EDP ANALYZER

APRIL, 1976
VOL. 14, NO. 4

INTEGRITY AND SECURITY OF PERSONAL DATA

Some rather simple words in existing and proposed privacy legislation—words such as "accuracy," "integrity," and "security"—will pose a challenge to many data processing executives. The challenge already exists for U.S. federal agencies; it will probably confront the remaining public and the private sectors in the U.S. in the next few years. Further, many other computer-using countries have adopted or are considering privacy laws. The challenge that comes from these privacy laws occurs because these words cannot be put into practice as easily as might first be imagined. Moreover, if they are *not* put into practice, the threat of both civil and criminal penalties arises. This is not meant to be a scare discussion of privacy legislation. Rather, we are just pointing out what the privacy legislation says and what the consequent implementation problems appear to be. Based on what is shaping up, you may want to revise your data processing plans.

We will quote some selected portions of the Privacy Act of 1974 (Reference 1), which applies to agencies of the U.S. federal government and to private contractors acting as agents for those agencies:

PORTIONS OF PRIVACY ACT OF 1974

1. The opportunities for an individual to secure employment, insurance, and credit, and his right to due process and other legal protections are endangered by the misuse of certain information systems (preamble).
2. Federal agencies (will) . . . be subject to civil suit for any damages which occur as a result of willful or intentional action which violates any individual's rights under this Act (preamble).
3. Each agency . . . shall . . . establish appropriate administrative, technical, and physical safeguards to insure the security and confidentiality of records and to protect against any anticipated threats or hazards to their security or integrity which could result in substantial harm, embarrassment, inconvenience, or unfairness to any individual on whom information is maintained.
4. Whenever any agency . . . fails to maintain any record concerning any individual with such accuracy, relevance, timeliness, and completeness as is necessary to insure fairness in any determination relating to the qualifications, character, rights, or opportunities of, or benefits to the individual that may be made on the basis of such record, and consequently a determination is made which is adverse to the individual . . . the individual may bring a civil action against the agency, and the district courts of the United States shall have jurisdiction . . .
5. Any officer or employee of an agency who by virtue of his employment or official position, has possession of, or access to, agency records which contain individually identifiable information the disclosure of which is prohibited by this section or by rules or regulations established hereunder, and who knowing that the disclosure of the specific material is so prohibited, willfully discloses the material in any manner to any person or agency not entitled to receive it, shall be guilty of a misdemeanor and fined not more than $5,000.

The message of these quotes is clear. The agencies of the U.S. federal government are responsible for effective management of their data systems, including the integrity and security of the personal data in their files. They are to take appropriate steps to assure the integrity and the security of the records. Any willful or intentional action which results in harm to an individual, under this Act, is subject to a civil suit. Any willful but unauthorized disclosure of personal information can lead to a criminal penalty of a fine not exceeding $5,000. It is not clear whether a willful decision *not* to institute a security measure, which later leads to the unauthorized disclosure of personal information, will make the decision maker subject to this criminal penalty.

This is all very well, you say, but it only applies to agencies of the U.S. federal government. What does it have to do with me?

Well, the U.S. House of Representatives is considering H.R. 1984 (Reference 2), which would extend privacy legislation to state and local government units and to all private organizations. There is no telling if and when this legislation might be passed, or the form it might take before passage. Probably the Congress will wait for a year or two of experience under the Privacy Act of 1974 before deciding on legislation for the private sector; Representatives Goldwater and Koch say they will not push their bill (H.R. 1984) for awhile. But as H.R. 1984 is currently drafted, it has words that are quite similar to those in the Privacy Act of 1974. It calls for maintaining personal information with the accuracy, completeness, timeliness, and pertinence necessary to assure fairness in determinations. It calls for establishing categories for maintaining personal information to operate in conjunction with confidentiality and access controls. It imposes civil liability for unfair personal information practices, with damages to include actual damages, punitive damages where appropriate, and the costs of bringing the action. And for willful violations in the handling of personal information, criminal penalties can be imposed, of a fine not to exceed $10,000 for each instance or imprisonment of not more than five years, or both. And then there is a proposed federal privacy board, with broad-ranging regulatory powers.

We suspect that somewhat similar laws will be drafted and enacted in other computer-using countries. There seems to be a public fear of the computer, as representing a threat to privacy, even though the computer can be an ally in the protection of privacy.

So the challenge of meeting the integrity and security requirements is already here for agencies of the U.S. federal government. That same challenge may not be many years away for the rest of us.

Integrity of data

The Privacy Act of 1974 calls for federal agencies to establish appropriate safeguards to protect against any anticipated threats to the security and integrity of data records (which could result in substantial harm, etc. to the data subjects). H.R. 1984 says that an organization that collects, maintains, uses, or disseminates personal information should assure its reliability.

What is integrity of data? We have seen a variety of definitions, some of which confuse it or overlap it with security of data. As we use the term, it means undistorted by error—the condition of representing what it purports to represent.

As used in the context of data systems, integrity is somewhat different from accuracy—although our dictionary uses the synonym "truthful" for both. Accuracy is considered to mean: conforming exactly to the truth. Integrity is considered to mean: the *data system* does not inject error into the data. If erroneous data is entered into the system, the system acts with integrity if it does not inject more error into the data.

(But H.R. 1984 says that an organization must assure the reliability of the personal data—and synonyms for reliability are trustworthy and true. So the burden may be on the organization to see that it does not enter erroneous personal data into its systems.)

Most computer users already are taking steps to insure the integrity of data—that is, inhibiting the injection of error during initial recording, transmission, processing and storage, as well as during recovery from system failure. After all, errors in the data lead to errors of output, waste, wrong decisions, and such. But errors can never be completely eliminated, and every organization has found its own error tolerance level.

The big question is, of course, whether that error tolerance level—which has apparently been satisfactory for operational purposes—will be

good enough to meet the demands of the privacy legislation.

