The Inference Problem: A Survey

Csilla Farkas
Dept. of Computer Science and Engineering
University of South Carolina
Columbia, SC 29208
farkas@cse.sc.edu

Sushil Jajodia
Center for Secure Information Systems
George Mason University
Fairfax, VA 22030-4444
jajodia@gmu.edu

ABSTRACT

Access control models protect sensitive data from unauthorized disclosure via direct accesses, however, they fail to prevent indirect accesses. Indirect data disclosure via inference channels occurs when sensitive information can be inferred from non-sensitive data and metadata. Inference channels are often low-bandwidth and complex; nevertheless, detection and removal of inference channels is necessary to guarantee data security. This paper presents a survey of the current and emerging research in data inference control and emphasizes the importance of targeting this so often overlooked problem during database security design.

Keywords

Inference control, data security, semantic modeling, access control, external knowledge

1. INTRODUCTION

Working as a secretary for a computer manufacturer, Jane is eager to be promoted to marketing agent. Unfortunately, her competition, John, always has innovative ideas about how to improve the business to increase revenue. If only she could look at his files on market research! Since John is very careful to protect his data and uses encrypted e-mail, accessing these files is not an option. However, as a secretary, Jane has access to John's phone bills and Web usage logs. She notices that John visited Web sites of printer manufacturers and made several phone calls to ink cartridge suppliers. According to these Web sites, printer manufacturers gain large profits from cartridge sales for their printers. Jane deduces that John's recommendation at the annual company meeting will be to expand the company's profile to manufacture and market printers. Using this information and the data available on the Web, Jane writes a proposal recommending her company to manufacture printers. She sends her proposal, containing the expected profit to be gained, to her boss. By the time John presents his ideas at the annual company meeting, his ideas, assumed to be originated from Jane, are already being considered by the company leaders. Although fictional, the above example illustrates the possibility of security (as well as privacy) violations via inferences. Despite of the difficulties to develop techniques to detect potential inference vulnerabilities, no system can be called secure without them. Access control models offer protection against direct accesses to sensitive information; however, indirect accesses to sensitive data may still be possible via inferences. The inference problem in databases occur when sensitive information can be disclosed from non-sensitive data and metadata (see Figure 1). Metadata may refer to database constraints, like database dependencies and integrity constraints, or outside information, like domain knowledge and query correlations. Depending on the level of accuracy by which the sensitive information is revealed, full disclosure or partial disclosure may occur.

This paper surveys the research efforts seeking to address the inference problem. Sections 2 and 3 provide a brief review of research efforts in statistical and multilevel secure databases, respectively. In Section 4, we review works focusing on the inference problem in general purpose databases. In Section 5, we show the inference risk raised by data mining, and the current security focus. Section 6 addresses inference problems in Web-based applications. Finally, in Section 7 we conclude by analyzing the current state of inference protection and identify necessary security tools to protect tomorrow's database systems .

2. STATISTICAL DATABASES

Privacy violations via inferences were first considered in statistical databases. The security requirement in statistical databases is to provide access to statistics about groups of entities while protecting the confidentiality of the individual entities. The problem is that a user might obtain confidential information about an individual entity by correlating different statistics. Inference control in statistical databases

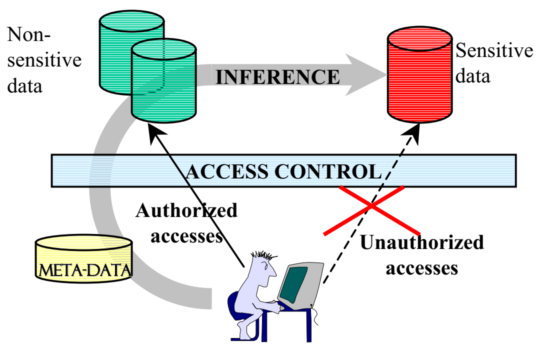

Figure 1: Indirect information access via inference channels

has been extensively studied [25; 53; 44; 30; 40]. A number of inference control mechanisms, such as query size and query overlap control, data swapping, and multidimensional transformation, were developed and their limitations established. The main problem is that simple inference control mechanisms are easily subverted and, therefore, insufficient. Mechanisms that provide confidentiality with high assurance are often complex and difficult to implement, or not applicable in general purpose databases and thus are limited to certain applications (e.g., U.S. Census Bureau database).

