532

Capuano, Robert M.
Conference on the Boundaries of
Privacy in American Society
Woodrow Wilson School of Public
and International Affairs
January 3, 1972

TECHNOLOGY
and the
CONTROL OF STORED DATA

Robert Capuano

I. STATEMENT OF THE PROBLEM

The problem of this paper is to describe the uses of the computer, particularly in the area of stored information, and to study the technologically feasible means of safeguarding this information.

II. SUMMARY OF PRINCIPAL FINDINGS

1. Computer technology has rapidly advanced from its early stages of purely mathematical operations to its present stage of information processing. The operational speed and storage capacities of computers have been steadily increasing while the physical size and cost of computing have been decreasing substantially. These trends are expected to continue for at least the next five years.

2.. Through the advancements made in programming languages, the computer is now easier to use. Programming languages have been moving closer and closer to written English, and more people are now able to use computers directly than ever before.

3. More and more data is continuously being stored in computer banks. Much of this data is useful, such as harmless statistical data, which aids the government in forming policy. Experts predict major contributions from the computer in the fields of Law-enforcement, Administration, Education, and Health. The question of privacy, however, arises from the misuse of personal data, which is potentially dangerous to ~~the~~ the individual.

4. Gaining improper access to computer files is relatively easy and highly profitable because of three factors: (1) the development of on-line, remote-access time-sharing systems, in which it is possible for a user to have direct contact with the central computer from a terminal thousands of miles away; (2) a lack of safeguards to protect private information; and, (3) a vagueness in the legal status of private information.

5. While a number of safeguards are technologically feasible, a major problem concerning their implementation is the cost factor. This overriding cost factor may prevent computer manufacturers and computer users from voluntarily safeguarding their systems. Therefore, the Federal Government may have to intervene to protect privacy.

6. No safeguarding system will be perfect, but, through a combination of technological safeguards and laws pertaining to the misuse of information, the price of snooping may be made sufficiently high to deter potential misusers.

7. To prevent a safeguarding system from becoming obsolete soon after implementation, careful attention must be given now to possible developments in computer technology and in snooping devices in the future.

III. POLICY RECOMMENDATIONS

1. Implementation of the following safeguarding devices in any system which handles potentially harmful information: (a) partitioning of the memory banks (pp. 28-29); (b) relatively simple encryption codes (pp. 23-24); and, (c) real-time monitoring and random auditing of the security system (p. 27).

> Access to our present systems is relatively easy (pp. 16-19) and the dangers from the misuse of information are clear (pp. 20-21a). Because of the cost factor, self-regulation by the information industry does not seem feasible (pp. 29-29a) and therefore the Federal Government must intervene to make certain that the above safeguards are installed.

2. Enactment of Federal legislation requiring the installation of the above three safeguards on all computer systems which handles personal information which could be misused. This law would apply to all computer systems both old and new.

3. Enactment of Federal legislation requiring the lincensing of computer systems, and their operating personnel. This legislation should also create a federal regulatory agency to oversee the issuance of operating licenses.

> This agency should be staffed by career civil servants and experts in the fields of privacy and computer technology. The agency will issue and revoke licenses along guidelines set up by Congress which are flexible enough to enable the agency to bend as technology advances. W.H. Ware suggests

such guidelines. To obtain an operating license the owner of a system must demonstrate the following to the federal agency: 1) the nature and purpose of the data bank including the uses of the data and the class of customers it serves; 2) the precise identification and description of sources and the checks for validity of the information; 3) complete description of physical safeguards in the system for protection and control of divulgence; 4) Description of procedural safeguards to edit information for errors, to assusre posting information to the correct file, to resolve ambiguity in identification of an individual, to treat information of doubtful validity, and to establish confidence levels on information derived or inferred from fragmentary data; 5) description of audit processes and audit information made available for periodic review; 6) a procedure for an individual to review his dossier and its sources, to challenge its contents and to correct errors; and, 7) the tests and inspections performed on the system to assure that it operates properly.[1]

The regulatory agency should also set up a committee of computer experts to conduct research on technological safeguards, in order to keep the agency up-to-date with the latest technological advances in the computer field. Recently a New Jersey engineering firm has developed an identification device which minutely measures the dimensions of a person's hand and is supposed to be more reliable and easier to administer than fingerprints. The Strategic Air Command has already installed such devices at its highest security bases.[2] Research of this type could be conducted by the agency's committee on research.

[1] W.H. Ware, "Computer Data Banks and Security Controls," RAND Corp. Report P-4329, pp. 7-8.

[2] Mel Most, "The Hand Is Quicker Than an ID Card," The Record (Bergen County), December 14, 1971.

IV. TABLE OF CONTENTS

V.. DISCUSSION

Computers--History and Development

Brief History

> The Analytical Engine has no pretensions whatever to originate anything. It can do whatever we *know how to order it* to perform. It can *follow* analysis, but it has no power of *anticipating* any analytical relations or truths. Its province is to assist us in making *available* what we are already acquainted with.[1]

The Orwellian fear of the all-powerful, intelligent computer subordinating man, monitoring his every action, and running worlds, seems to be a modern phenomena based in numerous works of science fiction.[2] However, the above quote was written in the mid-nineteenth century to allay the fears of people in England concerning a newly-designed machine, the Analytical Engine. The Analytical Engine was the prototype of the modern computer minus our present highly-sophisticated technology. The Engine was the invention of Charles Babbage, who spent from 1833 to 1871 designing and trying to build a machine, which would perform numerous and fairly complex mathematical operations. The Engine was a very complex, intricate system of finely calibrated gears, wheels, and levers.

[1] Jeremy Bernstein, *The Analytical Engine: Computers--Past, Present and Future*, p. 46.

