1

Security issues in data warehousing and data mining: panel discussion

Bhavani Thuraisingham[1], Linda Schlipper[1], Pierangela Samarati[2], T. Y. Lin[3,4], Sushil Jajodia[5], Chris Clifton[1]
[1]The MITRE Corporation/MS K329, Burlington Road, Bedford, MA 01730,USA
thura@security.mitre.org
[2] Computer Science Laboratory, SRI International, 333Ravenswood Ave, Menlo Park, California 94025, USA
samarati@csl.sri.com
[3]Mathematics and Computer Science Department, San Jose State University, San Jose, California 95120, USA
tylin@cs.sjsu.edu
[4]Department of Electrical Engineering and Computer Science, University of California, Berkeley, California 94720
[5]Department of Information Systems and Systems Engineering, George Mason University, Fairfax, Virgnia 22030-4444
jajodia@gmu.edu

Abstract

This paper describes the panel discussion on data warehousing, data mining and security.

Database Security XI T.Y. Lin and Shelly Qian (Eds)

INTRODUCTION BY BHAVANI THURAISINGHAM

Having a data warehouse for managing the data is becoming a necessity with many enterprises. Several organizations are building their own data warehouses. Commercial database system vendors are marketing data warehousing products. In addition, some companies are specializing in developing data warehouses. The idea behind a data warehouse is that it is often cumbersome to access data from multiple and possibly heterogeneous databases. Several processing modules need to cooperate with each other for processing a query in a heterogeneous environment. Therefore, a data warehouse will bring together the essential data from these diverse data sources. This way the users need to query only the warehouse. In addition, a data warehouse also contains information such as summary reports and aggregates that are determined by the applications using the warehouse and the queries posed.

A related technology, which is used to convert the data in the warehouse into some useful information is data mining. That is, data mining is the process of posing a series of appropriate queries to extract information, often previously unknown, from large quantities of data in the database or the data warehouse. Data mining technology is a combination of various other technologies including machine learning, database management, statistics, and parallel processing.

This panel discussion will focus on security aspects of data warehousing and mining. Data warehousing security issues include security architectures, integrating multiple security policies for the warehouse, inference problem, administrating and auditing the warehouse. Data mining security issues include preventing unauthorized disclosure of information from mining as well as privacy issues for data mining. On the other hand data mining techniques could also be used to help with security including auditing and intrusions detection. The discussion will cover these various aspects of security for both warehousing and mining.

POSITION BY LINDA SCHLIPPER

For most enterprises there is no shortage of data. Operational data systems have been collecting data for decades; however, the process of getting and massaging this data for decision-making can be problematic. Applications and data usually reside on different systems, are managed by different software, and are owned by different organizational components. Further, these systems are usually organized around applications, such as scheduling or bill-of-materials, as opposed to being organized around business entities more suitable for analysis. Extracting data from these systems is often complicated and can require some amount of programmer support since users need to know what information to request and how to request it in order to generate even a simple report. Finally, decision support systems, which are targeted at analyzing data, interfere with the primary duty of operational systems' collecting data.

The strategic value of timely decision support information is becoming apparent to many organizations. Data warehousing is a technique to integrate an organization's distributed, autonomous, and heterogeneous data for use in analysis. It is a process which extracts information from sources then integrates, transforms, and summarizes the data for storage in a warehouse database. The warehouse database is seen as a unified data store and has the following characteristics [Inmo97, Devl97]:

- **Subject-oriented:** schema oriented around subjects such as customer, vendors, items, etc., instead of functions such as purchase orders.
- **Integrated:** common data format, variables, and naming conventions so that access to the data can be uniform.
- **Time-variant:** contains a series of "snapshots" of operational data and retains the "snapshots" so that trends over time can be examined.
- **Low volatility:** end users generally have read-only access to the data warehouse.

