ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

Inference Violation Detection and Collaboration with Inference Channels for Database Security

S.Kannan[1], Dr.G.N.K.Suresh Babu[2],

Research Scholar, Dept. of Computer Science, Bharathiar University, Coimbatore, India[1]

Asst. Professor, Dept. of Computer Applications, Bharath University, Chennai, India

Professor & Head, Dept. of MCA, GKM College of Engineering, Vandular, Chennai, India[2]

ABSTRACT: Many users can infer sensitive information from a series of data accesses. The inference violation detection system to protect sensitive data content. Based on data dependency, database schema, and semantic knowledge. The semantic inference model (SIM) that represents the possible inference channels from any attribute to the preassigned sensitive attributes. The semantic inference graph (SIG) is generating for query-time inference violation detection. The detection system is calculating the probability of inferring sensitive information for single user by using query history. The query request will be denied if the inference probability exceeds the prespecified threshold. The users may share their query answers to increase the inference probability for multiuser. The evaluation of collaborative inference based on the query sequences of collaborators and their task-sensitive collaboration levels.

KEYWORDS: SIM, SIM Detection, Inference, Task-sensitive collaboration.

I. INTRODUCTION

Access-control mechanisms are commonly used to protect users from sensitive information in data sources. This technique is insufficient for accessing large amount of information. For this problem, we develop an inference detection system for centralized system. The inference channel is providing a scalable and systematic sound inference. We construct a semantic inference model (SIM) by linking all the related attributes. The related attributes derived by attribute dependency from data dependency, database schema and semantic related knowledge. The SIM represents all the possible inference channels from any attributes in the system, The SIM provide a set of preassigned sensitive attributes. The violation detection system tracking the user query history by using SIM. A new query is posed, and then all the channels where sensitive information can be access and it are identified. If the probability of inferring sensitive information exceeds a prespecified threshold, then the current query request will be denied. This inference detection system are isolated the user and do not share information with one another. This system is not suitable for real-life, because many of the users are group together and worked as a team and access the information independently. We develop the collaborative system for the users. The users are merging their knowledge together and jointly infer the sensitive information. The collaborative system more general for single user, but this system is increasing the complexity for multiuser in inference detection system. We develop a collaborative inference system for a single-user case to a multiple-user case. The collaborators jointly infer the sensitive data. The inference violation detector as a tested to understand the characteristics of collaboration and the effect of collaborative inference. The experimental study is to learn the specific task, and the amount of information flow from one user to another depends and their relationships and task. Tracking the query history of all the users and their collaboration levels can derive collaborative inference for a specific task. Deluges and Hinke [4] used database schema and human-supplied domain information to detect inference problems during database design time. Garvey [7] developed a tool for database designers to detect and remove specific types of inference in a multilevel database system. 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 [9] 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. In order to provide a multilevel secure database management system, an inference controller prototype was developed to handle inferences during

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

query processing. Rule-based inference strategies were applied in this prototype to protect the security. Farkas [5] Proposed a mechanism that propagates update to the user history files to ensure that no query is rejected based on the outdated information. To reduce the time in examining the entire history login computation inference, Toland [6] proposed using a prior knowledge of data dependency to reduce the search space of a relation and reduce the processing time for inference. The previous work on data inference mainly focused on deterministic inference channels such as functional dependencies. The knowledge is represented as rules, and the rules are able to derive sound and complete inference, much valuable nondeterministic correlation in data is ignored. Further, many semantic relationships, as well as data mining rules, cannot be specified deterministically.

II. INFERENCE DETECTION SYSTEM

The proposed system has a probabilistic inference approach to treat the query-time inference detection problem. The inference detection system is detects the problem of 1) Deriving probabilistic data dependency, relational database schema, and domain-specific semantic knowledge and representing them as probabilistic inference channel in a SIM. 2) Mapping the instantiated SIM into a Bayesian network for efficient and scalable inference computation 3) Proposing an inference detection framework for multiple collaborative users. It consists of three modules, Knowledge acquisition module, Semantic inference graph (SIG), and Violation detection module.

A. KNOWLEDGE ACQUISITION FOR INFERENCE

The Knowledge Acquisition module extracts data dependency knowledge, data schema knowledge, and domain semantic knowledge based on the database and data source.

Data dependency knowledge

This knowledge represents relationship and Nondeterministic. The dependency between two attributes represented by conditional probabilities. The nondeterministic data dependencies as defined in the probabilistic relational model (PRM).