[2] The works of Isaac Asimov (*I, Robot*), Harlan Ellison, and *2001* by Arthur C. Clarke are a few examples of this type.

Babbage failed to complete the machine because the technology of the time was unable to provide him with these intricate pieces of equipment. However, his designs influenced the pioneers of computer technology in the twentieth century.

The logical organization of modern computers is very similar to the logical organization of Babbage's engine and consists of four basic units: (1) andInput/Output unit; (2) a Memory unit; (3) an Arithmetic unit; and, (4) a Control unit. The Input/Output (I/O) unit involves getting data and instructions into the computer and getting answers out. The I/O equipment is the link between the human and the machine. The memory banks store all the data and instructions for some frequently used operations (e.g. taking square roots, logariithms, etc.). The Arithmetic unit and the Control unit make up the Central Processing Unit which manipulates stored numbbers and controls the sequence of operations.[1] The physical equipment of these four units comprise the "hardware" of the computer. The "software" refers to the programs or sets of instructions that control the storage, retrieval and manipulation of data in the computer's banks.[2] An examination of the general trends of computing power and of the specific

[1] Bernstein, op. cit., p. 48.

[2] Arthur R. Miller, The Assault on Privacy: Computers, Data Banks, and Dossiers, p. 13.

developments of both hardware and software over the past thirty years, would be helpful in comprehending today's computer capabilities and in anticipating tomorrow's computer potential.

General Trends

The most advanced computer of the mid-nineteen fifties was a machine which occupied 1,000 cubic feet of space, and which could perform 25,000 additions per second.[1] The cost of one million operations was ten dollars and the total computing power in the United States in 1955 was 500,000 additions per second.[2] By 1965, the size of computers decreased by a factor of 10 to 100 cubic feet, and the operational speed increased by a factor of 200 to 5 million additions per second.[3] At the same time the cost of one million operations decreased by a factor of 300 down to about 3.3 cents and the total computing power in the United States increased four hundredfold to 200 million additions per second.[4] The figures from this ten-year period represents the rapid advance which has taken place in computer technology and which computer experts expect to continue into the future. Paul Armer, a computer

[1] Alan Westin, Privacy and Freedom, p. 166.

[2] W.H. Ware, "Future Computer Technology and Its Impact," Rand Corp. Report P-3279, p. 15.

[3] Westin, op. cit., p. 166.

[4] Ware, op. cit., p. 15.

expert for the RAND Corporation, described these technological advances in 1967 in terms of order of magnitudes, by which he means an increase or decrease of a factor of ten:

> But the speed of the electronic portions of computers has been increasing by an order of magnitude about every four years, and it looks like that pace will continue for some time [Fig. 31]. Size (again I'm talking about the electronic portion of the computer) decreased by an order of magnitude in the last ten years, and will probably decline by three orders of magnitude during the next decade [Fig. 29]. More importantly, the cost of raw computing power has declined by an order of magnitude every four years, and this trend looks like it will hold for awhile [Fig. 30].[1]

Another RAND expert Olaf Helmer views the development of the computer as the first phase of a two-phased process:

> It took just twenty years for the first computer fevolution to be completed, from the mid-forties to the mid-sixties, during which time the computer grew up from being a bookkeeping device to becoming a highly versatile data processor and research tool. During that period the size and the cost of electronic computer components have gone down by factors of 100 and 100,000 respectively, and their speed has gone up by a factor of 100,000.[2]

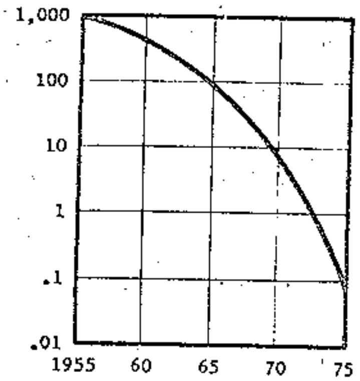
These trends have continued up to the present with the total amount of computing power available in the United States doubling annually, while costs have been decreasing rapidly.[3]

[1] Paul Armer, "Social Implications of the Computer Utility," RAND Corp. Report P-3642, p. 5.

[2] Olaf Helmer, "Prospects of Technological Progress," RAND Corp. Report, P-3643, p. 8.

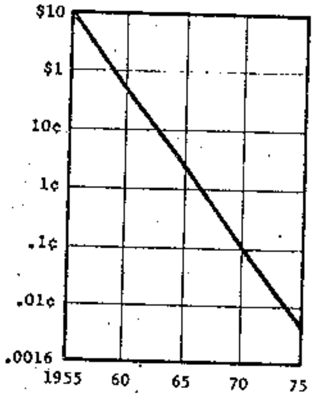
[3] Paul Armer, "Computer Aspects of Technological Change, Automation, and Economic Progress," RAND Corp. Report P-3478, p. I-229.

COMPUTING TRENDS IN THE UNITED STATES[1]



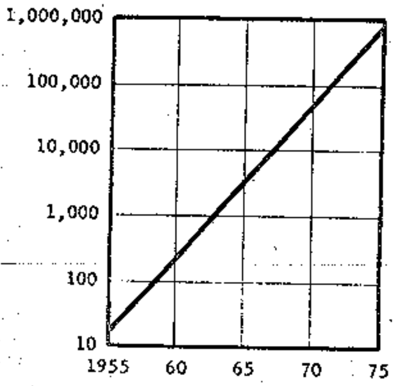
SIZE
CPU/Storage size in cubic feet

Fig.29



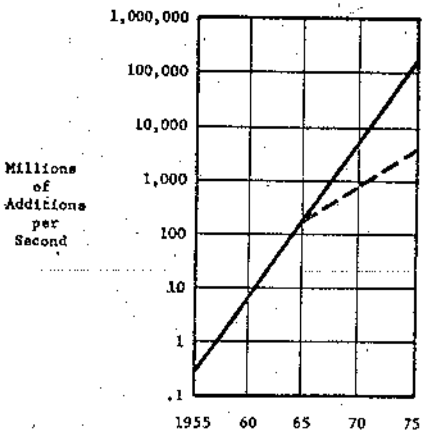
COST
CPU/Storage cost in dollars per million additions

Fig.30



SPEED
CPU/Storage speed in thousands of additions

Fig.31



COMPUTING POWER IN THE UNITED STATES

Fig.32

[1] Ware, op. cit., p. 14.

