probably no protection	Work factor, unless access to keys obtained	Possible alarms, post facto analysis
Core dumping to get residual information	No protection	Erase private core areas at swapping time	No protection unless encoded processing feasible	Possible alarms, post facto analysis
Physical acquisition of removable files	Not applicable	Not applicable	Work factor, unless access to keys obtained	Post facto knowledge form (sic) audit of personnel movements

V. Promising Research Problems

In this section, we discuss some technical problems which offer promising avenues for research in the future. We shall raise some relevant questions, but no answers are suggested in this paper.

1. Location in File Structure of Access Control Mechanism

For reasons mentioned in Section IV.A.3, the methods of protection which effectively pass privileges from one program to another are fairly unsatisfactory. We also saw there that protecting data by associating controls with the data at the file level only is not sufficient. What is really needed is some means of controlling access to each individual datum. Such a means should (1) be efficient, and (2) not unduly penalize the user who only wants a small part of his file protected. The mechanism may reside in program, data, indexes into an inverted file, authority items[38], or elsewhere. Parker[52] claims that this kind of protection can be expensive. I agree, but I have the feeling that it can also be inexpensive, and see in this subject a very interesting area for research.

2. Dependency of Access Control Efficiency on File Structure

The structure of a file is not independent of the method used to control access to it-- they may affect each other very strongly. For example, one might consider physically separating sensitive data in a hierarchical file (e.g., a tree-structured file). The more sensitive data could be stored in a memory which was logically at a low level and physically removed from higher-level data. This solution would not be feasible in certain types of associative memories, since the control would require all data to be at the same level. As another example, the existence of indexes into a tree-structured file (i.e., using an inverted file) might strongly alter the operating characteristics of the access control mechanism by allowing control information to reside in the indexes rather than (say)

with the data itself. Further investigation of this relationship is warranted.

3. Costs of Various Proposed Methods

Several types of countermeasures have been proposed to insure privacy: various types of threat monitoring, privacy transformations, access management, etc. Some hardware countermeasures, such as physical keys which record on a file or protocol the key number have also been suggested. Unfortunately, no systems, hardware or software, simulated or actual, have been built which enable us to evaluate the various costs of processing time, storage space, etc., of these methods. Why haven't these systems been built? Is it just that no one has gotten around to it yet? Is it only that no one needs a certain countermeasure (yet)? Is it that we don't really know how to implement what we theorize about in the literature? It is true that the literature on this is sparse. Even worse, there is almost a complete absence of implementation of nearly all of the proposed techniques.

Consider just one of these techniques, privacy transformations. Petersen and Turn discuss the further work that is needed:

> "Special attention must be devoted to establishing the economic and operational practicality of privacy transformations: determining applicable classes of transformations and establishing their work factors; designing economical devices for encoding and decoding; considering the effects of query language structure on work factors of privacy transformation; and determining their effects on processing time and storage requirements."[44]

The implementation of a (real or simulated) system using many countermeasure techniques, in order to evaluate them in practice, would be a very desirable undertaking.

VI. Summary

It is hoped that this paper may help increase awareness of the computer privacy problem and the need to investigate it further. Paul Baran puts

it well,

"What a wonderful opportunity awaits the computer engineer to exercise a new form of social responsibility. The advent of the new computer-communications technology need not be feared with trepidation as we approach 1984. Rather, we have in our power a force which, if properly tamed, can aid, not hinder, raising our personal right of privacy.

If we fail to exercise this unsought power that we computer engineers alone hold, the word 'people' may become less a description of individual human beings living in an open society and more a mere collective noun.

It may seem a paradox, but an open society dictates a right-to-privacy among its members, and we will have thrust upon us much of the responsibility of preserving this right."[49]

VII. Partially Annotated Bibliography

1. Crisman, P.A. (ed.), The Compatible Time-Sharing System-- A Programmer's Guide (second edition), MIT Press, Cambridge, 1965.

