Establishing Security Controls in a Distributed Data Base

Sooun Lee
Mohamed Nour
Soong H. Chung

Payoff
Distributed computing systems can provide benefits not available with centralized computing systems. However, it can be difficult to maintain the security and integrity of data on distributed systems. In fact, some characteristics of distributed data bases give rise to additional security concerns. This article provides a detailed analysis of information security issues in the distributed data base environment and suggests a model for establishing and implementing effective security controls.

Introduction
In recent years, the computing environment has shifted from the centralized model to a more decentralized one. This shift has been motivated primarily by the increased decentralization of most business operations as well as the perceived advantages of decentralized computing environments.

Distributed data base systems (DDBSs) are an important element in distributed processing systems. The concept of Distributed Data Base System has been evolving during the past few years and a great deal of uncertainty still exists about what the concept means and what these systems do. A particular challenge associated with distributed data bases concerns information security. Although distributed data base systems share some of the same security problems as centralized data bases, their unique characteristics add other security concerns not found in the centralized environment.

This article describes the features and benefits of the distributed environment and how the proliferation of this recently introduced technology has led to a greater need to establish an information security plan. The article presents a model for securing data, taking into consideration the special needs of this environment.

The Benefits and Features of DDBs
There are perhaps as many definitions of Distributed Data Base System as there are authors who write about them; this is partly due to the relatively recent development of the technology and the confusion surrounding its definition. A distributed data base environment has been described as a communications network comprising individual sites with their own standalone data bases and with local operation. The data base may be distributed because of geographical, performance or effectiveness, and efficiency reasons.

The objectives and benefits of distributed data base systems include:

- Increased system availability and better performance.
- Higher system throughput.
- Improved disaster recovery.
- System reliability.
- Reduced communications costs.
Information-sharing capabilities.

Distribution of certain applications.

Flexibility and extensibility.

Local control.

Incremental data base growth.

These benefits do not come without a price, however, and the distributed nature of the data base is the source of most of the problems of distributed data base systems. These problems include higher operating costs, communications delays, consistency (i.e., update) problems, concurrency control, and data security problems. Compared to a centralized data base, a distributed data base is much more vulnerable to possible security threats—for example, theft, invasion of privacy, updating, and unauthorized access—because of the dispersed structure of its data.

Data can be distributed over multiple computer systems in three ways: through replication, partitioning, or allocation. Replication implies that the same copy of the data base is stored in each of the sites or nodes in the computer network. Partitioning, however, means that data is stored in the site where it is most needed. Partitioning is sometimes known as data fragmentation, whereby logically related sets of data are stored in the same site. Allocation is a hybrid of partitioning and replication. Each method has its own implications for data recovery, updating and concurrency control, system reliability and availability, data security and integrity, and transmission costs.

Security Threats in a Distributed Data Base

The basic threats to the security of data in a distributed data base environment are not very different from those affecting centralized computer systems. They include crime, employee negligence, accidents, technical faults, viruses and worms, environmental factors, and natural disasters. However, certain characteristics of distributed data bases (e.g., increases in the points of entry to the system, in user autonomy, and in the number of users) create additional security concerns.

The consequences of security violations can be classified into the following:

- Loss or destruction of data.
- Theft of data.
- Corruption or modification of data.
- Disclosure of data.
- Delay or denial of use of data.
- Misinterpretation of data.

Each of these data security violations could have a dramatic effect on an organization's operations or even its survival. Therefore, the need for effective management and control of data security in a Distributed Data Base System is compelling. The conceptual security policy model, described in the following section of the article, is suggested to provide practitioners with guidelines for managing security in a distributed data base environment.
The Conceptual Security Policy Model

The information security model described in this article involves a three-step process:

- Classifying the data.
- Determining the organization's security needs.
- Selecting the appropriate security techniques.

Each step is described in detail in the following sections.

Step 1: Classifying the Data

The first step in the development of an effective security model in a distributed data base environment is to classify the distributed data into three dimensions: location, sensitivity, and vulnerability. These three dimensions serve as the basis for the security plan.

