ISSN:0975-9646

Protection of Data Base Security via Collaborative Inference Detection

Ajay Sharma, Alok Kumar Shukla, Dharmendra Kumar

Dept of Information Technology
Institute of Technology & Management, Gida, Gorakhpur

Abstract:-Throughout companies, government departments, and doctors' offices, database systems are used. This particular system stores and retrieves sensitive information such as social security numbers, financial statements, and highly classified data. Organizations with sensitive data in their hands need to be secured using different security techniques and policies. In order to secure the data on a computer, they need to implement techniques like access control, auditing, authentication, encryption, etc; however, malicious users are still breaking into companies' data. Needless to say, companies are not implementing poor security techniques but hackers are just getting smarter and smarter. This is where IT employees need to find new techniques or enhance previous ones. Thus, we develop an inference violation detection system to protect sensitive data content. Based on data dependency, database schema and semantic knowledge, we constructed a semantic inference model (SIM) that represents the possible inference channels from any attribute to the pre-assigned sensitive attributes. This model can work for single user as well as for multi user environment. 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 pre-specified threshold. For multi-user cases, the users may share their query answers to increase the inference probability. Therefore, we develop a model to evaluate collaborative inference based on the query sequences of collaborators and their task-sensitive collaboration levels.

Keywords: Database, Security, and Detection Inferences. Semantic Inference Model(SIM).

1. INTRODUCTION:-

In the present time data base security is the main problem. Modern database systems allow multiple users access to data. When users are not to be allowed accesses to every item of data in the database, an access control system is needed. Access control mechanisms are commonly used to protect users from the sensitive information in data sources. An access control system based on two component. the access control policy and the access control mechanism. The access control policy describe the allow are disallowed for each user in data base. The access control mechanism enforces the policy. Each user accesses the database system using queries. The allowed queries are processed by the database system, and the results are returned to the user. The disallowed queries can be handled in various ways. For example, the user may simply be noticed that the query violates the access control policy and is not processed by the database system, or the database system intentionally returns incorrect responses to the user in order to protect the data. The invalid accesses might also be recorded for further investigation.

2. WORK:-

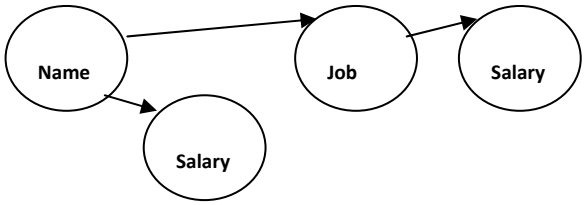
In this section there are three module used to protect the data base based on the inference channel.

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 semantic inference model can be constructed based on the acquired knowledge.

The Semantic Inference Model (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, Violation Detection module combines the new query request with the request log, and it checks to see if the current request exceeds the pre-specified threshold of information leakage. If there is collaboration according to collaboration analysis, the Violation Detection module will decide whether to answer a current query based on the acquired knowledge among the malicious group members and their collaboration level to the current user.

A multilevel database system is a database system that enforces the access policy. Early work on inference detection in multilevel database systems employed a graph to represent functional dependencies among attributes in the database schema. Each node in the graph corresponds to an attribute in the database schema. An edge from node A to node B in the graph indicates that the attribute corresponding to node A functionally determines the attribute corresponding to node B. An inference path is detected when there are two or more paths found in the graph that connect one node to another, and the paths are labeled at different classification levels. The classification level of a path is the least upper bound of the classification levels of the attributes corresponding to the nodes on the path. For example, consider the NSJ and JS tables. We can construct a graph as shown in Figure 2.1 to represent the database schema.



Since users may pose queries and acquire knowledge from different sources, we need to con-struct a semantic inference model for the detection system to track user inference intention. The semantic inference model requires the system to acquire knowledge from data dependency, data-base schema and domain-specific semantic knowledge.

