DETECTION OF MALICIOUS TRANSACTION IN DATABASE USING LOG MINING APPROACH

Ms. Apashabi Chandkhan Pathan
Department of Computer Engineering
D. Y. Patil College of Engineering, Akurdi
Pune, India
apashabi.pathan@gmail.com

Mrs. Madhuri A. Potey
Department of Computer Engineering
D. Y. Patil College of Engineering, Akurdi
Pune, India
mapotey@gmail.com

Abstract*— Data mining is the process of finding correlations in the relational databases. There are different techniques for identifying malicious database transactions. Many existing approaches which profile is SQL query structures and database user activities to detect intrusion, the log mining approach is the automatic discovery for identifying anomalous database transactions. Mining of the Data is very helpful to end users for extracting useful business information from large database. Multi-level and multi-dimensional data mining are employed to discover data item dependency rules, data sequence rules, domain dependency rules, and domain sequence rules from the database log containing legitimate transactions. Database transactions that do not comply with the rules are identified as malicious transactions. The log mining approach can achieve desired true and false positive rates when the confidence and support are set up appropriately. The implemented system incrementally maintain the data dependency rule sets and optimize the performance of the intrusion detection process.

Keywords: Data Mining, Database security, Intrusion Detection.

I. INTRODUCTION

Recently Database Management Systems have been developed which gives the guarantee of high assurance and security. The component which is very important for Intrusion Detection (ID) techniques to database security solution. These techniques are able to detect anomalous behavior of users and applications. There are many approaches used to protect the networks and data from attackers. To make data more secure by using Intrusion Detection Systems [IDS] on critical systems. The IDS's used for early detection of attacks and make the recovery of lost or damaged data simpler. Many researchers are working on increasing the intrusion detection efficiency and accuracy for database management system.

A. Intrusion Detection in Database Systems

To develop architectures of IDS, mechanisms and algorithms for a DBMS equipped with activity monitoring, intrusion detection and response capabilities. Within this broad context, the research issues that are as follows:

- Creating profiles that succinctly represent user/application-behavior interacting with a DBMS
- Developing efficient algorithms for detection of anomalous database user/ application behavior
- Developing strategies for responding to intrusions in context of a DBMS

Creating a system architecture for database intrusion detection and intrusion response as an integral component of a DBMS, and a prototype implementation of the same in PostgreSQL relational database.

II. LITERATURE REVIEW

Different approaches have been proposed by researchers to address the problem of identifying malicious database transactions. One approach by author William G, J. Halfond, Alessandro Orso, and Panagiotis Manolios [5] is to detect anomalous SQL query structures. The model proposed by Bandhakavi [3] is dynamically mines the programmer-intended query structure on any input and its detect the attacks by comparing the structure of the actual query issued. Security adds an extra defensive layer to the web application to detect and filter attacks such as SQL injection using a signature-based approach. The query structure intended by a programmer deduce at run time. Such type of techniques that promise a real scalable automatic solution to dynamically detect as well as prevent SQL injection attacks. These approaches mainly target at SQL injection attacks launched from web applications. Detecting malicious database transaction patterns was proposed by Ashish Kamra, Elisa Bertino in [2] to mine database logs to form user profiles and identify anomalous transactions in databases with a role based access control mechanisms. It is able to identify the behaviors of intruders that differ from the normal behavior of a role in a database. Elisa Bertino and Kamra was illustrated the model [6] that can use to identify intruders in databases but it has no roles associated with each user. To form concise pro-files clustering techniques representing normal user behaviors for identifying suspicious database activities. When the transactions that do not comply with rules these are identified as malicious transactions. Srivastava offered [7] a weighted sequence mining approach for detecting database attacks. However these models only consider sequential data dependencies and data dependencies at a single granularity level, i.e. attribute dependencies. Mining of the Data is very helpful to end users for extracting useful business information

978-1-4799-2102-7/14 $31.00 © 2014 IEEE
DOI 10.1109/ICESC.2014.50

from large database [11] The Query processing refers to the activities involved in extracting data from a database. Silberschatz, Korth, Sudarshan[12] proved the activities include translation of queries in high-level database languages into expressions that can be used at the physical level of the file system, a variety of query-optimizing transformations and actual evaluation of queries. Query optimization is a part of query compilation process [13] which consist of four step like parsing, simplification, cost-based optimization and plan preparation. The detection of intrusion is a passive approach [8] to database security and monitors information systems. Alarms raises when security violations are detected.