Generally, the driving force behind the implementation of a data warehouse is the goal of providing a more complete picture of an organization's operations to support management decisions. However, by its very nature, a data warehouse creates a security conflict [Kimb97]. On the one hand, the goal of every data warehouse is to make available to all concerned the information they need. On the other hand, an organization needs to ensure that this same valuable data is not exposed to unauthorized individuals or corrupted by hostile parties. If the correct balance between security concerns is not met, not all users that could benefit from the information will have access to it. Although the security concerns for a data warehouse are the same as those for any other information system integrity, access control, authorization, privacy, and confidentiality data warehouses present some unique and challenging issues.

Integrity. As a repository of information from operational systems, a data warehouse depends upon the integrity of its source systems for its own integrity. However, working with the data warehouse environment to achieve high quality is complicated by the processing needs, distance from sources, and different requirements of data warehouse users.

Poor quality data may exhibit many flaws: inaccurate, incomplete, or lacking in agreement with the data from other sources. Special processing may be required at the data warehouse to resolve inconsistencies in data from various sources. Current data warehouse tools aid in the derivation and propagation of the instance data and table definitions, but they do not address the root causes of poor data or attempt to apply any rigor to correction at the source. Such a capability would require myriad types of integrity metadata and events on at least two tiers: the source database tier and the derived database tier. Propagating this additional information will require a

huge amount of implementation, customization, and administration. To obtain automated help for this process, a framework is needed for organizing the pieces of the solution. A proper framework would make it possible to integrate integrity logic obtained from a variety of sources and would provide services to simplify the building of a system of integrity information. Ultimately, an organization must know what risks are associated with incorrect information and decide the level of acceptable risk.

Access Control. Controlling access to a data warehouse is particularly important since the data warehouse encompasses data from many systems and contributes to decision-making across organizational boundaries. In fact, access controls to a data warehouse need to be considered at a number of levels [Ross96, Rose97]. First, who will have access to the processes which extract the operational data. Secondly, who has access to the data and the processes that transform the operational data into a format suitable for inclusion in the data warehouse. Finally, who will access to the data in the data warehouse itself. The ease of access to large amounts of data raises concerns about attaching the appropriate level of security without inhibiting analysis.

Identifying and implementing an access control policy for a data warehouse involves a number of unique challenges. One is the dissonance between access control schemes for data models supported by operational DBMSs and those provided by data warehouse. For example, the relational model is the predominate data model in use today, while decision support systems tend to exploit analytical opportunities offered by non-traditional data models such as the star, temporal, snow flake, or multidimensional data models. The general lack of representation models for defining access controls further frustrates any process for deriving appropriate access controls at the data warehouse level from those used at the operational database level. In practice, the specification of security policies at the DBMS level is very rudimentary, and organizations rarely document their information system security policies. Finally, users of the operational systems are not the same as the users of the data warehouse, so an access control policy used for an operational system may have little resemblance to one appropriate for the data warehouse level. The solutions to these problems lie in a number of areas of database research. Research on security for heterogeneous databases can contribute to problems associated with integrating multiple and (potentially) inconsistent security policies. Better techniques for modeling access controls need to be considered, especially for the newer data models. At a conceptual level, the protection objects addressed by data warehouses are not tables and views, but key dimensions (e.g., shipments, locations, or dates), hierarchical paths (e.g., regions, countries, and cities), granularity distinctions between base facts and aggregates, and temporal concepts. Similarly, operations on the data warehouse are more complex than selects, projects, or joins; and include such terms as drill-down, pivot, and slice-and-dice.

Authentication. The security of a data warehouse rests on limiting access to it to those authorized to view the data. The internal process for controlling access to the data must match the external process for authenticating users. Most organizations have invested significantly in networks to link together hundreds, if not thousands, of computers that all have access to the data warehouse. The security administration burden associated with administering data warehouse accounts can be significant. Who determines who has access to what data? Is data access open by default or denied by default? How are passwords administered? How is the administrator notified when individuals leave the organization or move within organization? This issues are strongly reminiscent of those in federated database environments and research in that area will contribute significantly to data warehouse environments.