Security of data

Dr. Willis Ware, Vice Chairman of the Federal Privacy Protection Study Commission, uses the following definitions for security. *Computer security* is the totality of measures required to (1) protect a computer-based system, including its physical hardware, personnel, and data against deliberate or accidental damage from a defined threat; (2) protect the system against denial-of-use by its rightful owners; and (3) protect data and/or programs and/or system privileges against divulgence to or use by unauthorized persons. *Data security* is the safety of data from accidental or intentional but unauthorized disclosure, modification, or destruction. We would add to this definition by explicitly including the theft of data.

One important aspect of data security is access control. Access control includes the identification, authentication of identity, clearance, and "need to know" of each user who accesses data.

As with integrity, most computer users have taken some steps to safeguard their installations, their data files, and so on. The question is whether those steps will be adequate under the demands of privacy legislation.

What is the significance?

We believe that the great bulk of business and government transactions are handled satisfactorily; after all, organizations do continue to operate. When errors occur, they are generally corrected quite easily and no real harm is done. To the extent that this satisfactory handling of transactions occurs, privacy legislation would have relatively little effect.

The problem is the unusual case. Unfortunately, the unusual case cannot be anticipated nor identified in advance, most of the time. This is not a question of data system design that we refer to but rather one of human behavior. The unusual transaction generally is treated in the normal manner—until it suddenly explodes. The trouble is caused by differences of opinion, differences of interpretation, and so on.

Currently, it is hard to tell just how many of these unusual transactions occur and how much harm they cause. There are few statistics on them. In those instances where unusual transactions pertaining to personal information occurred, the file owners might just claim them to be "errors" and dismiss them—even though the data subjects might feel that they had been harmed.

However, should H.R. 1984 be enacted as currently drafted (admittedly, not too likely), then any person who violates the provisions of the Act would be liable for actual damages sustained by the data subject, punitive damages where appropriate, and court costs. Further, violations might be interpreted to include "insufficient" data integrity and security.

So, while many computer-using organizations may feel that their data integrity and security procedures are satisfactory because so many transactions are handled well, they may suddenly be confronted with law suits claiming large damages for a relatively few unusual cases.

If privacy legislation is enacted for the private sector, less than one year grace period may be granted before the law is enforced. In the case of the Privacy Act of 1974, the grace period was less than ten months. After the grace period, the *threat* of civil and criminal penalties comes into existence.

We are not saying that the civil and criminal suits will become an avalanche when the privacy laws come into force. But they can occur.

So the big question facing data processing management, in connection with the integrity and security of personal data, is: How much is enough, in order to meet the spirit of the law and avoid lawsuits?

"How much is enough?"

As mentioned, in the past the decisions on data integrity and security practices were completely up to the management of the organization. Each organization learned what integrity level it could tolerate and what security risks it could accept. When mistakes were made in these decisions, such mistakes usually led to loss of business—and in a small percentage of cases, to law suits.

But such decisions in the future will have to consider the privacy laws. Moreover, these laws do not say explicitly *how much* data integrity and security are required.

So how far should an organization go toward enhancing its data integrity and security prac-

tices, so as to bring the threat of lawsuits within acceptable bounds? Following is a range of alternative approaches that we will discuss briefly:

- Make no changes
- "Prudent man" principle
- Decision algorithms
- Accreditation of the installation
- Certification of the installation
- Licensing of the installation

Make no changes

With this approach, management decides that the present data integrity and security practices are adequate, and that no enhancements probably will be needed. In fact, the resistance to enhancements might be even more strongly expressed, such as, "We are doing OK now, so let's continue our practices; if we make a mistake, let them sue us; it will cost us less in the long run."

This appraoch obviously involves minimum extra cost for data integrity and security. It also probably has minimum credibility in court, in case law suits are entered under the privacy laws.

It seems to us that this is not a wise approach. The attitudes toward the handling of personal data are undergoing a change. We believe that most organizations (not just computer using organizations) will have to make some changes in their record keeping practices, including their data integrity and security practices.

"Prudent man" principle

Under this approach, the organization makes a threat analysis and risk analysis concerning the personal data in its files. (As Willis Ware points out, a threat is to the data and risk is to the organization.) It then decides which threats can be accepted and which ones will require additional safeguards, in order to protect the organization from lawsuits. The same criteria would be used as might be used by a "prudent man" who is concerned about protecting his own property.

One way of making the threat analysis and risk analysis is to perform a self-audit. The extensive AFIPS security checklist (Reference 3) would be helpful here, although it does not cover data integrity as thoroughly as it covers security. In addition, there are a number of other good quality checklists and manuals that pertain to integrity and security. Since there are so many and since we do not have space to give them adequate treatment in this report, we have prepared a special annotated bibliography on the subject (Reference 4). By the use of such material, an organization can audit its own integrity and security practices, to see what coverage is afforded. It is likely that outside sources will have to be consulted to determine the magnitude of the threats—but we do not have any leads to sources of good information on threats at this point in time.

Another way of making the threat and risk analysis is to use an outside organization, such as an auditing firm, a security consulting firm, or such. Such a firm would probably use the same type of checklist mentioned above, but based on experience with other clients, might be able to point out threats and risks that the client might miss. Such an organization probably would report on what is being done and not being done, and might make recommendations on what should be added or changed. But the organization probably would not "certify" the data integrity or security practices and would probably disclaim any legal responsibility.

It is at this point where the "prudent man" principle would be applied. Once the threats and the risks had been identified, the decision would have to be made on what to do and what not to do.

The main shortcoming of this approach is that it is wide open to differences of opinion. Whether an outside organization is used or not, the decisions on what to do and what not to do are largely the opinions of a relatively small number of people—those making the analyses and the decisions. To what extent the courts will side with the "prudent man" decisions is a matter of conjecture.

Decision algorithms

We have seen this approach discussed in the literature but have not seen instances of its use. It is very similar to the "prudent man" approach, except that formalized, quantitative methods are developed for making the decisions on what to do and what not to do. These methods will probably use the cost/benefit approach.

As with the "prudent man" approach, a threat and risk analysis should be made, either by a self-audit or by using an outside organization.

The results of this analysis would then be entered into the decision algorithms, to decide what

should be changed in the data integrity and security practices.