A. Analysis of Different Methodologies for Database Intrusion Detection

Table 1: Analysis based on different methodologies for database intrusion detection

Methods	Approach	Based on	Limitations
RBAC	Using Positive Tainting to Counter SQL Injection Attacks.	ID is based on mining database and its stored in log files.	Maintaining or updating the profiles for the large number of users is not a trivial task.
CANDID	To deduce at run-time the query structure intended by a programmer.	symbolic query computed on a program run.	Techniques that promise a real scalable automatic solution to the dynamically detect and prevent SQL injection attacks.
Weighted Sequence Mining.	Finding data dependencies RDBMS.	Mining Algorithm that mines user profiles based on the pattern of submitted queries.	Incapability in treating database attributes at different levels of sensitivity in particular.
Using Positive Tainting & SyntaxAware Evaluation	protecting existing Web applications against SQL injection.	Positive tainting and the concept of syntax-aware evaluation	The efficiency of the technique.

An approach for dynamic detection and prevention of SQLIAs is proposed by William G.J. Halfond, Alessandro Orso, and Panagiotis Manolios [5] Using Positive Tainting and Syntax-Aware Evaluation to Counter SQL Injection Attacks approach works by identifying trusted strings in an application and allowing only these trusted strings to be used to create certain parts of an SQL query, such as keywords or operators approach works by identifying trusted strings in an application and allowing only these trusted strings to be used to create certain parts of an SQL query, such as keywords or operators. Sensitivity of an attribute signifies how important the attribute is, for tracking against malicious modifications. This data mining techniques is proposed by the author Abhinav Srivastava, Shamik Sural and A.K. Majumdar [7]. Database Intrusion Detection using Weighted Sequence Mining mines dependency among attributes in a database.

The detection of malicious database transaction patterns was proposed by Bertino in [2] to mine database logs to form user profiles that can model normal behaviors and identify anomalous transactions in databases with role based access control mechanisms. The component which is very important for any strong security solution is represented by Intrusion Detection (ID) techniques. These techniques is able to detect anomalous behavior of users and applications.

Different approaches have been proposed by researchers to address the problem of identifying malicious database transactions. However these models only consider sequential data dependencies and data dependencies at a single granularity level, i.e., attribute dependencies. Data dependency rules generated reflect semantic relationships among data items and are less likely to change than SQL query structures and normal user behaviors. Therefore, they are ideal for profiling data correlations for identifying malicious database activities. Kamra illustrated [6] an enhanced model that can also identify intruders in databases where there are no roles associated with each user. Srivastava offered [7] a weighted sequence mining approach for detecting database attacks. Different granularity to represent the SQL queries appearing in the database log files and able to extract useful information from the log files regarding the access patterns of the queries. Evimaria Tezi, Ashis Karma, Elisa Bertino proposed in [4] when role information is available in the log records, use it for training a classifier that is then used as the basic component for our anomaly detection mechanism.

II. PROPOSED SYSTEM

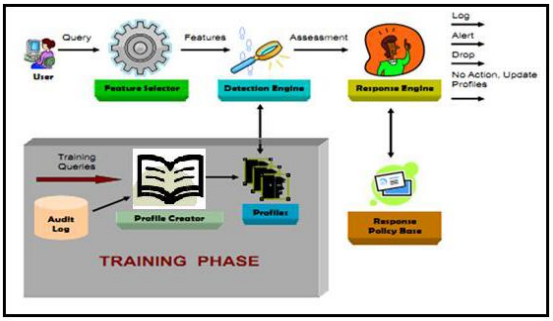

Fig 1: Proposed System Architecture

Proposed system roughly divides into following module.

- Module I: Training phase
- Module II: Detection phase

A. Intrusion Detection in Database Systems

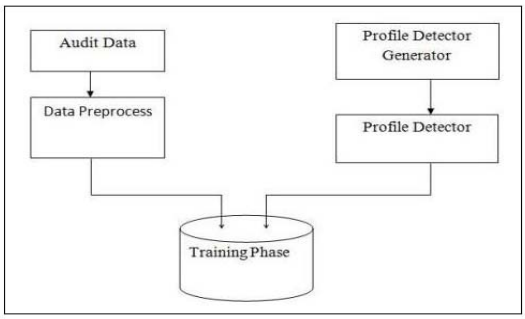

Fig 2: Training Phase

Fig. 2 shows the training phase for proposed system. To capture the behavior of database objects, this monitor and audit the system operation. This auditing system helps to collect necessary data for building database profiles. To be more accurate, whatever technique the profiler utilizes to build the profiles, data gathered by auditing system provides necessary input for it.