First, data must be classified from its current processing status as either data in place or data en route. This dichotomy reflects the two states of data in a distributed data base environment at any particular time. The category data in place refers to data currently residing in a distributed data base environment. Data en route refers to data in transmission. If the data is replicated, more attention should be given to data in place than to data en route, because most processing in this case is local. If, however, the data base is partitioned or allocated, more data is en route (i.e., transmitted), and consequently, more concern should be shown for protecting this transmitted data. Tolerance techniques can be used to protect data classified in this category. (These techniques are discussed in more detail in a subsequent section of the article.)

Second, the possible vulnerabilities or weaknesses in the distributed data base environment must be recognized and assessed. This assessment of potential sources of data security threats can contribute to finding the needs for the avoidance techniques discussed in a subsequent section. This assessment can also help managers establish the following managerial guidelines:

- Determining the types of security control required.
- Assessing the efficacy of existing safeguards.
- Identifying necessary additions and improvements in security measures.
- Determining the availability of resources.
- Determining the level of security coverage needed.

Managers should understand that at this point, probably the most serious threats are from the employees within the organization.

Finally, data must be classified according to its sensitivity and importance. For example, the data can be classified as routine, private, confidential, or strictly confidential. Depending on the sensitivity level, different types of security mechanisms may be needed. This classification can also be used as a significant determinant in measuring the resource requirements for the security model. Data must be examined and evaluated according to the three classifications—location, sensitivity, and vulnerability—to determine the organization's basic security needs.
Step 2: Determining Security Needs

After the data has been classified, the next step in the model is to determine and formalize the organization's general security needs. At this stage, general decisions regarding such issues as the security level of preventive and detective controls. Total security is hardly attainable or economical. It would be meaningless to spend, for example, $1 million to protect information that is worth only half that amount. An organization needs protection that is reasonable in view of the importance and value of the object of protection (i.e., data). As a general rule, data security should be acquired as long as benefits exceed costs.

The nature and types of security needs of an organization depend on such factors as the characteristics of the organization and its employees or users, and the sensitivity of its data to security violations. An organization that is not significantly affected by the disclosure of its information resource to competitors and adversaries would not be inclined to spend a significant amount of money on security. As this information becomes critical to the business and survival of the organization, the organization becomes more willing to spend for the sake of protecting its data.

Another important decision is to determine the security methods an organization needs (e.g., tolerance, avoidance, or a combination of both). This decision depends on the types and nature of the distributed data bases and available resources. These two methods involve different sets of protection tools with different costs, which is discussed in more detail in the following section of the article. The suggested model serves only as a guidepost and does not propose to define exactly how to select from among the two methods.

Step 3: Selecting the Appropriate Technique

Exhibit 1 illustrates the relationships between the various mechanisms of the protection methods and techniques. The basic concepts in the model are avoidance and tolerance.

Selection Model

Avoidance techniques pertain to internal as well as external threats to distributed data base security. Internal threats in turn are generally from employees and legitimate users. These threats may be intentional (i.e., caused by dishonesty) or accidental (i.e., caused by negligence and errors).

Tolerance, however, pertains to threats to the two types of data in a distributed data base environment: data in place and data en route, as discussed previously. The concept of risk or threat tolerance focuses on the idea that rather than denying a potential intruder access to data, the data should be made useless to the intruder.

This step involves specific security techniques using parameters described in the second step. The following section explores these techniques in detail.

Avoidance Techniques

Avoidance techniques pertain to internal threats, both intentional and accidental, and external threats. These types of threats and techniques for avoiding them are explained in the following paragraphs.

Internal, Intentional Threats

Internal, intentional threats come from employees who, for professional or personal reasons, try to defraud the organization by stealing, disclosing, or destroying data in
Distributed Data Base System. Preventive measures against these threats focus on ensuring that the potential user in a distributed data base environment is a legitimate user, who is trying to access an authorized data object or file and performing an authorized action. The first layer of this type of security ensures the legitimacy of a user by authenticating the individual's identity. The user is identified through a combination of some form of user ID and password. Other methods include the use of smart cards or tokens and biometrics systems.