A major shortcoming of this approach would be the same as for the "prudent man" approach. The decision of which practices should be used and which not used would still be a matter of the opinion of a small number of people. How much weight these opinions would carry in court is hard to say.

Accreditation of the installation

This approach requires that some independent agency develop standards for data integrity and security practices—although no such standards yet exist nor are they likely to appear in the near future. To be most useful, the agency should either represent the federal government or the industry-wide computer field. The agency would then appoint teams to inspect computer installations and evaluate their data integrity and security practices relative to the standards. The accreditation, if issued, would imply no legal responsibility on the part of the agency if the team failed to detect and report some sub-standard practices.

Webb and Abbott (Reference 5) have proposed that an Accreditation Commission be formed, similar to the Joint Commission on Hospital Accreditation, either by the federal government or by the computer industry. This commission would have two main responsibilities—to develop standards and to perform accreditation evaluations.

Standards. The commission would supervise the development of a data integrity norm, say Webb and Abbott, consisting of general principles, standards that support and define these principles, and guidelines that explain and give practical examples of applying the standards. These standards would then represent professionally developed and nationally applied criteria for data integrity.

Brian Ruder, of Stanford Research Institute, in a letter to us, suggests that *categories* of computer systems/centers be established, so that standards can be developed that are germane to a particular environment. For example, there might be different standards for the accrediting of large and of small organizations.

Accreditation. The commission would monitor the data integrity at computer centers by developing and applying a formal accreditation methodology, say Webb and Abbott. The commission would select groups of qualified people to visit data centers and make evaluations of the data integrity practices at those centers. Such evaluations could be done as frequently as daily or as infrequently as once every two to five years, depending upon the specific needs. Moreover, the evaluations could be mandatory for some sites and voluntary for others. Unlike certification, such accreditation would not imply a legal responsibility on the part of the accrediting agency.

In connection with accreditation, Ruder asks two important questions. For one thing, what sanctions should be imposed on those centers that fail to be accredited? And what motivation would there be for computer centers even to seek accreditation? One answer to the first question, he says, is that a public pronouncement may be sufficient to force the organization to seek compliance. And an answer to the second question may come from lawsuits claiming negligence or noncompliance.

So people are thinking about the development of standards for data integrity (and, we should hope, for data security). Those standards do not yet exist. But the new privacy laws may well provide the stimulus for their development.

Also, it is quite possible that if such data integrity and security standards did exist, casualty insurance companies might be willng to insure against civil penalty losses for accredited organizations. Question: would they be willing to insure in the absence of standards? And have they even been asked yet to provide such insurance?

Thus, with accreditation, a new element enters the picture—standards which are professionally developed and nationally applied. The decisions on what integrity and security practices to use would then be based on the opinions of a large number of qualified people. It seems to us that the courts would attach much greater weight to such standards than they would to the decisions of "prudent men." The courts might still rule in favor of the plaintiffs, but it would seem less likely than in the case of the "prudent man" principle.

Certification of the installation

As with accreditation, the certification of computer centers would involve the development of standards. Authorized individuals or organiza-

tions would then inspect data centers and certify (or fail to certify) that the centers meet the standards. The model that might be applied here is that of the certified public accountants who perform audits of client organizations and then certify that the financial statements truly reflect the financial conditions of those organizations "according to generally accepted accounting practices."

One way in which certification differs from accreditation is that the certifying individual or organization has a legal responsibility. If there is a failure to detect sub-standard practice, or if something is hidden that should have been disclosed, the certifying individual or organization can be sued. Because of this legal responsibility, certification would generally require more time and would cost more than would accreditation.

Some people are suggesting that the certified public accountants undertake the certification of data integrity and security, under the privacy legislation. An executive of the American Institute of Certified Public Accountants says that this is a function the CPAs are not seeking; it may be forced upon them but they do not really want it. Not only is a new technology involved (which many CPA firms are qualified to handle by way of their management service division consultants) but also it exposes the CPA firms to additional possible lawsuits.

We understand that most lawsuits today against CPA firms are of the malpractice variety, where the plaintiffs claim that the certifiers failed to apply the "generally accepted accounting practices" properly. But we also learned that some of the lawsuits are directed at those generally accepted accounting practices themselves, claiming that the practices are wrong or inadequate.

So, even if standards of data integrity and security are developed, this is no guarantee that they will not be challenged in court. But we suspect that the courts will usually side in favor of the standards.

Licensing of the installation

One possibility for control by licensing is that personal data could neither be collected nor maintained unless the organization obtained a license from the government to do so. And a license would be granted only if the organization agreed to meet standards of integrity and security. Further, the license could be withdrawn if violations could be proved.

Such licensing of installations is not being seriously considered in the U.S., to our knowledge. We understand that in Sweden, however, all mechanized files of personal information must be licensed.

Possible steps to take

Suppose privacy legislation for the private sector does come within the next 15 to 18 months (from state and/or federal laws). It is not too likely that data integrity and security standards will be developed by that time. What should an organization do to protect itself against lawsuits under the privacy laws?

Some people may think that it is too early to begin such consideration. The private sector privacy legislation may not actually materialize for years, and it may be greatly toned down when it does, based on the difficulties in the federal government of complying with the Privacy Act of 1974.

However, we do not think it is too early to begin such consideration. It would be wise, we think, to at least do some preliminary thinking and to begin exchanging ideas with other data processing executives.

One way to begin is to make a preliminary threat and risk analysis of the personal data in your files. We suspect that most of the problems will be related to the manually maintained records in the personnel department of an organization. But for the personal information in the mechanized files, it is worth asking how the subject individuals might be harmed, embarrassed, inconvenienced, or treated unfairly if that data were to be: (1) inaccurate, (2) irrelevant, (3) incomplete, (4) out of date, or (5) disclosed to someone who is not authorized to see it (assuming that the restrictions imposed by the Privacy Act of 1974 were applied to the private sector). We are not implying that this is an easy task; it is often hard to visualize what could happen and one tends to try to remember what *has* happened. But if a number of people start thinking about the problem and exchange their ideas, the threats and risks may start to come into focus.