All indications point to further increases in speed and decreases in size and cost for computers for at least the next five years. What technological developments have enabled such rapid advances?

Hardware

The first computers built in the 1930's were mechanical devices using several thousand electro-magnetic relays as switching devices in the Central Processing Unit. This type of relay had very definite physical limitations with regard to speed and accuracy. In the 1940's the first electronic vacuum tubes were used instead of the mechanical relays and increased both the speed and accuracy of computers.[1] However, these vacuum tubes had limitations of their own. They took up much space, they had to be heated up before operating and they were prone to burn out. The development of "solid state" devices in the late 40's led to smaller and more rugged computers. The most important of these devices was the transistor, which was developed by the Bell Telephone Laboratory in 1948, and, which replaced vacuum tubes as switching devices in computers. These transistors were very small, could be used cold, and could not burn out.[2] Solid state technology advanced into the mid-sixties when the typical switching device was, "a customized fully integrated circuit (or micro-

[1] Bernstein, op. cit., p. 51.

[2] Ibid., p. 68.

circuit)... [which] is roughly ½ by 5/8 in., and contains a total of 800 transistors and 350 other electronic components. The density of electronic components in these integrated circuits is about three thousand per square inch."[1] The latest generation of computers contain switching devices with 200,000 circuit elements per square inch.[2] These rapid advances in solid state switching devices has accounted for part of the increasing speed and decreasing size and cost of computers.

The development of a second solid state device, the magnetic core in 1950, has led to further decreases in size, larger capacity, and speedier retievals. In the late-fifties, the magnetic core of the computer's memory consisted of tiny rings of magnetic material (a few hundredths of an inch in diameter) strung together in a plane with four wires passing through the cemter of each ring. Three of the wires were for changing the orientation of the core (clockwise or counter-clockwise) by passing electrical impulses through the wires to "write" information into the memory bank. The fourth wire was used for "reading" information out of the memory bank. The "read-write" operation took two-millionths of a second.[3] By the late-sixties, the X-55 magnetic core had been developed, which is a ring of magnetic material with a 7/1000 inch inside diameter and a 12/1000 inch outside diameter threaded by the four wires. "To uniformly cover an area just 1 by 1 in.

[1] W.H. Ware, "The Computer in Your Future," RAND Corp. Report P-3626, p. 4.

[2] *Ibid.*, p. 6.

[3] Bernstein, *op. cit.*, p. 69.

would take about seven thousand such cores. These cores are assembled on a grid of wires to form what is called a magnetic core plane....A number of wires go through the hole in each core...."[1] Each magnetic core is capable of storing a "bit" of information in the form of a binary digit (for explanation see Software, below).

Other forms of storing data include magnetic disc storage and magnetic tape storage. Both of these forms of storage are pieces of equipment just outside the main computer which can readily be scanned for information by the computer. Information files are more likely to be stored on magnetic discs or tapes than on magnetic cores. The magnetic disc resembles a stack of phonograph records and can store from 60-200 million binary digits per disc depending on the size of the stack.[2] In the past five years, a new technique for imprinting data on magnetic tape has greatly increased its storage capacity. This new process uses a laser to burn minute holes in the opaque coating of plastic tape. A hole signifies the binary digit "1" and no hole signifies the binary digit "0". Electronic Design magazine reported the importance of this process in 1966 as follows:

> ...the new process is unique in the following way:
> * The capacity is 645 million bits per square inch. This compares with about 5600 bits per square inch for standard magnetic tape.

[1] Ware, "The Computer in Your Future," p. 8.

[2] Ware, "Future Computer Technology," p. 9.

> ■ The speed-recording rate is 12 million bits per second.
> ■ Permanence-holes cannot be erased and do not fade.
> ■ The accuracy of the recording is verified instantly, and an alarm indicates lack of correspondence between input and output.[1]

The importance of this discovery in the information processing field is startling. Alan Westin writes:

> One small unit, containing one 4,800 foot reel of one inch plastic tape, will be able to store in digital form about twenty pages of information (250 words of typing to the page) for every person in the United States, including women and children. Specific information from a person's twenty-page dossier on this reel could be retrieved in a maximum search time of four minutes, and the entire dossier could be printed out for dispatch to an inquiring source in a matter of a few more minutes.[2]

A more expanded estimate wasdgiven by Burton E. Squires, Jr., a visiting assistant professor of Computer Sciences at the University of Illinois. This estimate was given in testimony before a House subcommittee in 1966:

> A high-speed computer memory now under development can read and write electronically at the rate of 16 million characters per second. A typical 300 page book containsxabout 1 million characters.... Thin films and magnetic tapes are normally used for high capacity memories. A piece of magnetic tape about 0.0015-inch thick and 1-inch square, attached to a computer, can hold up to 3,200 alphabetic characters that can be read at rates exceeding 100,000 characters per second. This media packs information at a density of about 1½ million characters per cubic inch. Thus, a building, containing 10,000 square feet of storage space 10 feet high, could

[1] "Memory Process Puts 645 Million Bits on a Square Inch," Electronic Design (Dec. 6, 1966), p. 21.

[2] Westin, *op. cit.*, p. 167.