2. Schwartz, J.I., "The SDC Time-Sharing System", Datamation 10, 11(Nov. 1964), pp. 28-31.

3. Schwartz, J.I., "The SDC Time-Sharing System", Datamation 10, 12(Dec. 1964), pp. 51-55.

4. Computer Research Corporation, "Time-Sharing System Scorecard".

5. Parker, R.W., "The SABRE System", Datamation 11, 9(Sept. 1965), pp. 49-52.

6. Ramey, J.W., "Computer Information Sharing -- Threat to Individual Freedom", in Proc. Amer. Documentation Institute, 1967, pp. 273-277.

   Discusses, for a lay audience, why centralized data banks threaten privacy. Proposes licensing of computer professionals, much as CPA's are licensed now. Proposes legislation to allow an individual to inspect his entire dossier, delete inaccuracies via court order, and prohibit transfer of information identifiable with himself to a linked data bank without his express consent.

7. Dunn, E.S., Jr., statement in Reference 17.

8. Janssen, R.F., "Administration Studies Plan to Generalize Data, Hopes to Avoid 'Police State' Image", Wall Street Journal, 11 November 1966, p. 6.

9. Kaysen, C., "Data Banks and Dossiers" in The Public Interest, Spring 1967 (also in Reference 18).

   The case "for" a national data bank, in the light of the mauling this proposal got before the Gallagher subcommittee.

10. Davies, Lawrence E., "Computer Plan for Personal 'Dossiers' in Santa Clara Stirs Fears of Invasion of Privacy", New York Times, 1 August 1966.

11. Dunn, E.S., Jr., "The Idea of a National Data Center and the Issue of Personal Privacy", Amer. Statistician, 21, (Feb. 1967), p. 21ff.

An attempt by the author of the ill-fated BuBudget report to correct "certain obvious misinterpretations and set forth more explicitly some views on the very important issue of personal privacy." He says we have 10 or 15 years to figure out how to protect privacy while (right now) saving a lot of "harmless" data in his "statistical" data bank. The trade-offs on both sides are more clearly delineated than in the original report. I cannot be as sanguine over the prospects for protection of privacy.

12. Gallati, R.R.J., "The New York State Identification and Intelligence System", in Reference 17, p. 159ff.

13. Michael, D.N., "Speculations on the Relation of the Computer to Individual Freedom and the Right to Privacy", Geo. Wash. Law Review, 33, (1964-65), p. 270ff.

Between now and 1984, business and government will use extraordinary advances in computer technology to file and collate "personal" facts about private citizens and even to telemeter the populace. What are the implications for traditional ideas of freedom and privacy? Will such progress be met with constitutional objections or with public acquiescence? -- Author's abstract

A well-written paper with no technical content. However, it does make some valid and oft-overlooked points. It outlines factors which, in the past, have made privacy invasion difficult:

1. Data available but uncollected and uncollated.
2. Data not recorded with precision and variety necessary to gain new or deeper insight into the private person.
3. Difficulty of keeping track of a particular person in a large and highly mobile population.
4. Difficulty of access to already filed data about the private person.
5. Difficulty of detecting and interpreting potentially self-revealing private information within available data.

Points for a central data bank are validly and tellingly made, and the point is made that now, as in the past, people may give up some freedom to protect or enhance another freedom. How corruptible programmers may become privy quite legally to privileged information is discussed. In brief, a short, worthwhile paper.

14. Brictson, R.C., "Some Thoughts on the Social Implications of Computers and Privacy", System Development Corporation Document SP-2953/001/00 25 Sept. 1967.

This is a reprint of a talk presented to the American Society for

Industrial Security as part of a panel on "Problems in the Age of the Computer", 13th annual seminar, September 12-14, 1967, Los Angeles, California. Briefly discussed are (1) the computer as an innovation and tool along with some of the anxieties it creates, (2) a framework for an inquiry into the problem, (3) responsibilities of organizations and the establishment, (4) socialization--the preparation of new members for entry into society, (5) some examples reflecting issues, (6) possible remedies. In 11 short pages, a quite readable discussion understandable to the lay person is given. The framework suggested for investigation seems quite reasonable, and represents one of the few attempts to define the general problem before rushing off to tackle it. This structure considers information from the standpoint of (1) acquisition, (2) access, (3) dissemination, (4) retention, (5) revision, including updating, rejoinder and redress, (6) destruction, (7) time cycles. Brief examples are given for acquisition and protection. A good case (and a brief one) for the existence of professional ethics codes is made, much better than the discussion in Communications of the ACM, Vol. 11, No.3 (Mar. 1968) by Parker. Five guidelines for public policy makers are suggested: (1) specifications of benefits, (2) catalog of potential risks, (3) directory of preventive safeguards and controls (4) inventory of antidotes and countermeasures, (5) index of penalities and sanctions.