The next step would be to rank the threats and risks in order of priority. Where is the risk greatest to the organization, as a function of the threats, the probable effectiveness of the safe-

guards, and the possible loss to the organization if a violation occurs?

You might find that it would be possible to enhance your safeguards where the risks are greatest by working on improvements at the same time other changes are being made.

In short, we are suggesting that you consider following the "prudent man" approach at this point in time, since standards of data integrity and security are not yet available. You might be able to make significant progress at a not-great expense. However, if you wait until the privacy legislation is passed and have to make all the changes during a relatively short grace period, the costs of the changes may be higher. We suggest such action for two reasons. For one thing, it seems like a sensible thing to do to treat personal data carefully. And secondly, we believe that the odds favor privacy legislation, at either the state or federal levels, being enacted in the next two years or so that will affect the private sector. Any steps that you can take ahead of time, at not-great cost, will save you time and money when the legislation is finally enacted.

Possible integrity enhancements

Over the years, we have treated the subject of data integrity in a number of reports. In fact, our very first issue, in February 1963, was on error detection and control. So we will not attempt much of a discussion here but rather will mention certain types of changes and reference our past reports.

One step that we discussed in the November 1975 report was to develop data definitions for personal data, for meeting the privacy requirements. These definitions would include the accuracy, relevance, completeness, and timeliness characteristics. Some day there may be nationally applied standards in this area. For the present, organizations will have to rely on "prudent man" decisions on these characteristics.

A next step might be the redesign of forms and procedures for the collecting of personal data, in line with the data definitions just mentioned. If it is apparent that some piece of data is not really relevant to the purposes of the organization, the collection of that data can be stopped. Or it may be evident that much more care must be taken in collecting (and updating) some piece of personal information—so that when that data is used for its intended purposes, the individual is less likely to be harmed, embarrassed, or such.

The whole process of data entry for personal data, including the detection, correction, and reentry of erroneous items, might well be reviewed. We discussed data entry in our September and October 1971 reports.

The internal control system that is used for personal data could be reviewed. We treated internal control in our June 1967 and March 1975 reports.

The backup and recovery system, for recovering from system failures without damaging personal data, could be considered; we discussed this subject in our October 1968 and January 1972 reports. Also, the long term retention of personal data should be re-thought; we considered the mechanics of long term retention in our July 1973 issue.

At the "privacy mandate" conference, sponsored by the U.S. National Bureau of Standards and the Mitre Corporation in April 1975 (Reference 6), some suggestions were made for improving data integrity. One suggestion was to eliminate concurrent program development and data base accesses—since a program under development may inadvertently damage data in the data base. We have visited installations that do have such concurrent operations and which have reported no troubles of this type. But in fact, the threat is real and non-trivial, so the suggestion should be considered. Another suggestion was that dedicated equipment be used for particularly sensitive applications, both for integrity and security reasons. With today's mini-computer systems, this step may be more economically feasible, although it can play havoc with the integration of systems.

Dr. Douglas Webb, of Lawrence Livermore Laboratories, has pointed out in a letter to us that previously accurate data can be changed by programs. Hence, the question of accuracy and integrity of *programs* must be considered, as well as of data. Since one program in a multi-programming job stream can change another program, and the other program can change the data, the problem can become very complex to control.

There are a number of things, then, that can be reviewed in the way that personal data is handled and stored. Any improvements that can be made, perhaps when other aspects of these systems are being changed, will make the eventual com-

pliance with privacy laws that much easier.

The real quality of data

Up to this point, we have been addressing the subject of data integrity from a conventional standpoint. We have tended to pick up the subject where data has been recorded on source documents, and have then considered how errors can be minimized in the following steps. Most of the error control procedures have to do with the transcription and transmission of data.

Ivanov (Reference 7), in a doctoral thesis prepared at the University of Stockholm and Stockholm Institute of Technology (Sweden), points out that this conventional viewpoint has a glaring error in it. Error does not begin with the recording on a source document, he says. The basic question is: how well does the data in the system reflect truth and reality? A conventional analysis of error control does not begin at the right point and does not consider the right questions, he says.

The quality of data is determined by two major parameters, he claims—its accuracy and its precision. *Accuracy* is an indicator of truth and relevance for multi-purpose usage. It is measured by the amount of disagreement among several independent methods of measurement—and moreover, methods that have been selected on the basis that they are likely to give the greatest disagreement of measurements. Accuracy is affected by what is *not* under the control of the decision maker who will use the data; rather, it is controlled by the environment. Also, accuracy is characterized by a lack of bias.

Precision, on the other hand, is an indicator of the repeatability of measurements. It is the consistency (or lack of it) of one method of measurement repeated numerous times. Precision *is* affected by what is under the control of the decision maker—his operations of measurement, transcribing, values of attributes, and the stability of his processes. Most of the conventional concern with error control has dealt with precision and true accuracy has been largely ignored, as Ivanov sees it.

Humans have recognized that errors exist in the data that is used for decision making and have adopted strategies to try to get around these errors. One strategy has been to aggregate data, by making summaries, statistical analyses, and other such methods that tend to "average out" the errors. Another strategy has been to provide "catch-all" categories in our classification systems, for the events that do not fit into the standard categories properly. Instead of such strategies, says Ivanov, we should recognize quality of data for what it really is and strive for it.

What does this imply? As we see Ivanov's point, a piece of data in the file would have three elements: the value of the attribute, an estimate of the accuracy error, and an estimate of the precision error. He gives some simple examples to illustrate his thinking.

Consider "date of birth" data, says Ivanov. It is normally obtained from the data subject, taken at face value, and entered into our information systems. But in fact, errors in "date of birth" do occur. To get an estimate of the accuracy error, one might determine a statistical measure of the discrepancies where dates of birth have been investigated. Or one might get the date-of-birth data on an individual from several (hopefully independent) sources. The estimate of precision error might be a statistical measure of data entry and transmission errors. Together, they would give a measure of confidence in the recorded "date of birth" data.