***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. To learn the parameter of dependency within entities from relational data, from a relational table that stores entity E, we can derive the conditional probabilities $Pi/j = Pr(B = bi/A = aj)$ via a sequential scan of the table with a counting of the occurrences of A and B and the co-occurrences of A and B.

***Dependency between related entities*:**

Let A be an attribute in entity E1 and C be an attribute in E2. R relates E1 and E2, which is a relation that can be derived from the database schema. If C depends on A, then only for related instances of E1 and E2 would the value of attribute C in E2 instances depend on the value of attribute A in related instances of E1. Such dependency only exists for related instances of entities E1 and E2.

Database Schema

The entities specify the primary key and foreign key pairs. Such pairing represents a relationship between two entities. If entity E1 has primary key pk, entity E2 has foreign key fk, and e1.pk = e2.fk.

Domain-Specific Semantic Knowledge

The outside information such as domain knowledge can also be used for inferences. Domain-specific semantic relationships among attributes and entities can supplement the knowledge of general users and help their inference. The domain-specific semantic knowledge as extra inference channels in the SIM. Semantic knowledge among

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

attributes is not defined in the database. From the large set of semantic queries, we can extract the semantic knowledge.

B. SEMANTIC INFERENCE MODEL (SIM)

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. Three types of relation links connect the related attributes: dependency link, schema link, and semantic link. The dependency link connects dependent attributes within the same entity or related entities. Consider two dependent attributes: A and B. Let A be the parent node and B be the child node. The degree of dependency from B to A can be represented by the conditional probabilities. The conditional probabilities are summarized into a conditional probability table (CPT). Counting the co-occurrence frequency of events can derive the conditional probability. The schema link connects an attribute of the primary key to the corresponding attribute of the foreign key in the related entities. The Semantic link connects attributes with a specific semantic relation. We need to compute the CPT for nodes connected by semantic links. Let T be the target node of the semantic link, PS be the source node, and P1, . . . ,Pn be the other parents of T in fig.1. The semantic inference from a source node to a target node can be evaluated as follows. If the semantic relation between the source and the target node is unknown or if the value of the source node is unknown, then the source and target nodes are independent.

Source node

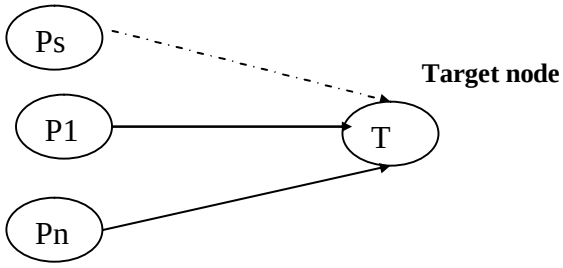

Fig.1 Target node T with Semantic link from source node Ps and dependency links from parents P1...Pn

Thus, the semantic link between them does not help inference. To represent the case of the unknown semantic relationship, we need to introduce the attribute value “unknown” to the source node and set the value of the source node to “unknown.” In this case, the source and target nodes are independent, that is, When the semantic relationship is known, the conditional probability of the target node is updated according to the semantic relationship and the value of the source node. If the value of the source node and the semantic relation are known, derived from the specific semantic relationship.

C. VIOLATION DETECTION

The Violation Detection module helps the access authorization and combines the new query request with the request log, and it checks if the current request exceeds the pre-specified threshold of information breach. In Fig.1 the schematic framework shows the Inference Detection System with collaboration level. The collaboration is 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. SIGs 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. Generalizing from the single-user collaborative system to the multi user collaborative system greatly increases the complexity and presents

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

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.

III. COLLABORATION

The combination of knowledge from collaborator on different types of inference channels. Based on the users query history, there are two different types of collaborative user pairs, ***Collaboration with nonoverlap inference channels:***

In this case, the two users pose queries on different nonoverlap inference channels. The inference probability will be computed based on their combined knowledge discounted by their collaborative level.

Collaboration with overlap inference channels:

In this case the query sets posed by the two users overlap on inference channels. Such overlap may cause the users to have inconsistent belief in the same attribute on the inference channel. Thus, we need to integrate the overlapping knowledge according to the collaborative level to compute the appropriate inference probability.