A very good paper for the layman and interested computer scientist.

15. Baran, P., "Remarks on the Question of Privacy Raised by the Automation of Mental Health Records", RAND Document P-3523, April 1967.

Remarks invited for presentation before the American Orthpsychiatric Association Workshop on "The Invasion of Privacy", held in Washington, D.C., 21-23 March 1967. A speech of Baran which presents in excellent fashion, to an intelligent group of computer laymen, a view of computer privacy invasion which only computer types heretofore have appreciated. Some horror stories are recalled, with emphasis on medical record leaks, in view of the audience. The famous tale of the MIT freshman who progammed the computer to dial up every telephone extension in the school simultaneously is retold, and thus is graphically illustrated what a real "bad guy" could do. A very good talk to alert intelligent people about the implications of the computer age for privacy.

16. Bauer, K.G., "Report on the Joint Center for Urban Studies Project for the Preliminary Design of a Health Information System for Boston for the Period October 1 through December 31, 1967", Cambridge, Mass., 1967.

A nine-page section on the privacy issue as it relates to a proposed health information system for the Boston ares. "...Right now our project has a unique opportunity to propose safeguards to privacy in the design of an information system at a time when the crucial operational decisions have not yet been made. ..." Discusses present safeguards to record disclosure. Currently privacy is not really insured, and only the excessive cost of getting sensitive information (because of the unwieldiness of current non-computerized systems) prevents almost all unauthorized

access. "...With proper safeguards computerization makes such information far easier to guard ..." -- Why this is the case is explained. A broad framework of new safeguards, combining legal, technological, and administrative measures is being urged, and these are gone into very briefly, with references to a couple of papers. The committee hopes during the coming months to secure staff help to define levels of security and to suggest specific access rules and rights of patients that should be kept in mind.

17. "The Computer and the Invasion of Privacy -- Hearings before a Subcommittee of the Committee on Government Operations, House of Representatives, 89th Congress, Second Session", (the Gallagher report), 26-28 July 1966, U.S. Government Printing Office.

Pro and con on a national "statistical" data bank-- the full testimony.

18. "Computer Privacy -- Hearings before the Subcommittee on Administrative Practice and Procedure of the Committee on the Judiciary, United States Senate, 90th Congress, First Session" (the Long report), 14-15 March 1967, U.S. Government Printing Office.

The full testimony before the Long subcommittee on computer privacy.

19. Ervin, Sam J., "The Computer -- Individual Privacy", Vital Speeches of the Day, 1 May 1967, p. 421.

Senator Ervin discusses the impact of the computer on national life in a speech to the American Management Association. He thinks the industry, to avoid strict legislative controls and denial of government research and development funds, must devise safeguards against improper data access, illegal tapping, and purloined data in shared systems. He evidently likes the idea of an industry ethical code.

20. Watson, T.J., Jr., "Technology and Privacy", speech given to Commonwealth Club of California, Hotel St. Francis, San Francisco, 5 April 1968.

An Address by Thomas J. Watson, Jr., Chairman of the board of IBM, to the Commonwealth Club of California. Watson discusses in general what the privacy problem is, advantages and disadvantages of centralized data banks, possible solutions to the problem, and gives suggestions for legal, ethical, and technological safeguards.

21. Warburton, P., "A National Data Center and Personal Privacy -- Resolution Proposed", Computers and Automation, 16, 5, May 1967, p. 8.

A resolution on the National Data Center and Personal Privacy proposed by the Washington, D.C. chapter of the Association for Computing Machinery.

22. Lickson, "The Right of Privacy in the Computer Age", IEEE Computer Group News, 2, 1(Jan. 1968).

A nontechnical five page paper which defines privacy, examines some historical court cases dealing with it, and tries to pinpoint current legislative trends in this area. "...Legislation and court decisions can catch up to the state of the art." A good general overview from a nontechnical standpoint, well-referenced.