The estimate of accuracy and precision may or may not be important, says Ivanov. It can be very important if a person's life were to depend on it.

We think Ivanov's views are very relevant to privacy legislation. For instance, the Privacy Act of 1974 says that an agency must maintain any record concerning any individual with such *accuracy,* relevance, etc., as to assure fairness in determinations of . . . rights . . . benefits, etc. (emphasis ours). Further, it says that each agency shall collect information to the greatest extent practicable from the subject individual when the information may result in adverse determinations about the individual's rights, benefits, and privileges under federal programs. H.R. 1984 has very similar wording—collect information to the greatest extent possible from the data subject directly.

These privacy provisions say, in essence, "Give the benefit of the doubt to the data subject; take the data subject's word for it." This may not have been the intent of the wording—but it is, we think, a valid interpretation. Obtaining the information from just one source is almost sure to inject bias into the data.

In light of the above discussion, the accuracy

provision might be restated in the following way: "Collect information from a variety of independent sources, including the data subject, when the information may result in adverse determinations about the individual's rights, benefits, and privileges. Moreover, these sources should be chosen on the basis of most likely disagreement. Then indicate the estimates of accuracy and precision for each data item recorded about the individual."

From a practical standpoint, most organizations already do and will continue to (1) collect appropriate data from the data subject, (2) collect that data from the data subject only, and (3) accept that data at face value (except, of course, the police, welfare agencies, credit bureaus, etc.). The reason is evident—it is usually much easier and less expensive than the alternatives. Moreover, privacy legislation already has and will probably continue to put an "official seal of approval" on this policy.

But the privacy legislation also requires that personal data be collected and maintained with such accuracy (and relevance, etc.) as is necessary to assure fairness in determinations. As Ivanov sees accuracy (and we believe he has a strong point), the policy that has been set for collecting the personal information—that of collecting it from the individual himself, to the greatest extent possible—*is inherently inaccurate.* Thus, the legislation is probably contradictory on this point. We suspect that there will be court cases in the not-distant future that will test this contradiction.

Interesting, isn't it? The words of the legislation seem reasonable at first but upon further study all sorts of problems arise. This is one reason, we think, why privacy legislation for the private sector should not be hurried but should be debated carefully.

Possible security enhancements

"We have discussed computer center security and data security in a number of previous reports, including May 1970, December 1971, January 1972, December 1973, and January 1974. Hence we will not attempt any in-depth discussion at this point but instead will provide a brief overview.

Many computer-using organizations still have substantial gaps in their physical and data security systems. When we visit installations, we are frequently invited to walk through the computer rooms; there are no guards on the doors and anyone can walk in. One computer field security expert related a test he made on this point. He simply walked into a computer room (at the site he was asked to inspect) and started looking around. An employee saw him, did not recognize him, and following instructions, she asked him to identify himself. He refused to do so. She asked him if he belonged there. He said no. She told him he would have to leave. He refused. She pleaded. He said he would leave if he could have one of the magnetic tapes. She said he could have anything in the place, if he would only leave. His conclusion: inadequate security.

As we have discussed in previous issues, there are a number of steps that organizations can take to inhibit the accidental or deliberate destruction or damage of the physical installation, or the theft of storage media, printouts, etc. The techniques of physical security are well within the state of the art.

On the other hand, effective access control in multi-access systems is *not* within the state of the art. A basic premise of most of the workers in the field of data security is: if data is on-line to the computer, it can be accessed by a skilled penetrator working at a terminal attached to that computer. All serious penetration efforts have been successful, we have been told by these people. At the same time, some new operating systems—such as IBM's vs2/Release 2, TENEX, and MULTICS, do in fact offer higher levels of security than previous commercial operating systems. While more secure than previous operating systems, we gather that they are still not impenetrable for skilled people.

James P. Anderson, a consultant in computer technology, pointed out to us that the success of the military information security system stems from the nearly 40 years of experience with it. One feature of this system is the use of sensitivity levels (confidential, secret, top secret), with standards for handling each level, both manually and in computer systems. So far, privacy legislation has not identified levels of sensitivity for personal data.

The policies of the U.S. Air Force, related to the multi-level security problem, are discussed in Reference 8. There are two basic policies. If jobs with more than one security level are run simultaneously on a multi-access computer system, one

approach is to clear all users, terminals, communications lines, etc., for the highest security level. If some top secret work is being processed, this means that even users with unclassified work would have to be cleared and operate under the conditions of top secret work. The other approach is to process one security level at a time—run all of the unclassified work, then all of the secret work, etc. Moreover, before changing to a different security level, all memory devices would have to be cleared—internal memory, disk units attached, magnetic tapes attached, even printer ribbons. This so-called "color changing" can consume 20 to 40 minutes, it is reported.

As with physical security, we gather that most users of multi-access computer systems may be unaware of how little data security they really have. If data is on-line to the computer, you *must* assume that it can be accessed by any program—either from batch programs or from on-line terminals, it makes no difference. If you do not want the data disclosed, do not keep it on-line to the computer. Make it available to the computer only when it is the only work being processed and all other programs are disconnected. At the conclusion of the work, go through a "color change" routine to wipe out all traces of the work.

If that sounds inconvenient, that is the way it is. If you do not take this precaution, you are running a real risk of disclosure of the data, assuming that someone else wants to see that data.

Well, you say, what about encryption? Why don't we encrypt the data we want to protect, both during transmission and in storage? Then if someone accesses it, they won't be able to understand it.

Should you use data encryption?

Following our January 1974 issue, in which we discussed data encryption briefly, we decided to do a complete issue on the subject. It sounded like a practical method to enhance data security. So we talked to a number of people, wrote lots of letters, collected a fair amount of literature—and then decided to forget the idea of a complete issue on data encryption.

What was the problem? First, we encountered a good number of conflicting claims about how secure a given encryption method or device might be. Secondly, there was no ultimate source of good information available to us—for instance, the U.S. National Security Agency. Thirdly, there was no way for *us* to analyze and evaluate the conflicting claims.