B. Intrusion Detection in Database Systems

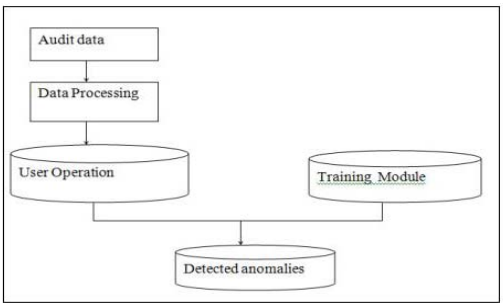

Fig 3: Detection Phase

Fig. 3 shows Detection system for Database. Depending on the suspicious level or sensitivity of intrusion, detection mechanism can contribute to access control system to deny access and prevent the intruder from causing malicious transaction. The log file consists the information about the committed transactions those are executed in the secure environment by the authorized users. Transactions profile are considered as authorized profiles and stored at the system, after that these authorized transactions profile are used at the detection phase.

Set of transaction of trained queries
S={s1, s2, ...sn}
Set of transaction of tested queries
T={t1, t2,.....tn}
For 1<i<n
Probability= P(si, ti)
If P(si, ti) < c
Then malicious transaction occurs.

III. Experiment Setup

Our experiments are based on an evaluation framework that we developed and has been used by us and other researchers in previous work [1]. The framework provides a database intrusion detection that consists two different categories database logs were generated, legitimate training transactions and malicious transactions. Different parameters are used to generate database log i.e., number of operations in a transaction, number of domains, number of data items in each domain, and number of transactions, and a large set of test inputs containing both legitimate transaction and malicious transaction. It consists of five database applications that accept user input via SQL and use it to build queries to an underlying database. Five applications are commercial applications i.e. Contact, Dataagent, Distribute, Stockdata, userdata developed by us. There are two sets of inputs: Training phase, which consists of legitimate transaction for the database application, and Detection Phase , which consists of legitimate transaction and malicious transaction.

IV. Result and Analysis

Fig. 4 shows the relationship between the support threshold of rules and true/false positive rate. The confidence threshold for the experiment generating this figure is set at 60%. By comparing Fig. 5 with Figure 4, it is observed that the true positive rate is more sensitive to the change of support, whereas the false positive rate is not really susceptible to the change. When the support changes from 10% to 30%, the enhance true positive rate changes from 50% to 60% whereas, and the enhance true positive rate drops sharply from 49% to 44% for the support value 40 to 50%. From the results of these experiments, it can be seen that the desired confidence threshold range is 60% and the ideal support threshold range is [10, 50] for our experiment settings. Relation between the support of rules and true/false positive rate Although malicious data read operations can cause information to be leaked to unauthorized users, illegitimate modification of data can cause greater damage. .

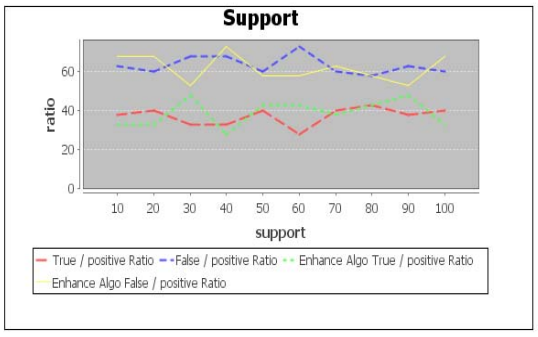

Fig. 4: Relation between the support of rules and true/false positive rate[S=10% C=60%]

To test how effective our approach is for identifying malicious data modifications, we conducted experiments

based on the mean number of write operations in a training transaction and observed false/true positive rates

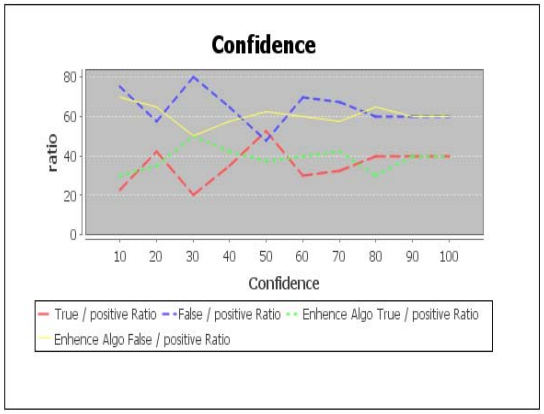

Fig. 5: Relation between the Confidence of rules and true/false positive rate[S=10% C=60%]