23. Westin, A.F., Privacy and Freedom, Atheneum, New York, 1967.

A comprehensive, well-written book on the relationship of privacy to freedom, tracing "privacy rights" from 1776 to the present. The emphasis is, by far, on the present and the future. The book has four parts: the functions of privacy and surveillance in society, new tools for invading privacy, American society's struggle for controls (five case studies), and policy choices for the 1970's. Each part is copiously documented, and in addition there are four bibliographies at the end: the functions of privacy, the new technology, the struggle for controls, and privacy in American law and policy. The section on computer technology and possibilities for it by 1975 was quite enlightening, even to a computer science graduate student. This is a must book for those concerned with the privacy problem. Westin is, at the time of this review, Professor of Public Law and Government at Columbia, and numerous legal decisions are cited. It is a seminal work in the field.

24. Prosser, W.L., "Privacy", California Law Rev., 48, 3(Aug. 1960), p. 385ff.

A review of court cases dealing with a "right to privacy". The review appears to be comprehensive (to this layman at law). The author, then Dean of the University of California Law School at Berkeley, contends that four distinct kinds of privacy invasion cases can be described: (1) intrusion upon seclusion or solitude, or into private affairs, (2) public disclosure of embarrassing private facts, (3) publicity which places the plaintiff in a false light in the public eye, (4) appropriation, for the defendant's advantage, of the plaintiff's name or likeness. The article is well-written and interesting. As a final fillip, I can not conclude without praising the author for making me aware of "a possible nomination for the all-time prize law review title, in the note 'Crimination of Peeping Toms and Other Men of Vision', 5 Ark. L. Rev. 388 (1951)."

25. McCarthy, J., "Information", Scientific American, 215, 3(Sept. 1966), p. 64ff.

McCarthy, in a very good survey article on computation, proposes a computer bill of rights which would guarantee privacy in computerized data files.

26. Berkeley, E.C., "Individual privacy and Central Computerized Files", Computers and Automation, 15, 10(Oct. 1966), p. 7.

Discusses a privacy bill of rights initially suggested by Professor John McCarthy in his lead article "Information" in Scientific American of Sept. 1966

27. Parker, D.B., "Rules of Ethics in Information Processing", Communications of the Association for Computing Machinery, 11, 3(Mar. 1968), p. 198ff.

28. Control Data 6400/6600 Computer Systems Reference Manual, Control Data Corp. St. Paul, Minn., 1966.

29. Corbato, F.J., and Vyssotsky, V.A., "Introduction and Overview of the Multics System", Proc. Fall Joint Computer Conference 1965, p. 185ff.

30. System/360 Model 67 Time-Sharing System Preliminary Technical Summary, IBM Corporation, White Plains, New York, 1966.

31. SDS 940 Computer Reference Manual, Scientific Data Systems, Santa Monica, California, Aug. 1966.

32. IBM System/360 Principles of Operation, IBM Corporation, Poughkeepsie, New York, 1966.

33. Baran, P., statement in Reference 14.

34. Daley, R.C., and Neumann, P.G., "A General-Purpose File System for Secondary Storage", Proc. Fall Joint Computer Conference, 1965, p. 213ff.

Control is placed on the branches of a tree-structured file directory. Five modes of control are allowed -- trap, read, execute, write, and append. Some of the best thinking about a practical, general solution to lower-level access control yet. One of the "Multics papers". Must reading for data base system designers.

35. Bowman, R.T., statement in Reference 14.

36. Reich, C.A., statement in Reference 14.

37. Squires, B.E., Jr., statement in Reference 14.

38. Hsiao, D.K., A File System for a Problem Solving Facility, dissertation in Electrical Engineering, Univ. of Pennsylvania, 1968.

An important new concpt is introduced and implemented on the file system at Penn. This concept, that of the authority item, allows control within files over data access. Each field in a file can be protected from unauthorized access. Data records need not be reprocessed if a change in a record's protection status or in a user's level of accessability occurs. The capability to read only, write only, etc., goes with a file and not with a record. Protected records are completely nonexistent as far as the unauthorized user is concerned. The system as currently implemented is dependent on the file structure (multi-lists). However, the idea of authority items is not and is an important new concept. This thesis should be examined by those who have the responsibility for access control in thier own file systems. It appears to be the first working system with protection below the file level.