3. MULTILEVEL SECURE DATABASES

Papers considering the inference problem in multilevel secure databases emerged during the early 1980s. Most of these works focused on defining the problem and providing frameworks to address specific types of inferences. Jajodia and Meadows give a comprehensive survey of the research efforts in [39]. They provide overview of the different inference channels and present techniques to detect and remove them. They show that most of the inference channels in multilevel secure databases are created by combining *metadata* (e.g., database constraints) with data in order to obtain information that has a higher security classification than the original data.

Techniques to detect and remove inference channels can be organized into two categories. The first category includes techniques that detect inference channels during database design. Such channels can be removed by either modifying the database design or increasing the classification levels of some of the data items [15; 33; 36; 66; 45; 48; 60; 63; 65]. These techniques often result in over-classification of data and, therefore, reduce the availability of data.

Techniques in the second category seek to eliminate inference channel violations during query time [26; 46; 61; 67]. If an inference channel is detected, the query is either refused or modified to avoid security violations. This technique allows inference detection at both data and schema level. While data level inference detection allows increased data availability (since only the data released to the user is considered for inference), it is computationally expensive.

Most of the early works on developing frameworks and actual inference algorithms addressed specific inference problems. For example, Denning [26] recommends an authorized view equivalence schema to remove any unauthorized data from the answers to select-only, select-project, and select-project-join queries. However, she does not address the inference problem in the presence of database constraints. Hinke [36] presents a semantic-graph based inference detection tool to represend semantic relationships and to detect inference channels. The work of Thuraisingham [67] presents a general and powerful logic-based framework.

4. INFERENCES IN GENERAL PURPOSE DATABASES

Early works on inferences in general purpose databases built the foundation for future research, identified the need to provide formal characterization for database inferences, and provide assessment methods of the achieved security. There is a recent increase in the research on the inference problem. Researchers in both statistical and general purpose databases have taken a fresh look to this problem and developed solutions that provide high assurance. New models to address the problem in a systematic, formal way and evaluation methods of inference algorithms are being developed. Recent research [8; 13; 17; 20; 21; 23; 34; 37; 45; 73] in inference control focuses on issues like minimal classification updating, partial disclosure, classifying existing data repositories, and to prevent inferences via knowledge discovery.

Research efforts led by Hinke and Delugach [66; 23; 37; 38] developed method for automated analysis of database inferences. Their aim is to develop general purpose inference detection techniques that is applicable to a wide variety of databases and corresponding semantics. Representation of external knowledge and domain knowledge is also addressed. The authors propose a conceptual graph-based method to detect illegal inferences. Database entities and activities, relationships between them, domain knowledge, and data sensitivity is represented in the graph. The graph is manipulated by inference rules to derive new inferences. Illegal inferences are detected if there exists a path from unclassified information to classified information.

Hinke et al. also developed several techniques based on the conceptual graph analysis and implemented prototypes of inference detection systems: *Merlin, AERIE*, and *Wizard*. For example, Wizard takes as input a database schema, its instances, and domain knowledge, and generates associations between entities and/or activities that may create an inference channel. To incorporate external knowledge, the system uses human-aided microanalysis technique to add semantic knowledge to the graph.

In addition to the need to develop automated inference detection techniques, it is necessary to provide assurance of the detected inferences. Two main directions can be observed: 1) techniques to handle imprecise inferences and 2) techniques to formally evaluate the correctness of the detected inferences.

Hale and Shenoi [35; 34] addresses issues related to partial (imprecise) inferences. Partial inferences occur when an unauthorized user is able to infer the value (or a set of values) for a data item with certain probability. The authors address the imprecise inference problem at the presence of functional dependencies (FDs) in relational databases. They use abduction and partial deduction techniques, similar to Morgenstern's sphere of inference [48], to derive probabilistic inferences. External knowledge, introduced as catalytic databases, is considered to generate probabilistic inferences. They authors argue that although any information derived via catalytic inferences is imprecise in nature, the granularity of the disclosed data item may be small enough to create a security breach.