> conceivably store a book of information about every man, woman, and child in the United States. Specific information about any particular person could be transmitted along any given telephone line within a few minutes.[1]

The technological advances in these hardware swithhing and storage devices have enabled an enormous amount of information to be stored and retrieved at speeds of up to a few nanoseconds (billionths of a second) per bit of information. This speed is even more startling whèn one realizes that the relationship between one nanosecond and one second is roughly the same as between one second and <u>30 years</u>. In the absence of adequate safeguards, such computer capabilities pose serious threats to the security of private information. I now turn to the developments in software which have tried to keep pace with these rapid advances in hardware.

Software

hThe internal language of the computer is the Binary System. All data is stored and manipulated in the form of the two binary digits, "0" and "1". For example, the numbers 0, 1, 2, 3, 4, 5 and 6 are represented in the binary system by 0, 1, 10, 11, 100, 101, and 110, respectively. The binary

[1] Burton E. Squires, Jr., testimony before subcommittee of the House Committee on Government Operations, July 27, 1966, in U.S. 89th Congress, 2nd Session, <u>The Computer and Invasion of Privacy</u>: Hearings, p. 136. (Further testimony from these hearings will be cited hereafter as <u>House Hearings</u>.)

system is the most sensible for computers because it can easily describe the on-or-off state of the hardware devices within the computer's units. A relay is either open (1) or closed (0); a vacuum tube either on(1) or off (0); a transistor either polarized (1) or uniform (0); and, a magnetic core either oriented clockwise (1) or counterclockwise (0). All of these devices are bi-stable in that, they are in some sense either on or off. Therefore, a "bit" of information refers to "the fact that a bi-stable device is on or off."[1]

Early users of computers who wanted a problem solved would first go to a professional programmer and explain the problem. After understanding the problem (which could take some time), the programmer would then translate the problem from English to the binary language which the computer could understand. Computer manufacturing companies and groups of large users began developing standard languages, which the user could insert directly into the computer. Inside a part of the computer's memory called a compiler, there would be a translating program, which would translate a program written in one of these standard languages to a set of instructions in binary language for the machine to follow. In this sense, "a programming language is an artificaal language that

[1] Bernstein, *op. cit.*, p. 67.

is structured in such a way that a user of a computer can communicate with the machine as a tool."[1] Between 1950 and 1961 over 2 billion dollars was spent in the development of programming languages. Many different languages resulted from this research, but the most frequently used languages are FORTRAN (Formula Translation), which is used mostly for scientific and engineering problems, and, COBOL (Common Business Oriented Language), which is used in data processing for such things as payrolls and inventories.[2] Olaf Helmer characterizes the Second Computer Revolution as, "the relative automation of the computer in the sense of doing away with many of the cumbersome aspects of computer programming and thereby facilitating direct communication between the individual researcher and the computer."[3] The gap between programming languages and the English language has been steadily narrowing and the new English-like programming languages will better be able to serve the new generation of computer uses, especially, the storing and retrieving of personal records, files and dossiers. But with the widespread and easier use of programmming languages, there also exists the more easier misuses of stored information. The hub of this new generation of uses and misuses is the recently-developed "time-sharing" system.

[1] Ware, "The Computer in Your Future," p. 18.

[2] Bernstein, op. cit., p. 75.

[3] Helmer, op. cit., p. 9.

Time-sharing and Useful Uses

"The major advancement in spreading the access of the computer to an even wider segment of society is 'time-sharing.'"[1] Eowever, time-sharing is also the area most vulnerable to illegitimate access to information files. Therefore, the remainder of this discussion will focus on the time-sharing system, its uses, misuses, security, and costs.

Time-sharing

Paul Baran, a computer and privacy expert for RAND Corporation, gave the following description of time-sharing durtng the House Hearings on the Computer and Invasion of Privacy:

> Both the form and uses of computers are now undergoing radical changes. They have become so powerful, can store so much data, and process this data so quickly that it becomes possible to ttime-share' a simple computer....Time sharing means literally that. Many people have access to the single computer installation. The computer has so much capability and is so fast that it creates the illusion that each user has his own computer.[2]

The physcial equipment of a time-sharing system consssts basically of the central computer, typewriter-like terminals, which can be located hundreds or even thousands of miles from the central computer, and communicationllines, which connect

[1] Paul Baran, "The Coming Computer Utility--Laissez-Faire, Licensing or Regulation?", RAND Corp. Report P-3466, p. 5.

[2] Paul Baran, House Hearings, testimony on July 27,1966, p. 122.

the terminals to the computer. When a person at a remote terminal wants to use the computer, he first "calls" the computer (by a special Bell telephone). The computer then indicates that it is ready. The user next types his account number and name into the computer, which checks the number and name with those in itbs memory. If the account number is valid (a password is needed in some systems), then the user has a direct connection to the computer banks. More specifically, the internal operation of a time-sharing system can be described as follows:

> The operation of an on-lіne time-shared information system is controlled by a set of master programs--the operating system. The basic tasks of the operating system are to:
> 1. Receive access requests from terminals and verify that the user is authorized for access--possesses a valid account number and/or password.
> 2. Control all communications with the terminals.
> 3. Schedule time slices to user programs or information requests.
> 4. Provide protections to users' programs and data (and to the operation system itself) against inadvertant destruction by other users.[1]

Useful Uses

While some experts predict major contributions from the computer in such areas as Government, Law-Enforcement, Education and Health,[2] time-sharing systems are already being

[1] H.E. Petersen & Rein Turn, "Security of Computerized Information Systems," RAND Corp. Report P-4405, p. 3.

[2] Armer, "Computer Aspects," pp. I-224--I-226.

used for a number of purposes. They keep insurance records, check automobile tags, locate outstanding criminal warrants, and make it more difficult to pass abbad check. The categories of information stored in remotely accessible time-shared computer systems include personal data on millions of individuals, proprietary industrial data, trade secrets, bank accounts and stock market transactions.[1] However, the use of the computer has been increasing rapidly in the areas of Government operation and law-enforcement, which has led to the controversial proposal for a National Data Center.