Similarly, efforts focused at combining research in role-based access controls with Internet security mechanisms may lead to a more rapid movement of technology out of the research arena into the commercial marketplace. In a data warehouse environment it is typically "what you are" rather than "who your are" that really determines a user's access rights. For example, a "policy" would be stated as allowing personnel staff or managers to have access to information relating to salaries. An implementation currently requires identifying individuals in those roles and establishing accounts for the individuals and mapping the accounts to privileges. Internet technologies are focusing on public key technologies, specifically X.509 certificates, not only as a basic authentication mechanism, but as a vehicle for providing attributes upon which access controls can be based. Examples of such attributes are organization, organization unit, and location.

Inference. Authentication, access control, and integrity are traditional security concerns for all systems. However, data warehouses, by their nature, are particularly subject to inference problems. Actually, inference is a dual-edged sword in the data warehouse environment. Decision support systems, such as data warehouses, rely on summary or aggregate information to feed the decision making process. Powerful data mining tools use inference techniques to infer information not explicitly available from the data. However, these same tools and techniques can lead to the unintended exposure of information. The inference problem is being addressed by on-going research. One of the more interesting approaches is that of using data mining tools on a collection of data to detect potential inference problems. Again, this is somewhat reminiscent of the techniques that have been useful in the development of network intrusion detection tools, which exploit the techniques used for computer break-ins as a means of detecting their occurrence (and, admittedly, this process has worked in the reverse).

Summary. Development of a data warehouse requires a major commitment on the part of an organization, in money and other resources. Poor data integrity and security together affect information timeliness, accuracy, and credibility; thereby, undermining the return-on-investment an organization can realize on its effort. This

paper has identified a number of problems that require solutions to achieve appropriate and workable security solutions in a data warehouse environment (see also [Kimb97]).

POSITION BY PIERANGELA SAMARATI

A data warehouse is a repository of integrated information obtained from distributed, autonomous, and possibly heterogeneous sources and made available for direct access. The peculiar characteristics of data warehousing can be summarized as follows:

- The warehouse stores data derived from the local data;
- Queries are processed at the warehouse, without need of consulting the local databases from which the data stored at the warehouse have been obtained;
- The warehouse must respect the autonomy of the local databases.

With respect to security enforcement, the first two aspects imply that control decisions on whether requests on data should be allowed, denied, or partially allowed must be taken at the data warehouse. Contacting the local databases to enforce access decisions would eliminate the efficiency advantages of having explicitly stored the data at the warehouse for direct processing. On the other hand, the third aspect requires the access control to take into consideration the protection requirements possibly expressed locally at the data sources.

An important issue to be addressed in this context is therefore the administration of access specifications. Access administration in particular determines the power left to the data warehouse (more properly to the administrator/s of the data warehouse with respect to the power of the local databases (more properly of the local administrator/s of the data) in deciding accesses to be allowed or denied. Questions to be answered include: Should the access restrictions specified at the data warehouse have an exact correspondence with the access restrictions specified at the local level? May the data warehouse allow access to some data even if the local sources from which the data have been obtained do not? May the data warehouse deny access to data even if local sources from which data have been obtained would like the access to be allowed? Who can specify access restrictions on data stored at the warehouse which do not have a correspondence to data at local systems (for instance historical data or data obtained at the warehouse by aggregating data from different sources). It is certainly difficult, if at all possible, to give a unique answer to these questions. The correct approach to be taken may depend on the specific application or data considered (see also [Morg92. Thur97]).

In some cases the protection requirements at the warehouse may have an exact correspondence with those stated locally. (Note that this does not necessarily require a one to one correspondence between the data stored at the data sources and the data stored at the warehouse; views can be defined locally for their inclusion in

