ISSN NO. 2320-5407

Journal homepage: http://www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH

RESEARCH ARTICLE

Protection of Data Base Security via Collaborative Inference Detection

Ms. Ankita. S. Chikhale[1] , Prof. S.S. Dhande[*2]

1. Department of Computer Engineering, Sipna College of Engineering and Technology, Badnera Road, Amravati, Maharashtra, India

2. Department of Computer Science and Engineering, Sipna College of Engineering and Technology, Badnera Road, Amravati, Maharashtra, India.

Manuscript Info

Manuscript History:

Received: 10 December 2014
Final Accepted: 20 January 2015
Published Online: February 2015

Key words:
knowledge processing; database; inference; probability; protection; security; query;

Corresponding Author

Ahmed Razaq Wajid .

Abstract

Malicious users can exploit the correlation among data to infer sensitive information from a series of seemingly innocuous data accesses. Thus, we will develop an inference violation detection system to protect sensitive data content. Based on data dependency, database schema, and semantic knowledge, we have to construct a semantic inference model (SIM) that represents the possible inference channels from any attribute to the preassigned sensitive attributes. The SIM is then instantiated to a semantic inference graph (SIG) for query-time inference violation detection. For a single user case, when a user poses a query, the detection system will examine his/her past query log and calculate the probability of inferring sensitive information. The query request will be denied if the inference probability exceeds the prespecified threshold. For multiuser cases, the users may share their query answers to increase the inference probability. Therefore, we will develop a model for evaluating collaborative inference based on the query sequences of collaborators and their task-sensitive collaboration levels.

INTRODUCTION

Inference is a technique a user can employ to defeat access control mechanisms in a database system. It poses a confidentiality threat to a database system, making it difficult to control access to sensitive information. An inference detection system is needed to determine if users can use legitimately accessed data to infer sensitive information. The design of an inference detection system is a trade among soundness, completeness, accessibility of the database, and efficiency of the inference detection process. ACCESS-CONTROL mechanisms are commonly used to protect users from the divulgence of sensitive information in data sources. However, such techniques are insufficient because malicious users may access a series of innocuous information and then employ inference techniques to derive sensitive data by using that information. To address this inference problem, we will develop an inference detection system that resides at the central directory site. Because inference channels can be used to provide a scalable and systematic sound inference, we construct a semantic inference model (SIM) that represents all the possible inference channels from any attribute in the system to the set of preassigned sensitive attributes. The SIM can be constructed by linking all the related attributes, which can be derived via attribute dependency from data dependency, database schema, and semantic related knowledge. Based on the SIM, the violation detection system keeps track of a user's query history. When a new query is posed, all the channels where sensitive information can be inferred will be identified. If the probability of inferring sensitive information exceeds a prespecified threshold, then the current query request will be denied. This inference detection approach is based on the assumption that users are isolated and do not share information with one another. This assumption, however, may not be the case in a real-life situation. Most users usually work as a team, and each member can access the information independently. Afterward, the members may merge their knowledge together and jointly infer the sensitive information.

Generalizing from a single-user collaborative system to a multiuser collaborative system greatly increases the complexity of the inference detection system. This motivates to extend our research from a single user case to a multiple-user case, where users may collaborate with each other to jointly infer sensitive data

I. Literatures Review & Related Work

Database inferences have been extensively studied. Many approaches to address the inference problem were presented in [1]. Particularly, Delugach and Hinke used database schema and human-supplied domain information to detect inference problems during database design time [2], [3]. Garvey et al. developed a tool for database designers to detect and remove specific types of inference in a multilevel database system [5]. Both approaches use schema-level knowledge and do not infer knowledge at the data level. These techniques are also used during database design time and not at runtime. However, Yip and Levitt pointed out the inadequacy of schema-level inference detection, and they identify six types of inference rules from the data level that serve as deterministic inference channels [6]. In order to provide a multilevel secure database management system, an inference controller prototype was developed to handle inferences during query processing. Rule-based inference strategies were applied in this prototype to protect the security [7]. Further, since data update can affect data inference, Farkas et al. [8].proposed a mechanism that propagates update to the user history files to ensure that no query is rejected based on the outdated information. Hinke et al. developed an inference analysis tool that factors domain knowledge into the inference detection system HDC94, HDW95, DH96. They group inference relevant information into three layers: entity layer, activity layer, and the entity-activity relationship layer[20].