This National Data Center would pool all the information and data, gathered by the numerous government agencies, into the central computer of a large time-sharing system. There are some real benefits which are to be derived from this type of system, including:

> 1. Cost savings should result from the elimination of duplication, from increased productivity in information processing, and from better utilization of resources through planning. More comprehensive, current, accessible, and accurate information should result in better decisionmaking.
> 2. Collection of more revenues should result through better audit, followup, and uniform application of complex rules.
> 3. Services to the public should be improved and expanded. This benefit is most important although difficult to measure.[2]

[1] Petersen & Turn, op. cit., p. 2.

[2] Armer, "Computer Aspects," p. I-221.

The cost savings for a California Statewide Information System were estimated as a reduction of personnel costs by 1 to 5 percent, which would easily be enough to pay for the system. The total cost savings per year were estimated at 10 times the annual costs of the system.[1] For a law-enforcement system using a time-shared computer, which was installed in the counties around San Francisco, it was reported that, in the first year of operation, an additional two million dollars in revenues was collected through the computer's better audit of taxes and traffic fines.[2]

Aside from the strictly monetary benefits, the government and society would benefit from a much more efficient bureacracy through the use of computer systems. Arthur Miller, in The Assault on Privacy, writes of this increase in efficiency:

> In addition, statistics are becoming increasingly crucial as a foundation for the social and economic research, policy-making, and environmental planning that go into the administration of federal programs. In these contexts, the computer's ability to manipulate huge bodies of detailed information concerning a large number of potentially relevant variables that may pertain to events occurring over long periods of time permits the testing of hypotheses in ways that have never been feasible. Without modern electronics, planners might wander aimlessly in the federal government's paperwork jungle.[3]

[1] Armer, "Computer Aspects,", p. I-221.

[2] Ibid., p. I-221.

[3] Miller, op. cit., p. 55.

An efficient management of Government programs could match the unemployed with job openings, help in the legislative process, and even expose some unsystematic and wasteful administrative procedures.[1]

While these benefits are somewhat impressive, it is necessary to compare them with the possible misuses of stored data, and with safeguards which are technologically and economically feasible, in order to determine whether such a centralized information system is desirable and, if so, how it is to be safeguarded.

Snooping and Misuses

There are several major points in a typical remote-access time-sharing system through which improper access to data may be obtained: (1) the physical files themselves; (2) the transmission lines between terminal and computer; and, (3) the operating system or monitor or control program.

Physical Files

The physical files themselves are susceptible to theft, copying or destruction during the several stages of their processing. After information is first collected, it must be transcribed from alpha-numeric form to machine-readable form, which means additional handling of the infor-

[1] Armer, "Computer Aspects," p. I-221.

mation. Also the punch cards and the magnetic tapes and discs are more susceptible to theft and duplication than are paper files. For example a single magnetic tape containing 50 million bits of data can be reproduced in a matter of minutes and without a trace of copying or tampering in the original tape. In addition, it is easier to destroy computerized information than paper files or record books. A simple magnet can erase magnetic tape. Finally, because computers are so intricate and delicate, a speck of dust may render them inoperative or make them function erratically. Consequently, data may be lost, distorted or misdirected to unauthorized users.[1] In their RAND Corp. Report, Petersen and Turn summarized these problems concerning the physical files:

> Storage of information in computerized form allows rapid retieval and updating of files and drastically reduces the required storage space. However, information previously in the form of printed documents in locked file cabinets is now replaced by magnetization patterns on tapes and discs--they can be anonymously read, altered or erased without a trace of evidence that this has occurred. Hence, anyone that has gained access to the information system could, in principle, manipulate any information in the files--perhaps plant damaging information on a competitor, change bank accounts or copy trade secrets.[2]

[1] Miller, op. cit., pp. 27-29.

[2] Petersen & Turn, op. cit., p. 2.

Transmission Lines

The communications lines, especially in a geographically dispersed time-sharing system, are particularly vulnerable to illegitimate access. In a long-distance system, the terminals are connected by communications lines to the switching center which,in turn, is connected by lines to the central processor, and wiretapping is possible almost anywhere along the hundreds of miles of telephone wires.[1] "Passive eavesdropping by wire tapping is a low cost approach to copying all information communicated over the line. It is necessary to gain physical access to the communication lines and sort out the correct wires. A pickup device, tape recorder and a terminal (or equipment that can emulate the terminal) are required for recording and uncovering the information."[2] Because of the difficulties and high costs of safeguarding transmissions lines (see Safeguards, below), these lines are the weakest link in the time-sharing system and they provide the "snooper" with a relatively easy and cheap means of access to the information flow. If this information flow were to contain a large amount of sensitive information, then the snooper is likely to make large profits at the expense of the individual's or business' privacy.

[1] Armer, "Social Implications," pp. 12-14.

[2] Petersen & Turn, op. cit., p. 3.

The Operating System

The third area of improper access concerns the operating system or central processor, which is the very center of the time-sharing system. Paul Armer noted the importance of the central processor when he wrote:

> The central processor represents the acme of system security since it not only has access to all files and programs, but also to necessary passwords, file access authorization lists, and keys to the encryption schemes. Clearly, here, we need the highest available level of protection.[1]

A similar view of the control program's importance is that of Arthur Miller:

> The key software item of a time-share system--the monitor or control program--seems to be particularly vulnerable to intrusion. It can be duplicated so that false signals are generated, altered to permit unauthorized people to enter the system, or destroyed. Once access to the control program is gained, the intruder has the ability to display and manipulate any part of the data stored within the system.[2]

A snooper, who wishes to gain access to the operating system usually acquires the cooperation (through bribery or coercion) of a person or persons who normally gain access to the control program and to the computer hardware--systems programmers and maintenance engineers. In the course of their work, these systems personnel will insert into the monitor

[1] Armer, "Social Implications," p. 13.