the warehouse. For instance, consider the case of local databases storing classified data on which an unclassified view is defined for its inclusion in the warehouse.) When the data warehouse is required to obey to the protection constraints locally stated, the security rules to be enforced at the warehouse may be derived from those locally stated with a derivation/integration process similar to the process followed for obtaining data. This process may, in the case of security rules, be complicated by the fact that the data or security models at the different sources and at the warehouse may differ. In particular some systems may use mandatory access control policies while other systems may use discretionary access control policies. Even in the case when all systems use a mandatory or a discretionary policy, heterogeneity problems may arise. Within mandatory policies, different systems may user different classification lattices or support a different granularity for data classification. Within discretionary policies, different systems may use different authorization subjects (e.g., roles, groups, or user identities) or use different approaches to authorization specification (e.g., closed, open, or other hybrid approaches). This wide range of possible policies, requires the security model/mechanism at the warehouse to be as flexible as possible so to be able to mirror the different constraints that may need to be expressed. In particular, the data warehouse should implement a policy neutral mechanism able to enforce different policies rather than a specific mechanism with a single policy wired in [Jajo97b]. A problem that may arise when the specifications at the warehouse are derived from those locally stated is that local specifications may be inconsistent with each other. Inconsistencies arise when the data at the warehouse are obtained from different sources (because of integration or because data are replicated) which disagree on the access decision to be taken: one source states the access is to be allowed while the other states that the access is to be denied. Conflict resolution policies must be devised to resolve these inconsistencies.

Not always the security rules at the warehouse can be derived from, or faithfully represent, the constraints locally stated. The reason for this is that there may not be a direct correspondence between the data stored at the data sources and the data stored at the warehouse (which may be derived through aggregation or integration, or may represent historical data). Also, in some cases, the data warehouse may actually represent a separate system with its proper protection requirements and corresponding security rules. Moreover, security rules enforced at the warehouse may evolve over time, as the need for access may change. In such cases the security rules enforced at the data warehouse may be independent from those enforced at the local data sources. If the autonomy of the local sources is to be respected, acceptance of the policy enforced at the warehouse by the local administrator/s may be required. For instance, the local database administrator may allow inclusion of data in the warehouse agreeing to the fact that the data are subject to the policy established by the warehouse administrator/s.

In the most general case, therefore, access rules can be specified both at the local site, expressing protection requirements locally stated, and at the warehouse, expressing protection requirements at the general level and corresponding to the

access restrictions that will actually be enforced on the accesses. The relationship, if any, between the protection requirements stated at the data sources and those enforced at the warehouse may depend on the specific data or application considered.

Note that, regardless of whether the local administrator dictates the security rules to be enforced at the warehouse (like in the first scenario illustrated) or simply accept them (like in the second scenario illustrated), there is an implied trust of the local source on the protection requirements that will be enforced at the data warehouse. Assurance is certainly one of the biggest issue in data warehousing. Once data taken from the data source are acquired and stored at the warehouse, the local source looses control on data access. Access requests are processed at the warehouse, which has data available for direct access and the data source cannot actively participate in the access decision. The local sources must therefore trust the warehouse with respect to the enforcement of the security requirements they have specified or they have accepted (depending on the administration policy supported). Note also that the fact that access control cannot be executed locally may make it difficult or impossible to enforce particular controls. For instance, it is not possible for local sources to impose constraints based on local identities (or grouping) of the requesters as it was possible in traditional distributed or federated systems [Deca97]. Moreover, access constraints that could be enforced locally may not be enforceable at the data warehouse. This happens for instance when local constraints impose conditions on the basis of data which are not available at the warehouse. To illustrate, consider a site that stores a table Sales with name, addresses, and purchase of customers. Suppose that from such a table, a table Largesales is stored at the data warehouse containing the name of the customers who have made purchase for more than $5,000. An access restriction such as: "Managers can read information about customers residing in California" which can be easily enforced at the local source cannot be obviously enforced at the warehouse, where the address information is not available. Another example where constraints may not be easily enforced at the warehouse is when data are obtained from aggregation of local data. For instance, consider again table Sales above. Suppose that, at the data warehouse, a table Totalsales is stored containing, for each branch, the total amount of merchandise sold. Again, since managers can read only information about purchases made by customers residing in California, only the sum of such purchases should be made available to the managers. However, such values cannot be obtained from the data available at the data warehouse.