A.N-collaborators

Therefore, for any two collaborative users, we can integrate one's knowledge to the other and detect their inference toward sensitive data. When any user poses a query, the system not only checks if the query requester can infer sensitive data above the threshold with a query answer but also checks the other team members to guarantee that the query answer will not indirectly let them infer the sensitive attribute. We can iteratively generalize the above approach Chen and Chu: protection of database security via collaborative inference detection 1021. The SIM for a transportation mission planning to an n-collaborator case In general, when there are n-collaborative users in the team, the violation detection system tracks the query posed by every team member. A query should be denied if the query answer will increase the certainty of any team member inferring the sensitive data above the pre specified threshold. Since the system needs to evaluate the inference probability for every collaborator, the time required for inference evaluation increases as the number of collaborator increases. In our test bed, on a sample Bayesian network with 40 nodes, after any user in a group of collaborators poses a random query, the time for inference evaluation ranges from 15 ms for a single user to 281 ms for five collaborators when their collaboration level is equal to 1. The evaluation time almost doubles when the CL is less than 1 because the system requires extra computation to insert virtual nodes.

IV. COLLABORATION LEVEL

Information authoritativeness, honesty, and communication channel fidelity are three components of the CL metrics. In the section, we shall first conduct a set of experiments to validate the premise of the proposed metrics and then propose the integration of these three components to estimate the CL.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

Block Diagram

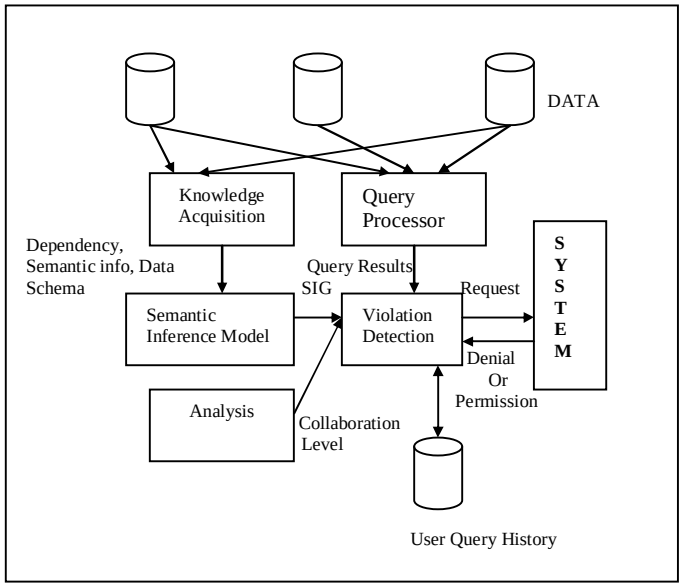

Fig.2: Framework for inference Detection System with collaboration

A. EXPERIMENTAL STUDY OF COLLABORATION LEVEL

We investigated the collaboration effectiveness under controlled authoritativeness and communication fidelity. This experiment1 was carried out similar to experiment2 (fig.2) except that it was conducted in another graduate class in the following quarter. Because of the small class size, we divided the students into three teams, each having three members. Since authoritativeness, honesty, and fidelity is user sensitive, we used the students in one of the authors' classes as test subjects. The experiment was used as homework for the class. A Web interface was developed for our inference test bed so that students could pose queries directly to the test bed and receive the answers. Before posing queries for inference, each student needed to register in the system and fill in the necessary background information, including their age, gender, major, year in school, courses taken, grade point average (GPA), skills, interests, teamwork ability, social activities, friends in the class, and so forth. The information gave us clues about the information authoritativeness and certain aspects of the fidelity of the test subjects.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

	A	H	F	CL
Exp.1 Team1:	1	1	1	0.989
Exp.1 Team2:	0.75	1	0.33	0.901
Exp.1 Team3:	0.6	1	0.17	0.8034
Exp.2 Team1:	1	0.7418	1	0.9741
Exp.2 Team2:	0.75	0	1	0.4936
Exp.2 Team3:	0.6	0.6107	1	0.8132
Regression Model: CL = 0.1449.A +0.4948.H+0.1988.F+0.0275 Residual Sum of Squares: rss = $8.124.10^{-3}$				

Table:1 Experiment 1 and Experiment 2 for 3 teams

***Estimation of Fidelity*:**

Fidelity measures the percentage of information sent by the provider that reaches the recipient side. Thus, fidelity depends on the quality of the communication channel and on the communication mode.