II. Framework for an Inference Detection System

Inference detection system consists of three modules, as shown in Fig.2 The Knowledge Acquisition module extracts data dependency knowledge, data schema knowledge, and domain semantic knowledge. Based on the database schema and data sources, we can extract data dependency between attributes within the same entity and among entities. Domain semantic knowledge can be derived by semantic links with specific constraints and rules. A SIM can be constructed based on the acquired knowledge. The SIM is a data model that combines data schema, dependency, and semantic knowledge. The model links related attributes and entities, as well as semantic knowledge needed for data inference. Therefore, SIM represents all the possible relationships among the attributes of the data sources. A semantic inference graph (SIG) can be constructed by instantiating the entities and attributes in the SIM. For a given query, the SIG provides inference channels for inferring sensitive information. Based on the inference channels derived from the SIG, the Violation Detection module combines the new query request with the request log, and it checks if the current request exceeds the prespecified threshold of information leakage. If there is collaboration according to collaboration analysis, the Violation Detection module will decide whether a current query will be answered based on the acquired knowledge among the malicious group members and their CL to the current user.

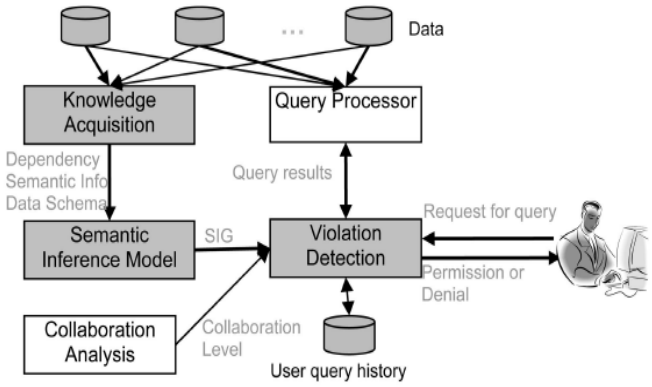

Fig 1 The framework for an inference detection system

III. Data dependency and database schema

Data dependency represents causal relationships and nondeterministic correlations between attribute values. Because of its nondeterministic nature, the dependency between two attributes. There are two types of nondeterministic data dependencies, as defined in the Probabilistic Relational Model (PRM) [13], [16]: dependency within entity and dependency between related entities.
Dependency within entity- Let A and B be two attributes in an entity E. If B depends on A, then for each instance of E, the value of attribute B depends on the value of attribute A with a probability value.
Dependency between related entities.-Let A be an attribute in entity E1 and C be an attribute in E2. E1 and E2 are related by R, which is a relation that can be derived from the database schema

IV. Design Criteria

Inference detection system is evaluated according to the following criteria:

- Soundness
- Completeness
- Accessibility of the database
- Efficiency

An inference detection system sound if it only reports inferences that exists.it is complete if it only reports all inferences that exist. The degree od accessibility of a database is measured by the amount of data that are legitimately accessible to users. The more data user can access from the database, the higher the accessibility of the database to the user. An un sound inference detection system leads to decrease in accessibility to the database. This is because queries that do not lead to any inference might be restricted, making the database less accessible to user. Similarly making more complete the detection system, lower the accessibility of the database if users are restricted access the detected inference paths. For schema-based detection system, efficiency of the detection system may not be an issue, as the detection system is run once-at the database design time. For detection system that detects inference using queries, efficiency can be an issue when the system needs to detect inference in real time.

V. Process of Inference Violation Detection System

:

In process extend our research model from single user to multiple user to inference secured data. We develop IVDS to identify the cooperation between the users and the information flows based on the cooperation. The cooperative inference for a specified query is based on the query history for the entire user with their cooperation levels. IVDS sets the threshold for every current queries, the IVDS gets the query result from the query log.

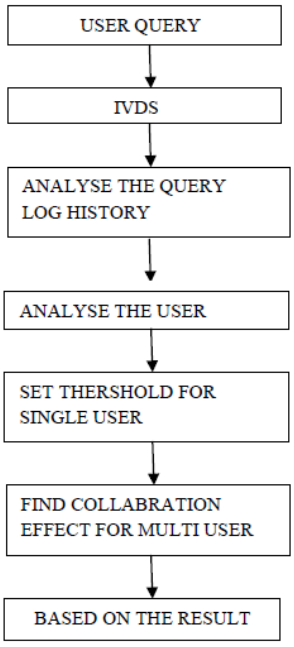

Fig 2. Overall process of IVDS

Many semantic relationships, as well as data mining rules, cannot be specified deterministically [5][17]. To remedy this shortcoming, we propose a probabilistic inference approach to treat the query time inference detection problem. The contribution of the paper consists of 1) deriving probabilistic data dependency, relational database schema, and domain-specific semantic knowledge and representing them as probabilistic inference channels in a SIM, 2) mapping the instantiated SIM into a Bayesian network for efficient and scalable inference computation, and 3) proposing an inference detection framework for multiple collaborative users. System owners have been required to put in place very rigid requirements for keeping their systems fully compliant with strict security policies and to have their systems scanned on a regular basis to guarantee that no security configuration has been altered. This strict security requirement has been accentuated by several government laws, regulations, directives, and publications.