POSITION BY T.Y. LIN

Rules Mining and Security: Over last decade the emphasis of databases has begun to shift from the technology of retrieving and storing simple data to that of high level information. This leads to the notion of data warehouses. Data mining, though has its own interests, is one of a very important technology in data

warehouses. A center piece of data mining is rule mining, and rough set theory is a theory on it. In this section, we will focus on rough set theory for very large databases [Lin96a, Lin96b, Lin97a] and its security implications.

What is rough set theory? In a concrete format, it is a theory of information tables (also known as Pawlak information systems). In its abstract form, it is a topological theory (e.g., approximation) of equivalence relations. There are plenty of literature devoted to rough sets (e.g., [Pawl91, Lin97a). It also have been discussed from the security prospect in a previous volume of this series [Lin96b]. Loosely speaking an information table is a relation instance. We will, however instead of giving a formal definition, illustrate how an information table is derived.

Rough Sets and Data Representations: Let U = {ID1, ID2, ... ID9} be the universe of discourse, in the example we are going to illustrate, it is a set of 9 balls. Assume U has been partitioned into equivalence classes by their colors. Recall that in rough set community, equivalence classes are called elementary sets and their names elementary concept. For this example, we name these elementary sets Red, Orange, and Yellow. Suppose these balls can also be classified by their weights {W1, W2, W3, W4}. We have an information table that represents two equivalence relations, color and weight.

BALLs	Color	Weight
ID-1	Red	W1
ID-2	Red	W1
ID-3	Red	W2
ID-4	Orange	W3
ID-5	Orange	W3
ID-6	Yellow	W4
ID-7	Yellow	W4
ID-8	Yellow	W4
ID-9	Yellow	W4

A universe together with a set of equivalence relations is called a knowledge base [Pawlak91]. So this table represents the knowledge base (U, color, weight). There is a one-to-one correspondence between knowledge bases and information tables.

Security and Rules in Very Large Databases: Rough set theory is an elegant and powerful theory in extracting and minimizing the set of rules from information tables. The central notions of the theory are cores, reducts, and knowledge dependencies. Skowron and Rauszer have shown us a discouraging phenomenon, namely, finding the minimal reducts is a NP-Hard problem [Skow92]. So the computational complexity have implicitly restricted its effective applications to a small data set. In [Lin96a] rough set theory is extended to very large databases with some sacrificing on its elegance. First, database searching techniques are used to

filter a family of small and clean information tables. Then we apply rough set theory to such a family of small nice sets of clean data. One of the main filtering criteria is the high frequency rates of the appearances of data [Lin1996c, Lin97a]. For example, assume that the information discussed earlier has about 10 millions rows and a particular tuple, say (Yellow, W4), appears repeatedly for 100,000 times. In such a case, the pair represents a very robust pattern.

It seems clear that the higher the frequency the more robust the pattern becomes. So one might think that the robust information are precious and reliable. In the context of security, one might want to classify them as high security information. However, somewhat surprisingly, several experiments indicate a twist in this direction of thoughts. It turns out that any "very" robust information is often known to many people from other sources too. So in security worlds, it means that security officers have to examine whether those newly extracted robust information are "common sense" to unclassified users or not. This implies that data mining of "high frequency rules" is quite important to database security from an unexpected prospect; further studies will be reported in near future.

POSITION BY SUSHIL JAJODIA

Data Warehousing and Threat to Privacy: Data warehouses typically do not contain data used for day-to-day data processing in organizations; they have become information systems that store everything, whether it is vital or not to an organization. With rapid advancements in computer and network technology, it is possible for organizations to collect, store, and retrieve vast amounts of data of all kinds quickly and efficiently.

The data warehouses represent a threat to personal privacy since they contain a great amount of detail about individuals. Admittedly, the information collection function is essential for an organization to conduct its business; however, indiscriminate collection and retention of data can represent an extraordinary intrusion on privacy of individuals. The basic principles for achieving information privacy are listed in [Jajo97a]. These principles are made more concrete when specific mechanisms are proposed to support them.