39. "A Storage Retrieval System for Real-Time Problem Solving", University of Pennsylvania Moore School Report No. 66-05.

40. Dennis, J.B., and Van Horn, E.C., "Programming Semantics for Multi-programmed Computations", Communications of the ACM, 9, 3 (March 1966), p. 143ff.

A number of meta-instructions are defined which relate to programming operations in multiprogrammed systems. These are related to parallel programming, protection of separate computations, sharing of files and memory. The meta-instructions are cumbersomely put into an Algol-like language, but nevertheless, some very good and long-neglected ideas are here. The capabilities are related to segments which are not quantitatively defined. In practice, these are still too large for a basic unit, and something else ought to be used, e.g., nodes of a tree. This may be possible by altering the Dennis and Van Horn scheme to acquire programs, rather than lists; these programs could call appropriate macros to set up the lists they need.

41. Graham, R.M., "Protection in an Information Processing Utility", Communications of the ACM, 11, 5, May 1968, pp. 365-369.

A good five page paper on the topic. A solution to the file

access problem is given which involves rings or spheres of protection for both data and programs (in particular, for segments, as at Project MAC). The main drawbacks are: (1) the method is tied to segments which in practice are fairly large blocks of memory, and protection of a smaller area wastes the rest of the segment; (2) parallel processes or processors may render invalid parameters or data if proper safeguards are not taken. Aside from these considerations, this may be a reasonable way to provide flexible but controlled access by a number of different users to shared data and procedures.

42. Mac Dougall, M.H., "Simulation of an ECS-based Operating System", Proc. Spring Joint Computer Conference 1967, pp. 735-741.

43. Lesser, V.R., "A Multi-Level Computer Organization Designed to Separate Data-Accessing From the Computation", Stanford Linear Accelerator Center Computation Group CGTM-37, Jan. 1968.

44. Petersen H.E., and Turn, R., "System Implications of Information Privacy", Proc. Spring Joint Computer Conference 1967.

"Various questions of providing information privacy for remotely accessible on-line, time-shared information systems are explored.... A range of protective countermeasures is discussed, and their choice and implication considered. It appears possible to counter a given level of threat without unreasonable expenditures of resources. The protective techniques discussed ... include: shielding to reduce electromagnetic emanations; use of once-only passwords for access control; application of privacy transformations to conceal information in user-processor communications and in data files; recording of attempted penetrations; and systematic verification of the hardware and software integrity." (authors' abstract)

A detailed and well-written paper on threats and countermeasures for file security. Problems at the processor, the files, the terminals, and the communication lines are discussed. A good bibliography is given. A must paper.

45. Peters, B., "Security Considerations in a Multi-programmed Computer System", Proc. Spring Joint Computer Conference 1967.

A specific list of desirable and necessary security safeguards in file systems for both hardware and software.

46. Baran, P., "On Distributed Communications: IX. Security, Secrecy, and Tamper-Free Considerations", RAND Corporation RM-3765-PR (unclassified), Aug. 1964 (DDC Accession Number AD-444839).

A consideration of the security aspects of a distributed communication system, written from the viewpoint that we should fully anticipate the existence of spies within our ostensibly secure communications secrecy protection structure; "Hence, our primary interest should be in raising the 'price' of espied information to a level which becomes excessive." The proposed system combines end-to-end and link-by-link cryptography, automatic error detection and repeat transmission, path changing, and use of a scheme requiring complete and correct reception of all previous traffic in a conversation in order to decrypt subsequent message blocks. It assumes enemy infiltration and takes these countermeasures: key bases split over N (>1) individuals, filtering tests, key change for each conversation, heavy system use for unclassified traffic. Contents:

I. Introduction
II. The Paradox of Secrecy about Secrecy
III. Some Fundamentals of Cryptography
IV. Implications for the Distributed Network System
V. A "Devil's Advocate" Examination

A clear, well-written discussion of an often "touchy" subject. Relevant points are brought out by good diagrams. One of the clearest expositions of real-to-life problems and solutions to be found in the open literature.