3. SEMANTIC INFERENCE MODEL:-

The Semantic Inference Model (SIM) represents dependent and semantic relationships among attributes of all the entities in the information system. The related attributes (nodes) are connected by three types of relation links: dependency link, schema link and semantic link. 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 pi|j =Pr(B=bi|A=aj). The conditional probabilities of the child node given all of its parents are summarized into a conditional probability table (CPT) that is attached to the child node. For instance, the CPT of the node "TAKEOFF_LANDING_CAPACITY" of the SIM The conditional probabilities in the CPT can be derived from the database content [FGK99, GFK01]. For example, the conditional probability Pr(B=bi|A=aj) can be derived by counting the co-occurrence frequency of the event B=bi and A=aj and dividing it by the occurrence frequency of the event A=aj.

Schema link connects an attribute of the primary key to the corresponding attribute of the foreign key in the related entities. For example, APORT_NM is the primary key in AIRPORTS and foreign key of RUNWAYS. Therefore, we connect these two attributes via schema link.

Semantic link connects attributes with a specific semantic relation. To evaluate the inference introduced by semantic links, 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. 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 node are independent. 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 node are independent, i.e., Pr(T=ti|P1=v1, … Pn=vn, PS=unknown) = Pr(T=ti|P1=v1, … Pn=vn). 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, then Pr(T=ti| P1= v1, … Pn= vn, PS=sj) can be derived from the specific semantic relationship.

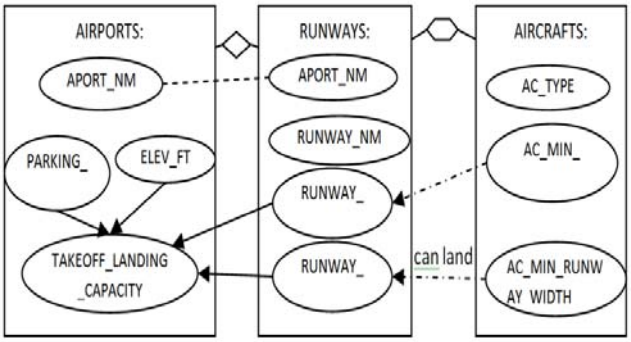


Fig. 1 Semantic Inference Model example for Airports, Runways and Aircraft

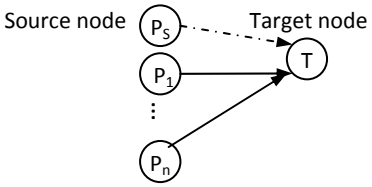


Fig. 1.1 Target node T with semantic link from source node P_S and dependency links from parents P_1, …, P_n.

For example, the semantic relation "can land" between Runway and Aircraft implies that the length of Runway is greater than the minimum required Aircraft landing distance. So the source node is aircraft_min_land_dist, and the target node is runway_length. Both attributes can take three values: "short," "medium" and "long." First, we add value "unknown" to source node aircraft_min_land_dist and set it as a default value. Then we update the conditional probabilities of the target node to reflect the semantic relationship. Here, we assume that runway_length has an equal probability of being short, medium or long. When the source node is set to "unknown," the runway_length is independent of aircraft_min_land_dist; when the source node has a known value, the semantic relation "can land" requires runway_length is greater than or equal to aircraft_min_land_dist.

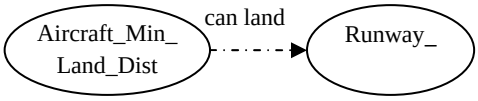


Fig. 1.2. The semantic link "can land" between "Aircraft_Min_Land_Dist" and "Runway_Length"

4. SEMANTIC INFERENCE GRAPH:-

To perform inference at the instance level, we instantiate the SIM with specific entity instances and generate a semantic inference graph (SIG). Each node in the SIG represents an attribute for a specific instance. Related attributes are then connected via instance-level dependency links, instance-level schema links and instance-level semantic links. The attribute nodes in SIG have the same CPT as in SIM because they are just instantiated versions of the attributes in entities. As a result, the SIG represents all the instance-level inference channels.