- Proper acquisition and retention is concerned with what information is collected and after it is collected how long it is retained by an organization.
- Integrity is concerned with maintaining information on individuals that is correct, complete, and timely. The source of the information should be clearly stated, especially when the information is based on indirect sources.
- Aggregation and derivation of data is concerned with ensuring that any aggregations or derivations performed by an organization on its information are necessary to carry out its responsibilities. Aggregation is the combining of information from various sources. Derivation goes one step further; it uses different pieces of data to deduce or create new or previously unavailable

information from the aggregates. Aggregation derivation are important and desirable effects of collecting data and storing them in databases; they become a problem, however, when legitimate data is aggregated or used to derive information that is either not authorized by law or not necessary to the organizations.

Aggregates and derived data pose serious problems since new information can be derived from available information in several different ways. Nonetheless it is critical that data be analyzed for possible aggregation or derivation problems. With a good understanding of the ways problems may arise, it should be possible to take steps to eliminate them.

- Information Sharing is concerned with authorized or proper disclosure of information to outside organizations or individuals. Information should be disclosed only when specifically authorized and used solely for the limited purpose specified. This information should be generally prohibited from being redisclosed by requiring that it be either returned or properly destroyed when no longer needed.

- Proper Access is concerned with limiting access to information and resources to authorized individuals who have a demonstrable need for it, in order to perform official duties. Thus, information should not be disclosed to those that either are not authorized or do not have a need-to-know (even if they are authorized).

Although there is overlap in principle between security and privacy, there are significant differences between their objectives.

Privacy protection is a personal and fundamental right of all individuals. Individuals have a right to expect that organizations will keep personal information confidential. One way to ensure this is to require that organizations collect, maintain, use, and disseminate identifiable personal information and data only as necessary to carry out their functions. In the U.S., Federal privacy policy is guided by following key legislations:

- **Freedom of Information Act of 1966** -- It establishes an openness in the Federal Government by improving the public access to the information. Under this act, individuals may make written requests for copies of records of a department or an agency that pertain to them.

- **The Privacy Act of 1974** -- It provides safeguards against the invasion of personal policy by the Federal Government. It permits individuals to know what records pertaining to them are collected, maintained, used, and disseminated.

- **Computer Matching and Privacy Protection Act of 1988** -- It states that agencies must follow specific procedures when engaging in the automated comparison of Privacy Act databases on the basis of certain data elements.

In 1996 and in 1997, U.S. Federal Trade Commission's Bureau of Consumer Protection (www.ftc.gov) conducted two public workshops to determine if federal legislation is required for privacy. Industry on the other hand would prefer self-regulation.

POSITION BY CHRIS CLIFTON

The goal of data mining technology is knowledge discovery, to find 'knowledge' that is otherwise hidden by large volumes of data. This immediately points to a potential security risk; if the knowledge is hidden, how do we know that a security risk exists? In many ways this is the opposite of raised by statistical/summary queries. With statistical queries, the value to be protected is known: the individual values of data instances. With data mining, the source is known (the data instances), but we do not know what it is we are trying to protect.

We are interested in problems where the security of individual data items is not a concern, but there may be patterns in the data that pose a security risk. A few examples of risks posed by patterns in the data include:

- **Prediction of sensitive information.** For example, suppose major corporate announcements required a face-to-face meeting of senior management from various locations. In addition, negative announcements required participation of senior public relations staff. Travel records (likely made available to an external travel agent) could then be used to predict the occurrence and type of corporate announcements. Here the individual travel records are not a concern, but their correlation with past announcements poses a risk.

- **Misuse of information.** Suppose we are in the business of gathering and packaging publicly available information. A bank may use this information in evaluating credit risks, finding patterns linking known good and bad risks to the information we provide. Note that these patterns will not necessarily be a causal relationship (and need not even be a strong relationship); public outcry over use of such information could well be directed against the information provided (witness credit reporting agencies current public relations dilemmas).