Brodsky et al. [13] focus on developing a general model to represent database and domain knowledge, and provide assurance of the inference detection technique. Their model focuses on precise inferences and uses logic-based techniques to derive inferences and to prove the correctness (soundness and completeness) of the developed inference disclosure algorithms. They present a security architecture, called Disclosure Monitor (DiMon), that is built upon lattice-based access control model. DiMon is able to enforce content, context, and history-based access requirements, while preventing illegal inferences. DiMon can function in data-dependent or data-independent mode. For data-dependent mode, the inference algorithm is sound and complete, for data-independent mode, the disclosure algorithm is complete but not sound.

Dawson et al. [20; 21] focus on the problem of classifying existing databases by using explicit classification constraints and association and inference constraints. Their model addresses both precise and imprecise inferences. An inference exists from an element in a set of data to an other element in a different set of data, if the second element can be derived from the first set, or if the set of possible values for the second element is reduced using the first set. The authors provide a minimal upgrading of data classifications to remove illegal inferences.

5. DATA MINING AND THE INFERENCE PROBLEM

External and domain knowledge plays a significant role in deriving inferences in databases. It is unrealistic to assume that all this information is known in advance, thus enable inference free database design, or is known by the security officer, thus incorporated in the security model. Moreover, data dependencies, specific to some databases, are unknown by the domain experts. Current research focusing on data mining may prove to be the right tool to extend security models for databases. Unfortunately, it is also a dangerous weapon, that can be used by malicious users to subvert security mechanisms.

With the increase of electronically available information, data mining represent an even greater risk than in centralized databases. Information originating from different sources can be analyzed. The goal of data mining applications is to extract pattern that support research and applications. Since data mining extracts higher-level information (metadata), it may represent serious security threats, similar to the inference problem.

Only a few researchers [50; 55; 51; 41; 59; 68] have addressed the problem of security threats via data mining. Security threats based on data mining can be addressed either before any mining activity is allowed (preprocessing) or during data mining (run-time). For preprocessing, a set of mining tools are applied on the database, to check whether sensitive information can be disclosed from the learned patterns. For run time, the inference controller evaluates the result to a user's request, and permits or rejects the release of the result based on this evaluation. In either mode, *data mining abilities* are reduced. Moreover, none of these approaches protect from inferences when the pattern discovered in one database is applied on a different database. It is unrealistic to assume that all (semantically) related databases would enforce the same security policy.

A different approach, presented in [31], classifies discovered metadata based on the range of its applicability. This approach assumes that, in most cases, it is not the discovery of a pattern that causes privacy violations but the unauthorized use of this pattern. Security threats associated with metadata are classified into two groups: 1) the discovered metadata is sensitive (e.g., pattern of drug usage of a patient), and 2) the discovered metadata itself is not sensitive, but when combined with additional data, sensitive information can be obtained (e.g., patients' high risk of disease based on matching a corresponding pattern).

Recently, several works addressing privacy preserving data mining [17; 3; 16; 43; 2] have surfaced that. Their main motivation is to allow data accesses for mining purposes, while preserving the confidentiality of the data. Techniques such as data estimation, perturbation and sample size restrictions are used to remove any unwanted inferences. The main aim of this research is to apply minimal modification to the original data without disturbing the data mining results. In addition to secure data mining, efficient methods of data sharing is important in distributed settings because of the large size of the involved databases.

6. WEB-BASED INFERENCES

During the 1990s, with the further development of World Wide Web, new privacy problems surfaced [1; 4; 9; 12; 18; 29; 47; 49; 52; 56; 58; 62; 69; 70; 71; 72]. Simultaneously, works to provide control accesses to documents in eXtensible Markup Language (XML) format surfaced [7; 6; 19; 42]. These models focus on defining access controls on XML documents, thus preventing privacy violations via direct data accesses. While these mechanisms are necessary, none of the above works provide technical solutions to enforce privacy requirements in the presence of possible inferences, or give assurance on the level of protection. A view-based XML access control model that preserves data semantics and eliminates inferences based on the existence of sensitive data is presented in [64].