[2] Miller, *op. cit.*, p. 29.

program a circumvention scheme or "secret door", through which the snooper may gain access to all files.[1]

Before I turn to a discussion of specific safeguards, I would like to look briefly at the possible misuses of personal information and at the relationship between the individual and his dossier stored in some remote computer, in order to gain some perspective of the seriousness of the privacy question.

Misuses and Errors

Blackmail is the most obvious but not the only misuse of stored personal information. A potentially more harmful effect could be caused by "inferential relational retrieval," in which a chain of relationships or inferences are drawn from a group of isolated facts.[2] For example, if a politician were seen at a restaurant dining with a political friend and with a reputed underworld figure, then the politician might be considered "associating" with criminals and his career may then be in jeopardy. However, the truth may be that the politician was unaware that the underword figure was going to dine with them and could not easily leave when he found out. Other misuses may arise as a consequence of

[1] Armer, "Social Implications," p. 14. (See also Petersen & Turn, op. cit., p. 3.)

[2] Paul Baran, "Communications, Computers and People," RAND Corp. Report P-3235, p. 11.

inaccuracies in the stored information. These inaccuracies may be of three types. Firstly, errors made by the key-punchers or programmers while transcribing the information into machine-readable language. Secondly, errors of context which occur when the raw unevaluated information is recorded in a cryptic form without recording any surrounding, extenuating or explanatory circumstances. And, thirdly, errors arising when ameliorating or supplementary information is not included in the dossier. For example, when a "non-conviction" is not entered following the recording of an "arrest".[1]

Many individuals are probably unaware that an information file on them exists, and if they were aware, they would naturally want the information to be handled properly so as not to inflict harm upon them. W.H. Ware made the following proposals concerning the individual and his file:

> Given the initial assumption that data banks serve useful purposes for the public, are cost effective, and will be in existence, it follows that each individual wants to make certain that: 1) information in the bank about himself is correct; 2) information is divulged only to those who will use it in his interest or to his benefit; and, 3) he has recourse for damages in the event the users or operators of the data bank willfully or negligently mishandle the information.[2]

A system set up along these three points will incur some costs, but so will any system of safeguards. The following section will contain a discussion of available and potential safeguards and their costs.

[1] Miller, op. cit., pp. 32-33.

[2] W.H. Ware, "Computer Data Banks and Security Controls," RAND Corp. Report P-4329, p. 6.

The problem of supporting or opposing a so-called National Data Center is understanding what such a center would encompass. A National Data Center basically conderned with statistical information, as was originally proposed, is highly desirable for reasons of efficiency and economy, and does not pose a serious threat to an individual's privacy. However, a full-scale National Data Center containing an up-dated dossier on every American man, woman and child is potentially dangerous and should not be constructed. Aside from the possible misuses of the highly-personal data itself, a full-scale, centralized information file could have grave psychological affects on many Americans. A person may be fearful that any action might be permanently entered on his record and possibly ruin his future. A fear of this type is an infringement of a person's freedom of thought, action and creativity and is harmful to our entire society.

Safeguards

There are three categories of safeguards corresponding to the three major points of improper access: (1) protecting the physical files; (2) securing the transmission lines; and, (3) safeguarding the central processor. However, there are two technological realizations, which must be made before implementation of an adequate safeguarding system can be accomplished. "First, it must be understood that the problem is a system one, which must be attacked from a system engineering point of view in the broadest sense. If handled in a bits-and-pieces fashion, the finest of safeguards in one part of the system can easily be circumvented by loopholes elsewhere."[1] And, secondly,

> The normal protective features of a remotely accessible information system are not designed to resist sophisticated penetration attempts. For increased security they must be augmented by additional programmed procedures or electronic devices. The objective is not absolute security --this can never be achieved, but rather an increase of the cost of penetration...to a level where the expected payoff becomes relatively small. At the same time, a balance must be maintained between the cost of countermeasures and the value of the protected information.[2]

What specific safeguards are available or potentially available under these three categories?

[1] Ware, "Computer Data Banks," p. 3.

[2] Petersen & Turn, *op. cit.*, p. 4.

Physical Files

The physical files are the easiest of the three areas of a time-sharing system to protect. There are a number of anti-intrusion devices already developed, which the manufacturing compamy or the system operator would install to protect the expensive machinery. These include: (1) electrical burglar alarms; (2) electromagnetic, optical, or acoustical barriers for protecting an area; (3) electronic motion detectors, whtch can register even the slowest motions; and, (4) vibration detectors. The output signals from these devices could be connected to local alarms or would activate automatic telephone dialing equipment to notify security forces.[1]

Transmission Lines

As I mentioned above (see Page 18), the communications lines of a time-sharing system are the weakest and most accessible link in the system. The most effective safeguard for these lines is also the most expensive and least feasible; namely: the cables can be buried and armored with locked and alarmed terminal boards and manhole covers.[2] An alternative safeguard which is more feasible but which can also incur an overriding cost factor is encryption of the transmitted data. Simply stated this entails some sort of

[1] Petersen & Turn, op. cit., p. 5.

[2] Ibid., p. 4.