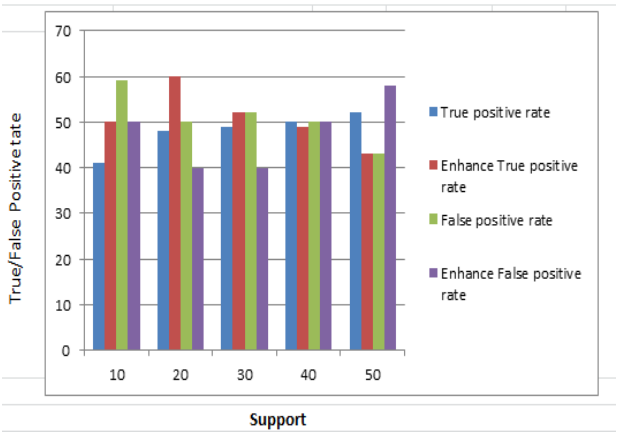

Fig. 6: Relation between the support of rules and true/false positive rate.

Fig. 6 presents the relationship between the confidence threshold of rules and true/false positive rate. The support threshold for the experiment illustrated by this figure is set at 10%. The true positive rates are generated by checking malicious transactions against data dependency rules generated from the legitimate training transactions. The false positive rates are derived by examining the log containing legitimate training transactions against rules generated from the same log. It can be seen that the false positive rate is sensitive to the change of confidence, whereas the true positive rate is not very susceptible to the change. When confidence changes from 50% to 100%, the false positive rate changes from 100% to 0% and the true positive rate only fluctuates between 100% and 85%.

V. Conclusion

An effective log mining approach for detecting malicious database transactions is presented Multi-level and multi-dimensional data mining are employed to discover data item dependency rules, data sequence rules, domain dependency rules, and domain sequence rules from the database log containing legitimate transactions. Database transactions that do not comply with the rules are identified as malicious transactions. The true positive rates are generated by checking malicious transactions against data dependency rules generated from the legitimate training transactions. The false positive rates are derived by examining the log containing legitimate training transactions against rules generated from the same log. The proposed work is to incrementally maintain the data dependency rule sets and optimize the performance of the intrusion detection process. The proposed work has the limitation regarding the processing power and the data storage issues to handle huge amount of information. Data dependency rule capabilities are limited. It must overcome many research challenges before it can make the rule for identify the malicious transaction.

References

[1] Yi Ru, Alina Campan, James Walden, Irina Vorobyeva, Justin Shelton, “An Effective Log Mining Approach for Database Intrusion Detection”, IEEE 2010.

[2] Ashish Kamra, Elisa Bertino, “Guy mechanisms for Database Intrusion Detection and Response”, Proceedings of the Second SIGMOD PhD Workshop on Innovative Database Research, ACM 2008

[3] Sruthi Bandhakavi, Prithvi Bisht, P. Madhusudan, V.N. Venkatakrishnan, “CANDID: Preventing SQL Injection Attacks using Dynamic Candidate Evaluations”, IEEE Nov 2007.

[4] Ashis Karma, Evimaria Tezi, Elisa Bertino, “Database Detecting Anomalous Access Patterns in Relational Databases”, IEEE 2007.

[5] William G.J. Halfond, Alessandro Orso, and PanagiotisManolios, “Using Positive Tainting and Syntax-Aware Evaluation to Counter SQL Injection Attacks”, IEEE Nov 2006.

[6] Elisa Bertino, Ashish Kamra, Evimaria Terzi, Athena Vakali, “Intrusion Detection in RBAC-Administered Databases”, Center for Education and Research in Information Assurance and Security, Purdue University, West Lafayette, CERIAS 2005.

[7] Srivastava A, Sural S, and Majumdar A. K, “Database Intrusion Detection Using Weighted Sequence Mining”, Journal of Computers, IEEE 2006.

[8] E.Ke savulu Reddy, Member IAENG, V. Naveen Reddy, P.Govinda Rajulu, “A Study of Intrusion De-tection in Data Mining”, Proceedings of the World Congress on Engineering 2011 Vol III WCE, IEEE July 2011.

[9] William A. R. Weiss, “An Introduction to Set Theory”, October 2, 2008.

[10] Morten Blomhoj, Thomas Hojgaard Jensen, “Developing Mathematical Modelling Competence: Conceptual Clarification and Educational Planning”, July 2003.

[11] Alex Berson, Stephen J. Smith, “Data Warehousing, Data Mining, OLAP”, Tata McGraw-Hill Edition 2004, page no-333.

[12] Silberschatz Korth, Sudarshan, “Database System Concepts”, Fourth Edition, Page no. 493-495.

[13] Silberschatz Korth, Sudarshan, “Database System Concepts”, Fifth Edition, Page no. 1069.