Estimation of honesty:

Honesty represents the willingness and truthfulness of the information release from the provider to the recipient. This is related with the evaluation of trust in peer-to-peer (P2P) which can be categorized by reputation-based or evidence-based approaches]. One approach is to use the reputation-based method, as proposed. The honesty level is recipient dependent. Therefore, for a given task, the honesty between two collaborators should also be estimated based on their closeness or friendship for a specific task. Therefore, the specific honesty of a provider to a specific recipient needs to be adjusted by the closeness between the collaborators for a given task.

B. ESTIMATING COLLABORATION LEVEL FROM A TRAINING SET

Since the CL is user and task sensitive, we propose using a regression method to study the task- and user-specific relations between the CL and its parameters. Specifically, for a group of users and a specific task, we can treat A, H, and F as the predictors and the CL as the response variable. We can then learn the coefficients of these variables from the regression model via the set of training data. As an example, let the results of collaborative inference from experiment1 and experiment 2 under a controlled environment, with selected A, H, and F values be a training set. Since the inference result obtained by a team reflects the collaboration effectiveness under the corresponding controlled environment, we can normalize the inference result (that is, the inference Result of the security attribute divided by the threshold) as the estimate of the CL. Using the six entries as the input for regression analysis, the CL can be fit by multiple regression method with residual sum of squares $8.124 .10^{-3}$ as shown in Table 1. Thus, we can estimate future CLS by substituting the parameters A, H, and F into the regression model for similar users and task.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

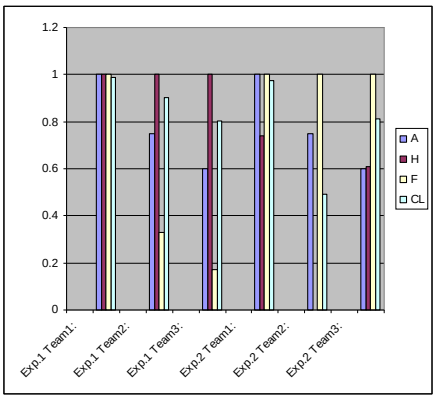

Fig: 2 Authoritativeness, honesty and fidelity from

Experiment 1 and Experiment 2

V. CONCLUSION

We have developed a detection system that prevents single users from inferring sensitive information by a series of innocuous queries. The knowledge acquisition done the dependency and semantic information. Based on the data dependency, the database schema and the semantic knowledge, we constructed a semantic inference model (SIM) that links all the related attributes and thus, represent all possible inference channels from any attributes to the set of pre-assigned sensitive attributes. The SIM is then instantiated by specific instances and reduced to a semantic inference graph (SIG) for inference violation detection to control query access. To reduce computation complexity for inference, the SIG can be mapped into a Bayesian network, where the nodes represent the attributes and links represent the relationships among attributes. Available Bayesian network tool can then be used for evaluating the inference probability along the inference channels. When a user poses a query, the detection system will examine by using query log and calculate the probability of inferring sensitive information from answering this posed query. The query request will be denied if it can infer sensitive information with probability exceeding the pre-specified threshold. We are currently extending the detection system to multiple collaborative users based on query history of all the users as well as their social relations. Then the SIM and Violation Detection are used for protection of database for the single user with query processing. The Collaboration level we done channel fidelity for CL metrics. We are currently extending the detection system for multiple collaborative users that is based on their query histories as well as their social relations.

REFERENCES

[1] 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.
[2] C. Duma, N. Shahmehri, and G. Caronni, "Dynamic Trust Metrics for Peer-to-Peer Systems," Proc. 16th Int'l Workshop Database and Expert Systems Applications (DEXA '05), pp. 776-781, 2005.
[3] R. Dechter, "Bucket Elimination: A Unifying Framework for Probabilistic Inference," Proc. 12th Conf. Uncertainty in Artificial Intelligence (UAI '96), pp. 211-219, 1996.
[4] 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.
[5] C. Farkas and S. Jajodia, "The Inference Problem: A Survey," SIGKDD Explorations, vol. 4, no. 2, pp. 6-11, 2002.
[6] T.S. Toland, C. Farkas, and C. Eastman, "Dynamic Disclosure Monitor: An Improved Query Processing Solution," Proc. Second VLDB Workshop Secure Data Management (SDM '05), 2005.
[7] 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. Sixth Ann. IFIP WG 11.3 Working Conf. Data and Applications Security, 1992.
[8] 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 and Applications Security, 1996.
[9] R.W. Yip and K.N. Levitt, "Data Level Inference Detection in Database Systems," Proc. 11th Computer Security Foundations Workshop (CSFW '98), 1998.