So instead of a complete issue on this subject, we will give a summary of our findings. Data encryption (for both transmission and storage) can be an important security measure for the protection of data—and, in today's environment, that includes personal data.

We see the status of data encryption as follows. First, there are a number of commercial data encryption/decryption hardware devices on the market. In addition, there are a number of software encryption methods in use, some by commercial timesharing companies. Also, we have seen critical comments about the effectiveness of some of these commercial offerings, comments that we are in no position to evaluate. Finally, the U.S. National Bureau of Standards has proposed one method for adoption as a federal standard. Should this adoption occur, this method probably would become the "standard" method throughout the computer field in relatively short order.

How good are commercial encryption devices?

James P. Anderson, of Fort Washington, Pennsylvania, heads a company that does consulting in computer technology. A fair amount of his work is in the area of data security and encryption. He performed one study for the U.S. Air Force (a copy of the report of which was given us by the Air Force) on how commercial encryption devices might be used in a multi-security-level operating environment. In this environment, it is desired to run multiple security levels of work simultaneously, ranging from unclassified to top secret. The unprotected terminals and communications lines, used for unclassified work, provide an entry means for the penetrator. Once entry is gained, system software changes can be made to make future entries easy. This study explored the idea of using hardware encryption devices for *all* communications. If the penetrator could not properly encrypt his messages, the messages would be rejected by the computer and he could not gain entry.

Anderson studied the idea of using relatively inexpensive commercial encryption devices to provide this protection. In general, these devices have a set of switches which are set by the user; from the setting of these switches, called a user

"key," a pseudo-random binary stream is generated. This binary stream is added to the message text to produce cipher text. Advertising literature for such devices tells how many billions of different possible binary strings can be used, with the implication that an exhaustive search is needed to find the right binary string to decode a message.

Anderson assumed that a very fast processor would be available for making such an exhaustive search, one that could make up to 5,000 tests per second. Further, he assumed that the key would be changed once a day. Further, he assumed the need for a very low probability (one chance in a million) that in one day's time, the key could be broken by such a search.

He pointed out that the log-on routine is generally quite stereotyped and provides an excellent sample of matched plain text and cipher text for the penetrator. The penetrator (via his computer) would continue to try different keys until a legitimate log-on message resulted.

The details about the commercial devices are not released by their suppliers. Anderson had to rely on the information that was released plus discussions with supplier personnel.

He concluded that one popular device provided a probability of penetration of only one in one thousand, far short of the goal of one in a million. Another gave a probability of one in 25,000. Another looked like it would meet the one in a million goal, based on some assumptions about the method used.

In short, he could not be optimistic about the use of these three commercial devices in the environment which he assumed. However, it would seem that personal data would not need the same degree of protection as would top secret military data. A greater risk might be tolerated in the case of personal data.

However, Baker (Reference 9) points out that low quality encryption devices may, in fact, degrade security. They call attention to sensitive data but provide inadequate protection from a skilled cryptanalyst. Furthermore, he says, they give a false sense of security and may lead to the sending of sensitive information that otherwise would not be sent. If data is worth encrypting, it is worth encrypting well, says Baker.

What are the problems with these low quality devices? Baker lists several. For one thing, a penetrator may be able to obtain one of the encryption devices, inspect its wiring, and analyze the binary stream that such a unit produces. Then the penetrator may be able to tell when a code setting is "almost right" by the way the cipher stream behaves. Further, such devices may generate repeated sequences of the binary stream which might be detected by cryptanalysis methods.

In this same line of thinking, Bryant Tuckerman of IBM has demonstrated that the popular Vigenere-Vernam ciphers are far weaker than even a skilled amateur at cryptanalysis might think.

We conclude from these studies that it is literally impossible for the lay person to evaluate encryption devices. Simply claiming that the devices may use billions of different keys is not sufficient. We also wonder, who is the expert in this field? Just because Expert A claims that an encryption method cannot be broken in less than some stated amount of work does not guarantee that Expert B cannot break it relatively easily.

So, it seems to us, it ends up as a matter of trust. Who do you trust? And how does the lay person evaluate the experts and make a decision on who to trust? That seems to be every bit as difficult as evaluating encryption devices.

Which leads us to a discussion of the federal data encryption standard.

A federal data encryption standard algorithm

The method (algorithm) that has been proposed as a federal data encryption standard was developed by IBM and submitted to the National Bureau of Standards for consideration as a federal information processing standard. Several patents are involved in the implementation of the algorithm. IBM offered to release any of its patents to the extent they must be used in complying with the standard if the method is adopted as a federal standard by September 1, 1976. IBM has agreed to make available royalty-free, non-exclusive licenses in accordance with this offer.

The method is known; it is described in detail in publicly available literature. Security of the data protected by using the algorithm is based on the security of a key. The concept is identical to keyed and combination physical locks. You may know how they work but if you do not have the key or the combination, it is very difficult to open the locks.

The method uses a block structure rather than a

string structure. The popular string approach generates a long sequence of random numbers and adds it, number by number, to the plain text to obtain cipher text. A message of cipher text must therefore be deciphered by starting at the beginning of the message.

With the IBM block approach, the algorithm operates on message-blocks of 64 bits, producing cipher-blocks of the same length. The user-selected key also has 64 bits, 8 of which are used for parity checking. The encryption of a block consists of an initial (input) permuation of the 64 bits, then 16 iterations of a key-dependent calculation to be described, and then a final (output) permutation.

Each of the 16 iterations consists of the following steps. First, 32 bits—that is, one half of an intermediate message (which is either the permuted input or the result of the preceding iteration)—are expanded to 48 bits by duplicating half of them. These 48 bits are then exclusive-OR-ed with 48 bits selected from the key; the selection is different for each iteration. The resulting 48 bits are broken into 8 groups of 6 bits. Each of these groups of 6 bits is subjected to a different nonlinear function which yields a group of 4 bits, for a total of 32 bits. These 32 bits are then subjected to a permutation, which mixes the groups, and are then exclusive-OR-ed with the other half of the intermediate message. The two halves are then interchanged, except on the last iteration.