The second layer of security focuses on ensuring that the user, having been identified as legitimate, is trying to access data that the individual is authorized to access. This is accomplished through the establishment of authorization rules involving such methods as user privileges and clearances, data classification and fragmentation, and restricted views.

Users are assigned clearances according to their privileges in using specific segments of the data base. Given the nature of the distributed data base environment, the assignment and tracking of clearances can be a difficult task because the population of potential users includes all those in the multiple nodes of the distributed data base systems. This procedure, however, helps determine who is authorized to access which object or segment of the data base. Data classification assigns sensitivity levels to different segments of the data base. Together, with the user classification, data sensitivity levels help determine the domain of each user.

The third layer of security is accomplished by the implementation of embedded controls in the distributed data base systems (the software that runs the distributed data base) and audit trails. Although access and authorization controls through audit trails are retrospective in nature, high-quality audit trails can provide a complete record of every activity that an individual performs.

**Internal, Accidental Threats**

Accidental threats from distributed data base systems users may be considered less serious than intentional employee threats because, unlike intentional threats, there is no malicious intent behind them. Nevertheless, the effects could still theoretically ruin an organization, especially if very sensitive, private data is involved. Measures to handle these threats include the implementation of extensive training programs, the heightening of user security awareness, and constant monitoring and assessment of security violations resulting from user negligence, ignorance, or omission of automatic error-checking methods.

**External Threats**

Protection against outsiders includes port protection measures, network firewalls, and audit trails. The purpose of installing port protection measures is to limit or prevent access to the system from outside. This could be implemented through a dial-back mechanism that checks the validity of the caller. Network firewalls can be established to automatically screen out external intruders.

The other method for protecting against would-be security violators is audit trails. The mere fact that the system traces intruders' paths limits the possibility of getting away without being detected and may prevent the intruders from even attempting to enter the system.

**Tolerance Techniques**

The storage and transmission of data in a distributed data base produces significant security risks. The tolerance approach allows potential violators to access data, but the data is made unmanipulable or unintelligible, thus preventing the likelihood of data abuse. Data in a distributed data base lies in one of its two states, in place or en route.
Techniques for Protecting Data In Place

For data in place, some methods have been developed to protect data by modifying the data in some fashion. Two of these techniques are noise addition and data dispersal. Noise addition involves adding noise to the data base objects, which are designated as $O_s$. If noise were represented by the variable $e_s$, the equation would read as follows:

$$O_s^* = O_s + (e_s)$$

For example, an organization might wish to protect the confidentiality of such data as employee salaries, ages, or sales per product or territory. To make these data elements not directly useful to an intruder, management could disguise these figures by adding fictitious data elements (possibly of the same data type) to these figures. Thus, age $a$ can be replaced with a disguised age ($a_1^*$). This is accomplished by subtracting a random number from the age as in the following equation:

$$a_1^* = a_1 + (--100) = a_1 -- 100$$

The negative nature of the resulting age figure would make it virtually useless for almost any purpose. Likewise, for example, sales by product can be disguised from possible competitor intelligence by adding a false component to it. Thus, if the sales of product $i$ is denoted by $S_i$, then the disguised sales figure ($S_i^*$) would be:

$$S_i^* = S_i + e_i$$

where $e_i$ could be a function (e.g., 75%) of $S_i$ or an independent and probably constant component, as in the age example previously stated. For example, if $S_i$ is $982,000$, the figure can be disguised as follows:

$$S_i^* = 982,000 -- 75\% (982,000) = \$245,500$$

To reconstruct the true figure, in the knowledge that 245,500 is 25% of $S_i$, $S_i^*$ can be divided by 25 and multiplied by 100 as follows:

$$S_i = S_i^* \times 100/25 = 245,500 \times 100/25 = 982,000$$

The important point to remember is that it must be preestablished whether a constant value or a function of the value is used. Therefore, the previous example would read as follows:

$$S_i^* = S_i = f(S_i) = S_i -- aS_i = S_i(1-a)$$

where $a$ happens to be .75. To reconstruct $S_i$, $S_i^*$ is divided by $1-a$.