VI. Inference Infraction Discovery for Single User

IVDS provide an integrated view of the relationships among data attributes, which can be used to detect inference violation for sensitive nodes. In such a graph, the values of the attributes are set according to the answers of the previous posted queries. Based on the list of queries and the user who posted those queries, the value of the inference will be modified accordingly. If the current query answer can infer the sensitive information greater than the pre specified threshold, then the request for accessing the query answer will be denied. The notion of imbedding policies into the database itself and altering these policies to closure every try to determine the land of the database, or to vary its shape in a way that opposes what has been accomplished and fed into the policy by the system owner. These policies can be accomplished at different graininess levels in such a way that the system owner can choose to raise coarse-grained policies to supervise and control the behavior of the database as a whole through the use of global settings, or invoke fine-grained policies that affect specific aspects or configuration settings. But the absolute core principle of our framework is the notion that the security policies, as well as all the database objects and logic that enforces them, are made an integral and inseparable part of the database that they are meant to protect.

VII. Inference Infraction Discovery for Multi User

From the single-user collaborative system to the multiuser collaborative system greatly increases the complexity and presents two challenges for building the inference detection system. First, we need to estimate the effectiveness of collaboration among users, which involves such factors as the authoritativeness of the collaborators, the communication mode among collaborators, and the honesty of the collaboration. In addition, we need to properly integrate the knowledge from collaborators on the inference channels for the inference probability computation. Database administrators or power users can alter security configurations in a way that could result in unauthorized access to and compromise of the database. An example would be that of granting privileged access to unprivileged users, or just simply misusing his privileged access[20] .Another example is one that pertains to security scans or audits of the database. Independent auditors are usually hired to perform a security scan of the database and they work with the DBA to get the database to a point where it is hardened enough to pass the scan. However, a database administrator can temporarily (or permanently) set some or all of the configuration parameters back to their original settings in order to achieve certain goals that he thinks are justified. The DBA can easily set that parameter to unlimited, change the password to the same one, and then set that parameter back to what it is supposed to be. By doing so, the DBA would have violated the rule that applies to reusing the same password over and over again. In this paper we describe a policy based approach for enforcing database configurations even to those who have privileged access. We do not advocate minimizing the role of the DBA or restricting his access. However, we do advocate that each action gets verified and approved by system owner embedded, predefined configuration policies before it is applied to the database. Unlike database security frameworks that exist today, which mostly detect imminent problems, generate an alert, and produce a report, our solution, which is an inseparable component of the database that it is meant to protect, mitigates any detected risk on its own without having to wait for human intervention.

VIII. Effectiveness of collaboration

Define CL as a metric for measuring the percentage of useful information flow from the information source to the recipient. The range of CL is from 0 to 1. CL = 0 and CL = 1 mean that none or all of the information is received by the recipient. By a series of experimental studies, we find that the CL depends on three components: the authoritativeness of the information provider A, the honesty of the collaboration H, and the fidelity of the communication channel between the provider and recipient F[9][11]. The authoritativeness of the information provider represents how accurate the information is. If a provider is knowledgeable and has high reputation in the field related with the task, then he/she can provide more accurate information. Honesty represents the honesty level of the provider and his/her willingness of releasing his/her knowledge to the recipient. For example, if user A is very knowledgeable, and A and B have a good communication channel, then both the authoritativeness and fidelity of user A are high. However, A is not willing to release his full knowledge to B. As a result, the useful information cannot reach B for inference. Further, A can deceive B with false information. Thus, we shall use the honesty measure as an indication of the honesty in collaboration [13] [18].

Fidelity measures the effectiveness of the communication between the provider and recipient. Poor mode of communication can cause information loss during the transmission, which reduces the effectiveness of the collaboration. Authoritativeness measures how accurate the provider can supply information, honesty describes the willingness of the provider to release the accurate information, and fidelity measures the percentage of information transferred to the recipient due to the limitation of the communication mode. Once we estimate these three components for a set of users on a specific task.

Conclusion and Future Research

In this paper we have implemented a technique to protect sensitive information content. Malicious users can exploit the correlation among data to infer sensitive information from a series of seemingly innocuous data accesses. This developed inference detection system can be used for any organization with very small changes as per their database.[18] Its ability to detect inference at the early stage rather than detecting after the attack is already committed. The developed Semantic Inference Model works for single user as well as for multi user environment. The developed system can be successfully deployed in any industry to deal with the threats that pose from internal users in an attempt to secure sensitive information. Further research and experiment in use of nested queries and use of multiple relations is needed.