Decryption consists of the same steps, except that the sixteen 48-bit selections from the key are used in the reverse order.

The protection and security of the key is vital. If the key gets lost, for all practical purposes the data will be lost.

To try to find the 64-bit key by exhaustive search might be a tremendous undertaking. We were given one estimate of approximately 10^{16} complete tests to effect the compromise of a single key. The U.S. National Security Agency was consulted by NBS on the effectiveness of this algorithm and, we were told, agreed that it was very effective. As we say, the question is: who do you trust? A layman cannot evaluate the effectiveness of the method.

We understand that the algorithm, when implemented by software, requires a lot of CPU cycles. One implementation requires one-half second of processing time per 64-bit block on a PDP-11, we were told. By refinements, the processing time might be cut to 50 to 100 milliseconds per block. When the algorithm is implemented via LSI chips (hardware), the processing time per block is 20 to 50 microseconds. So it is likely that most implementations will be in hardware, for which IBM has agreed to grant royalty-free, nonexclusive licenses.

Our opinion is that the method will become a federal standard and will be adopted as an *ad hoc* standard in the private sector.

To use encryption or not?

The decision to use data encryption is not a trivial one. There can be a lot of complications, both technical and social. We will briefly touch on some uncovered in our study.

Key handling. The same key must be used for both encryption and decryption. With an effective algorithm, if the decryption key differs from the encryption key by only one bit, the message text cannot be recoverd. For data transmission uses, the distribution, storage, and protection of keys will pose all sorts of problems; they can be lost, stolen, copied, may not arrive on time, and so on. Keys should be changed periodically, but this raises questions of what to do about encrypted data in storage. Must all encrypted data be retrieved, decrypted with the old key, encrypted with the new key, and then stored? Further, if keys are stored in the computer in software form, skilled penetrators can gain access to them and security will be compromised.

Cost/effectivenss. Friedman and Hoffman (Reference 12) have analyzed processing time requirements for a number of encryption methods implemented by software. The costs of encryption/decryption are real and must be applied to the processing costs for all sensitve data.

James Anderson, in a comment to us about software implementations methods, says that such methods are of value only in protecting data on removable media from being read if the media itself is stolen. Since skilled penetrators can get through all of today's operating systems, they can capture the cipher keys stored in memory.

We believe that, from a cost/effectiveness standpoint, only the hardware encryption/decryption methods will stand up. Further, we are assuming the use of something like the IBM method, implemented in hardware.

Data transmission. As indicated above, it may be necessary to encrypt *all* transmissions to and from remote terminals, to prevent penetrators from gaining any entry into the system. Data transmission must then be in the transparent protocol, so that encrypted characters will not be interpreted as data transmission control characters. On the other hand, if not all transmissions are encrypted, the processors will have to route some of the data stream to decryption devices while the rest need not be so diverted.

Social problems. Cogar (Reference 13) argues that the use of encryption involves a risk that society may find unacceptable. Encryption makes it harder for the thief to rob you, he says, but it also makes it harder for you to know that you have been robbed. He also questions whether it is wise, from a societal standpoint, to allow members of society to decide which data (and telephone conversations, etc.) will be encrypted. What if all of the underworld communications were encrypted, he asks? Can society afford not being able to know what is going on inside the information systems?

Integrity and security of personal data

If we can generalize about past practices, the following seems to have been the situation as far as integrity and security of personal data are concerned. In both the private and public sectors, only limited attention has been given to data integrity. Generally, each data item is obtained from one source and accepted at face value. Most personal data has been collected from the data subjects themselves. Other (different) data may be collected from third parties, such as credit bureaus, investigatory agencies, and so on. There has been no convenient way to carry different values for one data item, collected from different sources, or to handle those different values even if they were carried. Most organizations do want and need accurate, complete, and timely data for the efficient conduct of their business. They do tend to collect more data than they really need, and they do tend to tolerate a certain level of error in the source data. Most efforts on data integrity have dealt with detecting and correcting errors in data movement and transcription.

As far as data security is concerned, national security data has been subject to quite rigorous controls, including security clearances and need-to-know authorizations. (There still have been major violations, such as the famous Pentagon Papers case. Also, government agencies have used security as a shield for keeping possibly embarrassing information from the public eye, now made more difficult under the Freedom of Information Act.) In the private sector, really sensitive data has usually been kept in manual files, often under lock and key. "Normal" data has not been particularly safeguarded, probably because of the feeling that there would be no great loss if it were damaged or disclosed.

But with the arrival of privacy legislation, this picture is changing for personal data. U.S. federal agencies are already subject to a privacy law, and other legislation has been passed or is under consideration for state and local government units as well as the private sector. These laws will demand much more concern for the integrity and security of personal data.

As far as integrity is concerned, we believe that the privacy aspects of personal data will have to be carefully defined, as we discussed in our November 1975 report. The accuracy, completeness, relevance, and timeliness of the data must be carefully defined. Procedures must then be set up and enforced to see that those data definitions are met.

We suspect, too, that the question of the basic accuracy and precision of personal data will have to be considered, perhaps along the lines proposed by Ivanov. This could prove to be a very difficult problem area.

As far as security is concerned, organizations will have to make sure that their physical security systems adequately protect personal data. (Question: how much is adequate?) Data security will continue to be a major problem area for some years to come. If you have *any* personal data in on-line files, you must assume that it can be accessed by any program in the computer. It seems to us that an organization subject to privacy laws that adopts a "no need to change anything" policy for data security is taking a dangerous course. None of today's operating systems provide any real security. If it can be proved that unauthorized disclosure of personal data did occur, and that it *could* have occurred via the computer, then civil and perhaps criminal penalties might be imposed if the "make no change" policy had been followed.

The encryption of sensitive personal data is one

security mechanism that could be used. But, as we have pointed out, encryption brings with it some real problems. About all we can say is, if you choose this course, keep as much sensitive personal information out of your mechanized systems as possible and use just as good an encryption scheme as you can get.