Like noise addition, data dispersal, or scattering, is used to protect individual data objects in a distributed data base environment. An Information Dispersal Algorithm has been proposed that disperses a data object (e.g., an employee record) into various storage areas (e.g., the address, salary, and employee identification number would be stored in different data bases). The data object, represented by $o$, would be divided into $n$ pieces. The series of subvalues would be represented as $o_{i}$, where $i$ is greater than or equal to 1 but is less than $n$ (i.e., the number of pieces). The number of pieces must suffice for reconstructing the data object. Each one of these pieces, or a reasonable combination of them, is stored in a different node in the distributed system. This technique makes it virtually impossible for a would-be intruder to access all the pieces to reconstruct the original data object for several reasons. First, the intruder might not be aware of the
existence of all the nodes. Second, even if the nodes are known, accessing all of them might be too difficult or time-consuming. Finally, even if all nodes are accessed and the data objects are collected, a considerable amount of effort might be needed to combine them into a whole that makes reasonable sense.

The value of this technique is even greater when placed in a real-world context. For example, assuming that a record includes \( n \) data elements (i.e., fields), these data elements could be distributed over the \( m \) nodes of the distributed database. If \( n > m \), the remaining \( n-m \) elements could be distributed again over the nodes so that each node has one or more elements but less than \( n \). Therefore, by accessing only one node, an intruder will not be able to access the full record, assuming that only a complete record can be useful. All the records of a sensitive file can be distributed in this fashion. A way is needed to reconstruct the records, and therefore, the file, when desired. This is accomplished by having a control file within the operating system or an application program, where the locations, and probably the addresses of each distributed data item are stored. It should be relatively easy to implement these two approaches by means of systems software or applications programs.

**Techniques for Protecting Data En Route**

For transmitted data, or data en route, data encryption can be used to protect data transmitted in a communications network. Other articles in this Auerbach *Data Security Management* series provide detailed information on available encryption techniques for transmitted data.

**Limitations of the Model**

The model suggested in this article has at least three limitations. First, the model does not identify all possible data protection techniques; the lists are therefore not meant to be all inclusive. Second, the model does not compare the various techniques in terms of effectiveness and efficiency. Finally, some of the techniques or devices may not have been fully developed and made operational.

Despite these apparent deficiencies, the model provides a framework for analyzing security problems in a distributed database environment and perhaps implementing a policy model that reflects the security needs and characteristics of a business entity in a distributed database environment.

**Conclusion**

The problems of a distributed database can be mitigated if proper administrative and technical procedures are devised and rigorously implemented. The security problem, in particular, needs special consideration. The model used in this article can guide practitioners to design proper security strategies in a distributed database environment.

**Author Biographies**

Sooun Lee
Sooun Lee is an associate professor in the School of Business department of decision science and MIS) at Miami University School of Business, Oxford OH. His research interests are database design and development, end-user computing, expert systems, microcomputer applications, and systems development.

Mohamed Nour
Mohamed Nour is a PhD candidate in the School of Business at Kent State University, Kent OH.

Soong H. Chung
Soong H. Chung was formerly an assistant professor in the School of Business at the University of Wisconsin in Eau Claire and is now associated with Korea Electronics Technology.
Step 1

Avoiding Threats

Step 2

Internal Techniques

External Techniques

Tolerating Threats

Data Place Techniques

Data Route Techniques

Step 3

For Intentional Threats:
- User Authorization
- Data Classification
- Audit Trails

For Accidental Threats:
- User Training
- Security Awareness
- Violation Monitoring

For External Threats:
- Port Protection
- Audit Trails
- Network Firewalls

For Internal Threats:
- Noise Addition
- Data Dispersion

For External Threats:
- Data Encryption