The inference problem is a very harmful effect in securing the database. The attack may be happened along with the database architecture and the major consequences are handled by the database maintaining servers. For this we designed the IVDS (Inference Violation Detection System) which evaluates the query posted by every user and based on the analysis history of the every query (backlog) we can specify whether the IVDS answers the query or deny the query. This approach can be applied for both the single user as well as the multi users. We evaluate our approach in the real time experiments and obtain the results by giving various queries and different levels of users.

References

[1] C. Farkas and S. Jajodia, “The Inference Problem: A Survey,” SIGKDD Explorations, vol. 4, no. 2, pp. 6-11, 2002.

[2] H.S. Delugach and T.H. Hinke, “Wizard: A Database Inference Analysis and Detection System,” IEEE Trans. Knowledge and Data Eng., vol. 8, no. 1, pp. 56- 66, Feb. 1996.

[3] T.H. Hinke and H.S. Delugach, “Aerie: An Inference Modeling and Detection Approach for Databases,” Proc. Sixth Ann. IFIP WG 11.3 Working Conf. Dataand Applications Security, 1992.

[4] T.H. Hinke, H.S. Delugach, and R. Wolf, “Wolf: A Framework for Inference-Directed Data Mining,” Proc. 10th Ann. IFIP WG 11.3 Working Conf. Data andApplications Security, 1996.

[5] T.D. Garvey, T.F. Lunt, X. Quain, and M. Stickel, “Toward a Tool to Detect and Eliminate Inference Problems in the Design of Multilevel Databases,” Proc. SixthAnn. IFIP WG 11.3 Working Conf. Data and Applications Security, 1992.

[6] R.W. Yip and K.N. Levitt, “Data Level Inference Detection in Database Systems,” Proc. 11th Computer Security Foundations Workshop (CSFW ’98), [7] B.M. Thuraisingham, W. Ford, M. Collins, and J. O’Keeffe, “Design and Implementation of a Database Inference Controller,” IEEE Trans. Knowledge and Data Eng., vol. 11, no. 3, p. 271, June 1993.

[8] C. Farkas, T. Toland, and C. Eastman, "The Inference Problem and Updates in Relational Databases," Proc. 15th IFIP WG11.3 Working Conf. Database and Application Security, pp. 181-194, 2001.
[9] T.S. Toland, C. Farkas, and C. Eastman, "Dynamic Disclosure Monitor ðD2MonÞ: An Improved Query Processing Solution," Proc. Second VLDB Workshop Secure Data Management (SDM '05), 2005.
[10] Raymond W. Yip and Karl N. Levitt: Data Level Inference Detection in Database Systems.
[11] Y. Chen and W.W. Chu, "Database Security Protection via Inference Detection," Proc. Third IEEE Int'l Conf. Intelligence and Security Informatics (ISI '06), 2006
[12] M. Chavira and A. Darwiche, "Compiling Bayesian Networks with Local Structure," Proc. 19th Int'l Joint Conf. Artificial Intelligence (IJCAI '05), pp. 1306-1312, 2005.
[13] F. Cornelli, E. Damiani, S. De Capitani di Vimercati, S. Paraboschi, and P. Samarati, "Choosing Reputable Servents in a P2P Network," Proc. 11th Int'l World Wide Web Conf. (WWW '02), May 2002.
[14] A. Darwiche, "Recursive Conditioning," Artificial Intelligence, vol. 126, nos. 1 2, pp. 5-41, 2001.
[15] A. Darwiche, Class Notes for CS262A: Reasoning with Partial Beliefs. Univ. of California, Los Angeles, 2003.
[16] H. Li and M. Singhal, "Trust Management in Distributed Systems," Computer, vol. 40, no. 2, pp. 45-53, Feb. 2007.
[17] S. Marti and H. Garcia-Molina, "Taxonomy of Trust: Categorizing P2P Reputation Systems," Computer Networks, vol. 50, no. 4, pp. 472-484, 2006.
[18] R. Dechter, "Bucket Elimination: A Unifying Framework for Reasoning," Artificial Intelligence, vol. 113, pp. 41-85, 1999.
[19] Clifton, C. and Marks, D. (1996): Security and Privacy Implications of data mining
[20] Harry S. Delugach and Thomas H. Hinke: Wizard: A Database Inference Analysis and Detection System.
[21] T. H. Hinke, D. S. Delugach and R. Wolf: A framework for Inference-directed Data Mining.