47. Shannon, C.E., "Communication Theory of Secrecy Systems", Bell System Technical Journal, 28, 4(Oct. 1949), pp. 656-715.

In this classic paper, a mathematical theory of secrecy systems is developed. The theory is presented in a most readable form. First, basic mathematical structure of secrecy systems is dealt with. Examples of various types of ciphers are given. Measures of "how secret" a system is are introduced, and it is shown that "perfect" secrecy is possible but requires, if the number of messages is finite, the same number of possible keys. A measure of "noise" in a message is given, and strongly ideal systems where this cannot be decreased by the cryptanalyst are discussed. Finally, an analysis of the basic weaknesses of secrecy systems is made. This leads to methods for constructing systems which require a large amount of work to solve. Finally, a certain incompatibility among the various desirable qualities of secrecy systems is discussed. An excellent paper, and doubly so for the non-fainthearted in mathematics (particularly probability and modern algebra).

48. Earnest, L., private communication.

49. Baran, P., "Communications, Computers and People", Proc. Fall Joint Computer Conference 1965, Part 2, pp. 45-49.

A well-thought out general discussion of the privacy problem. Overlaps somewhat with his testimony before the Gallagher subcommittee. Contains some specific proposals.

50. Babcock, J.D., "A Brief Description of Privacy Measures in the RUSH Time-Sharing System", Proc. Spring Joint Computer Conference 1967, pp. 301-302.

A brief summary of the file security procedures in RUSH. Contains some good but short discussion of possible threats and associated countermeasures.

51. McLaughlin, F.X., private communication.

52. Parker, D.B., "Privacy in Resource-Sharing Computer Systems", Control Data Corporation Programming Technical Report TER-06, Nov. 1967 (company private).

An excellent down-to-earth paper on objectives of penetration of computer systems, threat types, countermeasures, and methods of protection. A survey of past and current privacy considerations of Control Data Corporation and its customers is made. Privacy methods are proposed for internal use and for products-- problems arising from oft-proposed, relatively simple (and in fact, too simple) methods are brought up.

53. Weissman, C., "Programming Protection: What Do You Want to Pay?", SDC Magazine 10, 7 and 8 (July, August 1967), System Development Corporation, Santa Monica, Ca.

54. Feldman, J.A., "Aspects of Associative Processing," MIT Lincoln Laboratory Technical Note 1965-13.

55. Duke University School of Law, "Privacy", Law and Contemporary Problems, XXXI, 2 (Spring 1966).

A Law Journal issue. Contents:

Foreword by Clark C. Havighurst
The Right to Privacy and American Law by William M. Beany
Privacy and The Law: A Philosophical Prelude by Milton R. Konvitz
Privacy: Its Constitution and Vicissitudes by Edward Shils
Some Psychological Aspects of Privacy by Sidney M. Jourard

Philosophical Views on the Value of Privacy by Glenn Negley
Privacy in Tort Law--Were Warren and Brandeis Wrong by Harry Kalven, Jr.
"The Files": Legal Controls Over the Accuracy and Accessability of Stored Personal Data by Kenneth L. Karst
Privacy in Welfare: Public Assistance and Juvenile Justice by Joel F. Handler and Margaret K. Rosenheim
The Privacy of Government Employees by William A. Creech

Nothing on computer methods except in the Karst paper, which has about four pages on the effect of automation. The possible solutions to this aspect of the privacy problem are dealt with in superficial detail, but relevant references are given for the reader interested in a more advanced technical discussion.

56. Harrison, A., The Problem of Privacy in the Computer Age: An Annotated Bibliography, RAND Corporation RM-5495-PR/RC, Dec. 1967.

A must document. This 300 entry bibliography is well-annotated and filed by author and by each of the following categories:

cashless-checkless society	computer utilities
time-sharing	congressional view of privacy
data banks	legal views
media	system security
social scientists' views	technologists' views
bill of rights	
electronic eavesdropping and wiretapping	

57. Humphrey, T. A., "Large Core Storage Utilization in Theory and in Practice", Proc. AFIPS 1967 Spring Joint Computer Conference, Vol. 30, Thompson Book Co., Washington, D. C., p. 719.