Development of technologies to support the Semantic Web [5] increases the risk of illegal data accesses via inference channels. Automated analysis allows software agents to integrate large amount, possibly distributed data. Such integration is impossible for humans because of the size of the data sources. The World Wide Web was designed for humans, where syntactic constructs allowed users to interact and share information. The envisioned Semantic Web is build upon the assumption of intelligent information processing, providing means for interoperations. Software agents, encompassing powerful reasoning abilities, and ontologies, to provide domain knowledge, will be present in future Web.

Few researchers have considered the security threats presented by technologies developed for interoperation. The main focus of Semantic Web research is to provide interoperation and intelligent query processing [14; 22; 24; 27; 28; 32; 54]. Semantically rich Web technologies, like RDF and ontologies, create inference possibilities. While these inferences are considered from the perspective of enabling machine processing of the Web, there is no comprehensive security analysis available. An initial analysis of security threats raised by the inferencing capability of semantic Web is given in [64].

Works that support interoperation and intelligent data accesses on the Web do not address the associated security issues. Most of the current works in access control for Web documents revolves around developing languages and techniques for XML documents. While these works are clearly needed, additional considerations addressing the problem of indirect accesses via inference channels need to be made.

Some of the illegal inference problems that need to be addressed are based on replicated data, RDF-based inferences, and ontologies. Techniques to detect data replication with inconsistent security classifications and in the presence of ontologies need to be developed. Furthermore, semantic-based data correlations, supported by ontologies, need to be addressed. The derived information may be sensitive and should not be derivable from the released (non-sensitive) data. Existing inference control technologies are insufficient to address the above problems in the Web environment.

Tracing user collaborations also pose a challenge for future security, where it might be desirable to support anonymous access to the information resources. However, methods to prevent users sharing their data or the same user to login with different virtual idenities must be developed. Anonymous collaborations have been studied by several researchers, however, the risk of illegal inferences have not been addressed.

7. THE FUTURE

Successful use of the Semantic Web will depend on the implementation and use of Web services. Information, available on the Web, will be gathered and analyzed by collaborating agents. Automated processing will allow agents to utilize large amount of information that is beyond human processing power. However, this enhanced processing power can be misused by malicious users and their agents to disclose sensitive information or sabotage other's information. Data replication and intelligent information correlation opens up new dimensions in the inference problem. Ontologies (metadata) allows data integration as well as secure and unsecure inferences. Scalability and data quality issues as well as inherent vulnerabilities of the underlying technologies need to be addressed to prevent security violations.

Agent technologies are the fundamental constructs of the Semantic Web. Security and reliability of these technologies are necessary to provide secure Semantic Web applications. Agent-based systems were designed for interoperability, distributed problem solving, and cooperation without the necessary security safeguards. Only recently, works to provide rudimentary protection for agents and agent platforms have emerged. These works focus on direct data accesses, and secure and privacy preserving agent events. Current research by [11; 10; 57] addresses issues related to secure communication and mobile-computing.

Collaboration and information sharing are highly desirable features for agent interoperation. However, they may represent security vulnerabilities. Inferences, based on agent behavior, usage monitoring, and application of "trick questions", may also occur. Also, even though each agent may behave in the desired and secure way, their combined knowledge may disclose sensitive data.

Security threat to agents by being monitored, tested on fake data, or supplied with malicious code has not been sufficiently addressed by the research community. Unfortunately, either of the above examples could easily happen to compromise the corresponding agent. Techniques, similar to the inference control in databases may be applicable to enhance multi-agent platform security.

8. ACKNOWLEDGEMENTS

Farkas and Jajodia were partially supported by the National Science Foundation under grants SFS-0112874 and CCR-0113515, respectively.

9. REFERENCES

[1] Genetic information and the workplace. Technical report, Department of Labor,Department of Health and Human Services, Equal Employment Opportunity Commission Department of Justice, 1998.

[2] D. Agrawal and C. C. Aggarwal. On the design and quantification of privacy preserving data mining algorithms. In *Symposium on Principles of Database Systems*, 2001.

[3] R. Agrawal and R. Srikant. Privacy-preserving data mining. In *Proc. of the ACM SIGMOD Conference on Management of Data*, pages 439–450. ACM Press, May 2000.

[4] J. G. Anderson. Clearing the way for physicians' use of clinical information systems. *Communications of ACM*, 40(8):83–90, August 1997.

