Journal of Official Statistics
Vol. 9. No. 2, 1993, pp. 501–502
© Statistics Sweden

Discussion
Computer Security

Martin H. David[1]

Keller-McNulty and Unger bridge a gap between the conceptual framework used by statisticians and the conceptual framework used by computer scientists in thinking about disclosure or compromise of data held in machine readable form. They stress that statistical thinking has been about compromising inferences; computer science thinking has been about compromising access. Access can be controlled by encryption, authorization, and intelligent parsing of queries to the data. Computer systems analysis also recognizes that vulnerability of data arises in three modes: unauthorized disclosure, unauthorized modification, and denial of authorized access.

A systems point of view makes it clear that security of data is insured by the intersection of decisions made at six levels: Policy and management, physical, transmission/communication, access, information flow, and inference. Attention by statisticians has focussed largely on the last two; release of sensitive data to a broader community of users must focus on all six. It is especially important to consider the responsibilities of super-users in computer systems and policies affecting the security of networks undergoing decentralized management and maintenance.

[1] University of Wisconsin, Department of Economics, 6440 Social Science Building, 1180 Observatory Drive, Madison, WI 53706, U.S.A.

Using the relational database model directs attention to controlling access not only by user-identity, but also by the intersection of user-identity and particular data objects, including attributes of particular entities and relationships. Security on the data object can be applied to sensitive attributes, to particular strata of entities, and to the relationship between different classes of entities. Implementation of the relational model offers data producers an environment in which control of access is substantially enhanced, relative to a process in which ASCII data on magnetic tape are submitted to statistical processors. (Implementation of the relational model for statistical databases involves tangible costs of database design but affords gains in addition to controlled access. The timeliness, accuracy, and feasibility of answering queries can be increased.)

Attention has been given to controlling views of data available to particular users and to limiting the nature of queries that may be executed by the user (read access only, prohibition on access to devices that can store copies of the data, etc.). While information flow from "one-at-a-time" queries can be controlled, no strategy has yet been devised to provide universal protection against sequences of queries that are aimed at compromising the data.

It is useful to regard the present federal

statistical system as a federated database. Conceptually, each agency has control of data that are now isolated. Emerging hardware, clustered computers and high-capacity communications will bring all of these isolated systems closer. The question of data management then becomes: How can the users of the data best be served by considering the federated database as a distributed database? What are the appropriate access and information flow controls? How can the database management system protect against attack? These questions are at the forefront of work by database experts who are currently adapting database management systems to the real problems of scale and independence that occur with distrubuted computing.

In the future, other technologies will add to the problems of system control, but they also create opportunities. Encrypted databases are being disseminated on CD-ROM. It will be difficult to maintain the integrity of data treated in that manner, as errors are detected and new versions released.

Some advances have been made in auditing computer use by expert systems to detect unusual requests that may reflect attempts to breach security on the part of routine users. In intelligence systems some databases have been created in which the datum that appears in answer to a query varies. This polyinstantiation assures that secret data are not compromised; the concept may be useful in dealing with readily identifiable entities in databases on business entities.