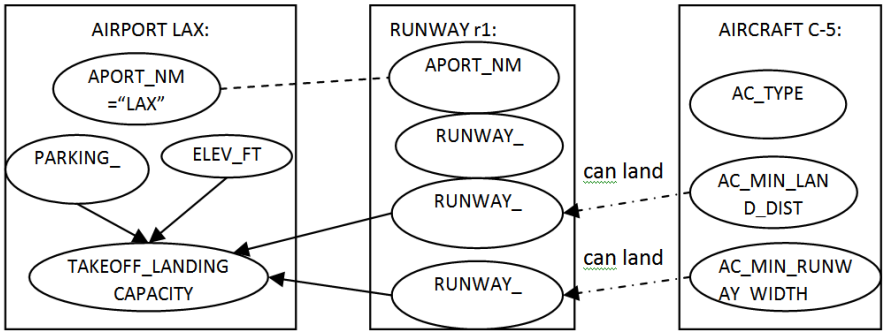


Fig.1.3 The Semantic Inference Graph for airport instance (LAX), with runway r1 and aircraft C-5.

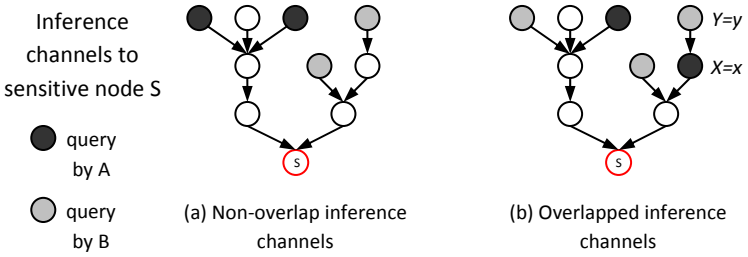


Fig.1.4. Types of collaborative user pairs in the social network posing query sequence on the inference

5. COMBINING KNOWLEDGE FROM COLLABORATORS:-

In this section, we study the combination of knowledge from collaborators on different types of inference channels. Based on the users' query history, there are two different types of collaborative user pairs: Collaboration with non-overlap inference channels and Collaboration with overlap inference channels

Collaboration with non-overlap inference channels: In this case, the two users pose queries on different non-overlap inference channels. The inference probability will be computed based on their combined knowledge discounted by their collaborative level.

For example, two collaborators A and B ask queries on non-overlap inference channels. In addition, the collaboration level from user A to user B is given by CLAB, and the collaboration level from B to A is CLBA. Therefore, to compute the inference probability to security attribute of user A, the query answers acquired by B (QB) can be combined with his/her own query answers (QA), but discounted by the collaborative level CLBA. On the other hand, to derive the inference probability of user B, A's query answers (QA) are discounted by collaboration level CLAB and then combined with QB. Because QA and QB are from independent non-overlap inference channels, their inferences to sensitive node S are independent and can be directly combined. Thus the inference probability for the sensitive node can be computed based on the user's knowledge from his past queries combined with his collaborator's query answers discounted by their respective collaborative level.

6. CONCLUSION:-

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. Usually database consists of User Profile of the Community websites and Employee database. These sensitive data should not be leaked if happened so the trust of the sites becomes lower. 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. Y. Chen and W. W. Chu, (2008), "Protection of Database Security via Collaborative Inference Detection", IEEE Transactions on Knowledge And Data Engineering, vol. 20, no. 8
2. A Data Level Database Inference Detection System By "Raymond Wai-Man Yip" B.Sc. (Chinese University of Hong Kong) 1988 .
3. Protection of Database Security via Collaborative Inference Detection By "Yu Chen and Wesley W. Chu".
4. A Novel Approach to Conquer Inference Problems and Inference Risks in Secured Database , Vol 3 (5), 1731-1735 , ISSN:2229-6093.
5. Chapple, Mike. "Database Security Issues: Inference." www. About.com/database security".