[5] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American*, May 2001.

[6] E. Bertino, M. Braun, S. Castano, E. Ferrari, and M. Mesiti. A java-based system for XML data protection, 2000.

[7] E. Bertino, S. Castano, E. Ferrari, and M. Mesiti. Controlled access and dissemination of XML documents. In *Workshop on Web Information and Data Management*, pages 22–27, 1999.

[8] J. Biskup and P. A. Bonatti. Controlled query evaluation for known policies by combining lying and refusal. In *Foundations of Information and Knowledge Systems*, pages 49–66, 2002.

[9] J. Biskup and H. H. Bruggemann. The personal model of data - towards a privacy oriented information system (extended abstract). In *Proc. of the Fifth International Conference of Data Engineering, February 6–10, 1989, Loas Angeles, California, USA*, pages 348–355. IEEE Computer Society, 1989.

[10] P. Bonatti, S. Kraus, and V. S. Subrahmanian. Secure agents. Technical Report CS-TR-4068, 1999.

[11] P. A. Bonatti, S. Kraus, J. Salinas, and V. S. Subrahmanian. Data-security in heterogeneous agent systems. *Lecture Notes in Computer Science*, 1435, 1998.

[12] B. Braithwaite. National health information privacy bill generates heat at SCAMC. *Journal of Americal Informatic Association*, 3(1):95–96, 1996.

[13] A. Brodsky, C. Farkas, and S. Jajodia. Secure databases: Constraints, inference channels, and monitoring disclosure. *IEEE Trans. Knowledge and Data Eng.*, 12(6):900–919, November/December 2000.

[14] J. Broekstra, A. Kampman, and F. van Harmelen. Sesame: An architecture for storing and querying rdf data and schema information, 2001.

[15] L. Buczkowski. Database inference controller. In D. Spooner and C. Landwehr, editors, *Database Security III: Status and Prospects*, pages 311–322. North-Holland, Amsterdam, 1990.

[16] C. Clifton. Using sample size to limit exposure to data mining. *Journal of Computer Security*, 8(4), 2000.

[17] C. Clifton and D. Marks. Security and privacy implications of data mining. In *Workshop on Data Mining and Knowledge Discovery*, number 96-08, pages 15–19, Montreal, Canada, 1996.

[18] N. R. Council. For the record: Protecting electronic health information. Technical report, National Academy of Sciences, 1997.

[19] E. Damiani, S. D. C. di Vimercati, S. Paraboschi, and P. Samarati. Design and implementation of an access control processor for XML documents. *WWW9 / Computer Networks*, 33(1-6):59–75, 2000.

[20] S. Dawson, S. D. Capitano, and P. Samarati. Specification and enforcement of classification and inference constraints. In *Proc. of the 20th IEEE Symposium on Security and Privacy*, May 1999. Oakland.

[21] S. Dawson, S. D. C. di Vimercati, P. Lincoln, and P. Samarati. Minimal data upgrading to prevent inference and association. In *Proceedings of the Eighteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, May 31 - June 2, 1999, Philadelphia, Pennsylvania*, pages 114–125. ACM Press, 1999.

[22] S. Decker, S. Melnik, F. van Harmelen, D. Fensel, M. C. A. Klein, J. Broekstra, M. Erdmann, and I. Horrocks. The semantic web: The roles of XML and RDF. *IEEE Internet Computing*, 4(5):63–74, 2000.

[23] H. Delugach and T. Hinke. Wizard: A database inference analysis and detection system. *IEEE Trans. on Knowledge and Data Engineering*, 8(1):56–66, 1996.

[24] G. Denker, J. R. Hobbs, D. Martin, S. Narayanan, and R. J. Waldinger. Accessing information and services on the DAML-enabled web. In *SemWeb*, 2001.

[25] D. Denning. *Cryptography and Data Security*. Addison-Wesley, Mass., 1982.

[26] D. Denning. Commutative filters for reducing inference threats in multilevel database systems. In *Proc. IEEE Symp. on Security and Privacy*, pages 134–146, 1985.

[27] A. Deutch, M. Fernandez, D. Florescu, A. Levy, and D. Suciu. A query language for XML. In *Proc. In International Conference on World Wide Web*, 1999.