REFERENCES

1. *Privacy Act of 1974*, Public Law 93-579, U.S. Congress; order from your U.S. senator's or congressman's local office.
2. *House Bill H.R. 1984*, U.S. House of Representatives, 94th Congress; order from your U.S. congressman's local office or write Document Room, House; H-226; U.S. Capitol, Washington, D.C. 20515.
3. *Systems Review Manual on Security*, AFIPS Press (210 Summit Avenue, Montvale, New Jersey 07645), 1974, price $10.
4. *Annotated Bibliography on Data Integrity and Security*, EDP ANALYZER (925 Anza Avenue, Vista, Calif. 92083), April 1976, price $10.
5. Webb, D.A., and R. P. Abbott, "Obtaining computer data integrity through accreditation of computer centers," Lawrence Livermore Laboratory Report UCRL-77093, paper prepared for submission to *Computerworld*, August 1975; see CW, October 8, 1975, page 16.
6. "The Privacy Mandate," summary report of a symposium/workshop held in April 1975, co-sponsored by U.S. National Bureau of Standards and the Mitre Corporation. Order from: The Mitre Corporation, Westgate Research Park, 1820 Dolly Madison Blvd., McLean, Virginia 22101, Attn: Mr. Gene Raichelson; price $4.
7. Ivanov, Kristo, *Quality-Control of Information* (doctoral thesis), 1972; order from U.S. National Technical Information Service (Springfield, Virginia 22151), PB-219297; price $2.25 microfiche, $9.75 paper.
8. "Computer security developments summary," Electronic Systems Division, U.S. Air Force Systems Command (Hanscom Air Force Base, Massachusetts 01730), December 1974.
9. Baker, H. C., "Voice and data scramblers," *Business Communications Review* (800 Enterprise Drive, Suite 115, Oakbrook, Illinois 60521) September-October 1974, p. 31-37; price $45 per year.
10. Girsdansky, M. B., "Cryptology, the computer, and data privacy," *Computers and Automation* (815 Washington Street, Newtonville, Massachusetts 02160), April 1972, p. 12-19; price $11.50 per year. Based on IBM Research Report, Vol. 7, No. 4, 1971, "Data privacy."
11. Feistel, H., W.A. Notz, and J. L. Smith, "Cryptographic techniques for machine to machine data communications," IBM Research Report No. RC 3663, December 27, 1971.
12. Friedman, T. D. and L. J. Hoffman, "Execution time requirements for encipherment programs," *Communications of the ACM* (1133 of the Americas, New York, N.Y. 10036), August 1974, p. 445-449; price $5.
13. News comment on "security and IBM's encryption algorithm," *Data Processing* (Dorset House, Stamford Street, London SE1 9LU, U.K.), September-October 1975, p. 255-256; price £1.75.

One of the interesting phenomena of recent years in the computer field has been the growth in popularity of the APL *programming language. Interesting because one of its big growth areas has been in business, in support of management problem solving and decision making. But just how are decision makers using* APL *and other decision support systems to help them make decisions? We discuss this question next month and report on some recent research studies of these types of systems and their use in business environments.*

SUBJECTS COVERED BY EDP ANALYZER IN PRIOR YEARS

1973 (Volume 11)

Number

1. The Emerging Computer Networks
2. Distributed Intelligence in Data Communications
3. Developments in Data Transmission
4. Computer Progress in Japan
5. A Structure for EDP Projects
6. The Cautious Path to a Data Base
7. Long Term Data Retention
8. In Your Future: Distributed Systems?
9. Computer Fraud and Embezzlement
10. The Psychology of Mixed Installations
11. The Effects of Charge-Back Policies
12. Protecting Valuable Data—Part 1

1974 (Volume 12)

Number

1. Protecting Valuable Data—Part 2
2. The Current Status of Data Management
3. Problem Areas in Data Management
4. Issues in Programming Management
5. The Search for Software Reliability
6. The Advent of Structured Programming
7. Charging for Computer Services
8. Structures for Future Systems
9. The Upgrading of Computer Operators
10. What's Happening with CODASYL-type DBMS?
11. The Data Dictionary/Directory Function
12. Improve the System Building Process

1975 (Volume 13)

Number

1. Progress Toward International Data Networks
2. Soon: Public Packet Switched Networks
3. The Internal Auditor and the Computer
4. Improvements in Man/Machine Interfacing
5. "Are We Doing the Right Things?"
6. "Are We Doing Things Right?"
7. "Do We Have the Right Resources?"
8. The Benefits of Standard Practices
9. Progress Toward Easier Programming
10. The New Interactive Search Systems
11. The Debate on Information Privacy: Part 1
12. The Debate on Information Privacy: Part 2

1976 (Volume 14)

Number

1. Planning for Multi-national Data Processing
2. Staff Training on the Multi-national Scene
3. Professionalism: Coming or Not?
4. Integrity and Security of Personal Data

(List of subjects prior to 1973 sent upon request)

PRICE SCHEDULE

The annual subscription price for EDP ANALYZER is $48. The two year price is $88 and the three year price is $120; postpaid surface delivery to the U.S., Canada, and Mexico. (Optional air mail delivery to Canada and Mexico available at extra cost.)

Subscriptions to other countries are: One year $60, two years, $112, and three years $156. These prices include AIR MAIL postage. All prices in U.S. dollars.

Attractive binders for holding 12 issues of EDP ANALYZER are available at $4.75. Californians please add 29¢ sales tax.

Because of the continuing demand for back issues, all previous reports are available. Price: $6 each (for U.S., Canada, and Mexico), and $7 elsewhere; includes air mail postage.

Reduced rates are in effect for multiple subscriptions and for multiple copies of back issues. Please write for rates.

Subscription agency orders limited to single copy, one-, two-, and three-year subscriptions only.

Send your order and check to:
EDP ANALYZER
Subscription Office
925 Anza Avenue
Vista, California 92083
Phone: (714) 724-3233

Send editorial correspondence to:
EDP ANALYZER
Editorial Office
925 Anza Avenue
Vista, California 92083
Phone: (714) 724-5900

Name_____

Company_____

Address_____

City, State, ZIP Code_____