"scrambling" or "garbling" of the information, so that, if it were intercepted, it would seem unintelligible. More specifically, there are two basic classes of encryption:

> 1. *Substitution* of the characters in the message with other characters or groups of characters. The replacement characters may come from an alphabet different from that used for the message.
> 2. *Transposition* of the sequencing of the characters in the message.[1]

The set of rules that specify the particular substitution or transposition is called the "key" or "code". In the simplest type of cryptographic system the data is scrambled at one end by adding the key, transmitted in its scrambled form, and unscrambled at the other end by subtracting the same key. But the simplest encryption system is also the easiest to crack, and will not prevent much snooping as long as highly-sensitive data (highly-profitable to the snooper) is being transmitted. Snoopers would be willing to incur the costs of breaking simple codes for the profitable returns on the data. More complex systems can be devised which use switching stations along the way to change the key, which can also be very complex. The most complex system is called "doubling encryption" because it uses both the simple end-to-end encryption and the switching-center encryption, along with elaborate keys.[2] These more complex systems are only economically feas-

[1] Petersen & Turn, *op. cit.*, p. 5.

[2] Paul Baran, "On Distributed Communications: IX: Security, Secrecy, and Tamper-Free Considerations," RAND Corp. Memorandum RM-3765-PR, pp. 10-16.

ible for use in military communications, where the highest degree of security is needed. As the complexity of the system increases, higher costs result from the use of larger and more complex keys and from the necessity of using more sophisticated transmitting and receiving equipment to handle these keys. However, in the time-sharing information system, there seems to be a dilemma: the simplest encryption system may be economically feasible but will not provide adequate protection; while a more complex system will provide better protection but is not economically feasible. The solution to this dilemma lies in the use of a simple encryption system while at the same time not transmitting highly-sensitive information over the lines. The profits to be gained from the less-sensitive information should not provide an incentive for the snooper to incur the costs of breaking the code. Until highly-sensitive information can be effectively protected in an economically feasible encryption system, this solution should be implemented.[1]

The Operating System

There are two problems associated with the safeguarding of the Operating System. One involves the actual hardware and software safeguards, and the other involves the

[1] "Computers: Safeguarding Time-Sharing Privacy--an All-out War on Data-Snooping," Electronics (February 6, 1967), p. 158.

prevention of the insertion of "secret doors" by systems personnel. The latter is beyond the scope of this paper but is nevertheless a very important problem, which could possibly be handled through the licensing of personnel or through the software safeguards which I shall now discuss. Safeguards in the central processor should cover three basic areas: (1) recognition of authorized users and terminals; (2) real-time monitoring and auditing of the system; and, (3) partitioning of the memory along a hierarchical range from highly-sensitive to public information.

The weakest area of security in the operating system, at present, is the recognition of authorized users at remote terminals. In the present system, a user must identify himself by name, account number and, in some cases, a password. A snooper can learn these facts from discarded printouts or from random guessing. Once the unauthorized user acquires these passwords, he has the same access to the computer banks as the authorized user. Brian W. Pollard, president of Radio Corporation of America's Electronic Data Processing Dept., suggests the following system of changing passwords:

> Passwords can be easily changed....Suppose a file were set up in such a way that a different password was required each time the file was used. Only the person authorized to use the file would know that it had been used, and he would know the sequence of passwords, so that he should have no trouble. Anyone else who perhaps learns the sequence wouldn't know which one to use at any

> particular time, and trying them all on a random basis would be easily detectable.[1]

For highly-sensitive files, a more effective safeguard would be a "fail-safe" system, which would require several users to be present with their individual passwords before access to the file is gained. However, a more effective safeguard should be available in the near future, recognition of fingerprints or voiceprints. In this system, the user's voiceprint or thumbprint would be scanned at the terminal, transmitted to the central processor, and compared with the prints of authorized users on file.

The second area of safeguards is "real-time monitoring" and random auditing of the system. Real-time monitoring could be very effective:

> A capability to continuously monitor the system activity--access requests by users, granting or refusing of access, status of the lists of current users and terminals--provides a further increase in the system's security. Attempts at deception can be detected when two users claim the same identity or when two identically labeled terminals are connected. Unusual activity in a file or abnormally large numbers of access rejections may indicate attempts to penetrate the system.[2]

A random external audit of the operating program could be made to check for "secret doors."[3] The monitor program

[1] Brian W. Pollard, quoted in "Computers: Safeguarding Time-Sharing Privacy," p. 158.

[2] Petersen & Turn, op. cit., p. 4.

[3] Baran, "Communications, Computers and People," p. 17.

could also be set up to review computerized records periodically, and to erase obsolete and outdated information.[1]

An important safeguard which should be implemented is the partition~~g~~ing of the memory banks. Partitioning would serve two useful purposes: (1) it would stop an unauthorized user who got past earlier safeguards and thereby gained access to the memory; and, (2) it would prevent authorized users from exploring off-limits sections of the files. Because partitioning can stop a snooper who has passed earlier safeguards, segmenting the memory with "bulkheads" can be considered the cornerstone of any safeguard system and should definitely be implemented. Professor Robert Fano, a computer expert at the Massachusetts Institute of Technology, is a strong advocate of the partitioning system:

> 'You cannot rely on a single wall of security,' Fano notes, 'there has to be a sequence of hurdles.' In the new system, for example, there will be no one inner sanctum which, once gained, will provide access to all information. Instead, there will be successive walls and partitions to be passed only under specified conditions.[2]

Fano believes that partitioning will also protect the system from mistakes. He says, "If you accidently find a hole in the program, you can go just so far before you are blocked by another partition."[3] Although partitioning seems to be a desirable and effective safeguard, it too incurs added costs.

[1] Miller, op. cit., p. 66.

[2] "Computers: A Question of Privacy," Electronics (2/6/67), p.38.

[3] Robert Fano, quoted in "Computers: A Question of Privacy," op. cit., p. 38. (For other descriptions of partitioning see "Computers: Safeguarding," op. cit., pp. 157-158.)