[28] A. Deutsch, M. Fernandez, D. Florescu, A. Levy, and D. Suciu. A query language for XML. *Computer Networks (Amsterdam, Netherlands: 1999)*, 31(11–16):1155–1169, 1999.

[29] L. C. J. Dreyer and M. S. Olivier. Dynamic aspect of the infopriv model. In *Proc. 9th Database and Expert Systems Applications DEXA 98*, pages 340–345. IEEE Computer Society, Los Alamitos, 1998.

[30] G. Duncan and S. Fienberg. Obtaining information while preserving privacy: a markov perturbation method for tabular data. In *Statistical Data Protection*, pages 351–362, 1998.

[31] C. Farkas, S. Fenner, and M. Valtorta. Medical privacy versus data mining. In *Proc. Fifth Multiconference on Systemics, Cybernetics and Informatics*, pages 194–200, July 2001.

[32] G. Gardarin and F. Sha. Using conceptual modeling and intelligent agents to integrate semi-structured documents in federated databases. *Lecture Notes in Computer Science*, 1565:87–99, 1999.

[33] J. Goguen and J. Meseguer. Unwinding and inference control. In *Proc. IEEE Symp. on Security and Privacy*, pages 75–86, 1984.

[34] J. Hale and S. Shenoi. Catalytic inference analysis: Detecting inference threat due to knowledge discovery. In *Proc. of the 1997 IEEE Symposium on Security and Privacy*, pages 188–199, May 1997. Oakland.

[35] J. Hale, J. Threet, and S. Shenoi. A practical formalism for imprecise inference control. *IFIP Trans. Computer Science And Technology*, 60:139–156, 1994.

[36] T. Hinke. Inference aggregation detection in database management systems. In *Proc. IEEE Symp. on Security and Privacy*, pages 96–106, 1988.

[37] T. Hinke, H. Delugach, and R. Wolf. A framework for inference directed data mining. In *Proc. 10th IFIP WG11.3 Workshop on Database Security*, pages 229–239, 1996.

[38] T. Hinke, H. S. Delugach, and R. P. Wolf. Protecting databases from inference attacks. *Computers and Security*, 16(8):687–708, 1997.

[39] S. Jajodia and C. Meadows. Inference problems in multilevel secure database management systems. In M. Abrams, S. Jajodia, and H. Podell, editors, *Information Security: An integrated collection of essays*, pages 570–584. IEEE Computer Society Press, Los Alamitos, Calif., 1995.

[40] A. Karr, J. Lee, A. Sanil, J. Hernandez, S. Karimi, and K. Litwin. Web-based systems that disseminate information from data but protect confidentiality. *IEEE Computer*, February 2001. http://www.niss.org/dg/technicalreports.html.

[41] W. Klosgen. Knowledge discovery in databases and data privacy. In *IEEE Expert*, April 1995.

[42] M. Kudo and S. Hada. XML document security based on provisional authorization. In *Proc. of the 7th ACM Conference on Computer and Communication Security*, November 2000.

[43] Y. Lindell and B. Pinkas. Privacy preserving data mining. *Lecture Notes in Computer Science*, 1880:36–??, 2000.

[44] T. Lunt. Current issues in statistical database security. In C. Landwehr and S. Jajodia, editors, *Database Security, V: Status and Prospects, IFIP WG 11.3*, pages 381–385, 1991.

[45] D. Marks. Inference in MLS database systems. *IEEE Trans. Knowledge and Data Eng.*, 8(1):46–55, February 1996.

[46] S. Mazumdar, D. Stemple, and T. Sheard. Resolving the tension between integrity and security using a theorem prover. In *Proc. ACM Int'l Conf. Management of Data*, pages 233–242, 1988.

[47] B. N. Meeks. Privacy lost, anytime, anywhere. In *Communications of ACM*, volume 40/8, pages 11–13, 1997.

[48] M. Morgenstern. Controlling logical inference in multilevel database systems. In *Proc. IEEE Symp. on Security and Privacy*, pages 245–255, 1988.

[49] U. S. G. A. Office. Medical records privacy, access needed for health research, but oversight of privacy protections is limited. Technical report, United States General Accounting Office, Report to Congressional Requesters GAO/HEHS-99-55, 1999.