We are investigating ways to limit the use of data for mining, while preserving usefulness for the intended function. There are a number of approaches, such as 'preemptive' data mining (find sensitive patterns before making the data available), ensuring that no unnecessary information is provided (e.g., using U.S. Social Security numbers as personal identifier; these encode the issuing office) or adding

'false' information to generate false patterns (useful where the intended use of the information is to get specific values, e.g., a telephone book). Many of these have the drawback that either the intended use of the data, or the knowledge to be protected, must be clearly specified.

Our current focus is on limiting the amount of data available; patterns found in a small enough sample are likely to be artifacts of the sample size rather than reflecting real-world patterns. The question we are trying to answer is how small is small enough. Our goal is to provide provable limits on the reliability of information given the sample size and characteristics of the data.

REFERENCES

[Deca97] S. De Capitani di Vimercati and P. Samarati. Authorization Specification and Enforcement in Federated Database Systems. Journal of Computer Security, 1997.

[Devl97] Barry Devlin, Data Warehouse from Architecture to Implementation, Addison-Wesley, 1997.

[Inmo97] W. H. Inmon, J. D. Welch, and Katherine L. Glassey. Managing the Data Warehouse, John Wiley & Sons, Inc., New York, 1997.

[Jajo97a] S. Jajodia "Database security and privacy," ACM Computing Surveys, 50th anniversary commemorative issue, Vol. 28, No. 1, March 1996, pages 129-131.

[Jajo97b] S. Jajodia, P. Samarati, and V.S. Subramanian. A Logical Language for Expressing Authorizations. In Proceedings of the 1997 IEEE Symposium on Security and Privacy, May 1997.

[Kimb97] Ralph Kimball, Hackers, Crackers, and Spooks, Ensuring that Your Data Warehouse is Secure, DBMS, April 1997, pp. 14-16.

[Lin96a] T.Y. Lin, Rough Set Theory in Very Large Databases, Symposium on Modeling, Analysis and Simulation, IMACS Multi Conference (Computational Engineering in Systems Applications), Lille, France, July 9-12, 1996, Vol. 2 of 2, pp. 936-941.

[Lin96b] T. Y. Lin and Rayne Chen, Supporting Rough Set Theory in Very Large Database Using ORACLE RDBMS, Soft Computing in Intelligent Systems and Information Processing, Proceedings of 1996, Asian Fuzzy Systems Symposium, Kenting, Taiwan, December 11-14, 1996, pp. 332-337 (Co-author: R. Chen).

[Lin96c} T. Y. Lin, T. Hinke, D. Marks, B. Thuraisingham, Security and Database Mining, Database Security IX Status and Prospects, Edited by D. L. Spooner, S. A. Demurjian and J. E. Dobson, 1996, pp. 391-399.

[LIN97a] T. Y. Lin and Rayne Chen, Finding Reducts in Very Large Databases, Proceedings of Joint Conference of Information Science, Research Triangle Park, North Carolina, March 1-5, 1997, pp. 350-352. (Co-author: R. Chen).

[Lin97b] T. Y. Lin, N. Cercone, Rough Sets and Data Mining: Analysis of Imprecise Data, Kluwer Academic Publishers, 1997.

[Morg97] Matthew Morgenstern, Teresa F. Lunt, Bhavani Thuraisingham, and David L. Spooner. Security Issues in Federated Database Systems: Panel Contributions. In C. E. Landwehr and S. Jajodia, editors, Database Security, V:Status and Prospects, pp. 131-148, 1992.

[Pawl91] Z. Pawlak, Rough sets: Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, 1991.

[Rose97] Arnon Rosenthal, Paul A. Dell, Pamela D. Campbell, Integrity and Security in Data Warehousing, AFCEA,1997.

[Ross96] Steven J. Ross, "Control Issues in Data Warehousing," Infosecurity News, July/August 1996, pp. 22-24.

[Skow92] A. Skowron, C. Rauszer, The discernibility matrices and functions in information systems, Decision Support by Experience - Application of the Rough Sets Theory, R. Slowinski (ed.), Kluwer Academic Publishers, 1992, pp. 331-362.