Costs

All of the above safeguards have attached to them some sort of cost, either in dollars and cents or in lost computer efficiency and storage space. Petersen & Turn report that, "Each increase in the protective features of the operation system is accompanied by a decrease in the efficiency of the information system as more and more computer time is diverted to these nonproductive tasks."[1] Walter Bower, an executive in a computing firm on the West Coast, "anticipates a day when 10% of a computer's memory will be devoted to routines needed to qualify users requesting information. In some applications...the figure could be as high as 20%."[2] Who is going to bear these added costs? Paul Baran believes that the manufacturers, operators and users of computerized information systems should regard "such costs as necessary costs--a price [paid] to society for the privilege of building a potentially dangerous system."[3] These added costs could be greatly reduced if they were part of the original design of new systems.[4] However, because of the overriding cost factor, there is no motivation for voluntary safeguarding by the information industry itself.

[1] Petersen & Turn, op. cit., p. 3.

[2] "Computers: Safeguarding," op. cit., p. 159.

[3] Baran, "Communications, Computers and People," p. 15.

[4] Petersen & Turn, op. cit., p. 7.

Federal financial assistance, either in the form of outright grants or tax incentives, to computer manufacturers and systems operators is a possible means of self-regulation. However, a program of financial incentives has too many loopholes and would not be as effective as Federal intervention. On a practical level, incentives would not affect Government computer systems unless Federal, state and local governments buy their computers only from manufacturers which build in safeguards. An incentive program would also offer computer manufacturers and computer systems operators a choice of installing safeguards or not, but the choice of protecting privacy should not be left to businesses whose main concern is profits. The major argument against an incentive program is the idea of paying companies for privacy. Privacy should be an individual's right and not a commodity to be bought with taxpayers' money. The computer information industry should be made responsible for handling personal information and for protecting the individual subject's privacy. Ideally, this sense of responsibility should be accomplished through self-regulation , but I feel that the problem of privacy needs strong action now to be able to cope with potentially bigger problems accompanying future technological advancements. Federal intervention (p. iv) would be the most effective means of protecting privacy now, while self-regulation is an ideal which hopefully will replace government regulation in the future.

Conclusion

The above discussion on safeguards was based on several assumptions. Firstly, that the General Trends in computing power will continue; secondly, that there are very real and beneficial uses for time-sharing systems; and, thirdly, and most importantly, that more centralization of information is inevitable. Because of the reluctance of the information industry to voluntarily safeguard; because of the interstate nature of the information flow; and, because of the potential dangers of misused information, the Federal Government must intervene and regulate the implementation of safeguards now, before unregulated centralization takes place. As Professor Fano remarked:

> You can never stop these things [computerization]. It is like trying to prevent a river from flowing to the sea. What you have to do is to build dams, to build waterworks, to control the flow.[1]

[1] Fano, quoted in Westin, *op. cit.*, p. 326.

VI. BIBLIOGRAPHY

Bibliographies Prepared by Other Writers

Harrison, Marie. *The Problem of Privacy in the Computer Age: An Annotated Bibliography*. Vol. 1. Santa Monica, California, RAND Corp. Memorandum RM-5495-PR/RC, 1966.

Books and Pamphlets

Armer, Paul. "Computer Aspects of Technological Change, Automation, and Economic Progress." (Reprinted by the RAND Corporation from *Technology and the American Economy*, the report of the National Commission on Technology, Automation, and Economic Progress. Appendix Volume I, *The Outlook for Technological Change and Employment*, U.S. Government Printing Office, Washington, D.C. February, 1966.).

———. "Social Implications of the Computer Utility." Santa Monica, August 1967. (RAND Corp. Report P-3642.).

Bernstein, Jeremy. *The Analytical Engine: Computers--Past, Present and Future*. New York, Random House, 1964.

Baran, Paul. "The Coming Computer Utility--Laissez-Faire, Licensing or Regulation?". Santa Monica, April 1967. (RAND Corp. Report P-3466).

———. "Communication Policy Issues for the Coming Computer Utility." Santa Monica, 1967. (RAND Corp. Report P-3685).

———. "Communications, Computers and People." Santa Monica, November 1965. (RAND Corp. Report P-3235).

———. "Remarks on the Question of Privacy Raised by the Advent of Automation." Santa Monica, 1966. (RAND Corp. Report P-3523).

Helmer, Olaf. "Prospects of Technological Progress." Santa Monica, August 1967. (RAND Corp. Report P-3643).

Maron, M.E. "Computers and Our Future." Santa Monica, 1966. (RAND Corp. Report P-3501).

Miller, Arthur R. *The Assault on Privacy: Computers, Data Banks, and Dossiers*. Ann Arbor, University of Michigan Press, 1971.

Petersen, H.E. and Turn, Rein. "Security of Computerized Information Systems." Santa Monica, July 1970. (RAND Corp. Report P-4405).

Ware, W.H. "Computer Data Banks and Security Controls." Santa Monica, March 1970. (RAND Corp. Report P-4329).

______. "The Computer in Your Future." Santa Monica, November 1967. (RAND Corp. Report P-3626).

______. "Future Computer Technology and Its Impact." Santa Monica, March 1966. (RAND Corp. Report P-3279).

Westin, Alan. *Privacy and Freedom*. New York, Atheneum, 1967.

Periodicals

"Computers: A Question of Privacy." *Electronics*. February 6, 1967. 36-38.

"Computers: Safeguarding Time-Sharing Privacy--an All-out War on Data Snooping." *Electronics*. April 17, 1967. 157-159.

"Memory Process Puts 645 Million Bits on a Square Inch." *Electronic Design*. December 6, 1966. 21.

Government Publications

U.S. 89th Congress, 2nd Session. Subcommittee of the House Committee on Government Operations. *The Computer and Invasion of Privacy*. Hearings, July 26, 27, & 28, 1966. Washington, Government Printing Office, 1966.