[50] D. O'Leary. Knowledge discovery as a threat to database security. In G. Piatetsky-Shapiro and W. Frawley, editors, *Knowledge Discovery in Databases*, pages 507–516. AAAI Press/The MIT Press, Menlo Park, California, 1991.

[51] D. O'Leary. Some privacy issues in knowledge discovery: OECD personal privacy guidelines. In *IEEE Expert*, April 1995.

[52] D. E. O'Leary. Some privacy issues in knowledge discovery: Oecd personal privacy guidelines. *IEEE Expert/Intelligent Systems and Their Applications*, 10(2), April 1995.

[53] G. Ozsoyoglu and T. Su. On inference control in semantic data models for statistical databases. *Journal of Computer and System Sciences*, 40(3):405–443, 1990.

[54] Y. Papakonstantinou and V. Vianu. DTD Inference for Views of XML Data. In *Proceedings of the Nineteenth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, pages 35–46, Dallas, Texas, 2000.

[55] G. Piatetsky-Shapiro. Knowledge discovery in databases vs. personal privacy. In *IEEE Expert*, April 1995.

[56] T. C. Rindfleisch. Privacy, information technology, and health care. *Communications of ACM*, 40(8):93–100, August 1997.

[57] V. Roth and M. Jalali-Sohi. Concepts and architecture of a security-centric mobile agent server. In *ISADS*, 2001.

[58] A. D. Rubin, D. Geer, and M. J. Ranum. *WEB Security Sourcebook*. John Wiley and Sons, Inc., 1997.

[59] P. Selfridge. Privacy and knowledge discovery in databases. In *IEEE Expert*, April 1995.

[60] G. Smith. Modeling security-relevant data semantics. In *Proc. IEEE Symp. Research in Security and Privacy*, pages 384–391, 1990.

[61] P. Stachour and B. Thuraisingham. Design of LDV: A multilevel secure relational database management system. *IEEE Trans. Knowledge and Data Eng.*, 2(2):190–209, June 1990.

[62] L. D. Stein. *Web Security - A Step-by-Step Reference Guide*. Addison-Wesley Longman, inc., 1998.

[63] M. Stickel. Elimination of inference channels by optimal upgrading. In *Proc. of the 1994 IEEE Symposium on Research in Security and Privacy*, pages 168–174, May 1994. Oakland.

[64] A. Stoica and C. Farkas. Secure XML views. In *Proc. IFIP WG11.3 Working Conference on Database and Application Security*, 2002.

[65] T. Su and G. Ozsoyoglu. Inference in MLS database systems. *IEEE Trans. Knowledge and Data Eng.*, 3(4):474–485, December 1991.

[66] T.H.Hinke, H. S. Delugach, and A. Chandrasekhar. A fast algorithm for detecting second paths in database inference analysis. *Jour. of Computer Security*, 3(2,3):147–168, 1995.

[67] B. Thuraisingham. Security checking in relational database management systems augmented with inference engines. *Computers and Security*, 6:479–492, 1987.

[68] B. Thuraisingham. Security issues for data warehousing and data mining. In *DBSec*, 1996.

[69] T. C. Ting. Privacy and confidentiality in healthcare delivery information systems. In *Proc. of the 12th IEEE Symposium on Computer-Based Medical Systems*, 1998.

[70] G. Wiederhold, M. Bilello, and C. Donahue. Web implementation of a security mediator for medical databases. In T. Y. Lin and S. Qian, editors, *Database Security XI Status and Prospent*, pages 60–67. Chapman and Hall, 1998.

[71] G. Wiederhold, M. Bilello, V. Sarathy, and X. L. Qian. Protecting collaboration. In *Proceedings of the NISSC 1996 National Information Systems Security Conference*, pages 561–569, 1996.

[72] G. Wiederhold, M. Bilello, V. Sarathy, and X. L. Qian. A security mediator for health care information. In *Proceedings of the 1996 AMIA Conference*, pages 120–124, 1996.

[73] R. Yip and K. Levitt. Data level inference detection in database systems. In *Proc. of the 11th IEEE Computer Security Foundation Workshop*, pages 179–189, 1998. Rockport, MA.