Intrusion Detection and Forensic Analysis on Database using Log Mining Approach

Agrata Jain, Sneha Saswade, Yogesh Phalke, Chaitanya Gholap

Abstract— The demand for secure storage of data has become necessity of our time. Financial records, medical records and legal information are all in need of secure storage.In the dynamic world economies and the era of globalization, data outsourcing is unavoidable. Security is the leading concern in data outsourcing environment, since data is under the custody of third party web servers.In current scenario, third party can access and view data even though they are not authorized to do so and allowing the employee of the organization to update the database. This may lead to serious data tampering, data theft or data leakages causing severe business loss to data owner. An important element of any strong security solution is represented by intrusion detection (ID) systems, which detects anomalous behavior by applications and users. In our project, we have proposed a novel solution to detect database intrusion using Log Mining technique. Log files are unalterable files at runtime, automatically created by Web servers. Main use of log file is to keep trace of transactions performed on any web applications. We consider purchaser database at server-side and compare this with the transactions traced from the log files, with the help of which database tampering can be determined for any indifference found. Finally, with the help of forensic analysis algorithm, we will figure out who did the tampering. Hence the system administrator and the data owner will have a secured system with our model.

Index Terms— Forensic Analysis, Database Intrusion, Log Mining, Security, Tampering, Web Application, Web Server.

1 INTRODUCTION

An intrusion detection system (IDS) is a device or software application that monitors network or system activities for malicious activities or policy violations and produces reports to a Management Station. Some system may attempt to stop an intrusion attempt but this is neither required nor expected of a monitoring system. Intrusion detection system can also be system specific.

There are three main types of Intrusion Detection System viz Network intrusion detection system (NIDS), Host-based intrusion detection system (HIDS) and Stack-based intrusion detection system (SIDS). Our proposed system is a type of Host-based intrusion detection system. It aims to provide detection of tampering made in databases, which contains client transactions for any particular web application, on the server side by unauthorized users. The transactions made by customers through the web application are stored in a database on the server which necessarily should remain unaltered and consistent.

The project aim is to develop an Intrusion Detection System (IDS) for tampered data in web services on the server-side for Databases using Log Mining Algorithm. The proposed system aims to simplify the task of intrusion detection and simultaneously provide better speed up solution to find perfect intrusions. This project implements an E-Commerce website using Java Servlet Pages. The

client transactions database will be maintained using MySQL 1.2.1 beta. For maintaining the consistency of purchaser database, the project also provides a feature of implicitly mailing a forensic analysis report as per the time interval set by the owner, using the concept of triggers. In an effort to avoid further damage, it provides database restoration of purchaser.

2 BACKGROUND MOTIVATIONS

We referred several IEEE papers during the initial stages of our project. 'An Effective Log Mining Approach for Database Intrusion Detection'[1] published with IEEE paper standards describes that due to sudden proliferation of networked applications, database centered applications are facing a rapidly growing number of threats. Malicious outsiders launch attacks to access or corrupt data by stealing access control credentials or exploiting application vulnerabilities. On the Server side, server admin may sabotage databases by abusing priviliges. Although various intrusion prevention and detection mechanisms are employed to protect against outsider and insider attacks, they are not very effective in detecting attacks targeted on the database at the server-side. The above mentioned paper exhibits a novel scheme for identifying malicious transaction patterns to detect attacks launched either by outsiders or insiders on server database.

The IEEE paper titled 'Development of Host Based Intrusion Detection System for Log Files'[2] published with IEEE paper standards describes host based intrusion describes host based intrusion detection by using pattern matching technique on log files. The system will recognize two types of attack and its pattern. If an attack is unknown pattern, the ystem needs to keep that pattern in the database for the future assessment. Then, if an attack knows pattern, the system will match that pattern in their database and alert the host user about the attack or intrusion.

Another IEEE paper titled 'The Tiled Bitmap Forensic Analysis Algorithm' [2] published with IEEE paper standards describe the forensic analysis algorithm to detect database tampering by using cryptographically strong hash functions. The applied forensic analysis algorithms helps in determining when the intrusion occurred, what data fields got tampered, and perhaps ultimately who did it and why was it done. The tiled bitmap algorithm is more efficient than prior forensic analysis algorithms. It introduces the notion of a candidate set and provides a complete characterization of the candidate set and its cardinality.

Taking these researches into consideration we have developed our intrusion detection mechanism giving more stress to the comparision of server-side database with the application server logs, rather than identifying any kind of malicious transaction patterns. A major advantage of our approach is that it does not require any modification and enhancement to the storage system software.

3 PROJECT SCOPE

The aim of our project is to develop an intrusion detection system on server side for a web application hosted on a server. Intrusion in database which will consist of all transactions taking place in the web application would be detected on demand by the owner with the help of application logs. The project will automate the process of forensic analysis on tampered data. The data owner and system administrator can have secured system with our model. In case of intrusion, the date, time, intruded fields in the database can be detected with the help of the system. The user credentials of the server side database can be used to detect who was responsible for the tampering. Also, the tampered data would be successfully restored.

The project would be explained through

developing a locally hosted online shopping application. The protection of the database from insider attacks in the scope of our project.

4 SYSTEM DESCRIPTION

The basic partitioning of the system is done in two parts namely, client side and server side. Both the client side and the server side are efficiently operable on Windows operating system platform. The server side database is maintained by MySQL 5.0. It contains various tables for storing admin log-in credentials, purchaser information as well as sales information. The UI of the system is majorly developed using Java Servlet Pages. The UI provides a simple and user friendly interface for any type of user. The coding of the project is done using Java Platform which provides various utilities available in Java packages.

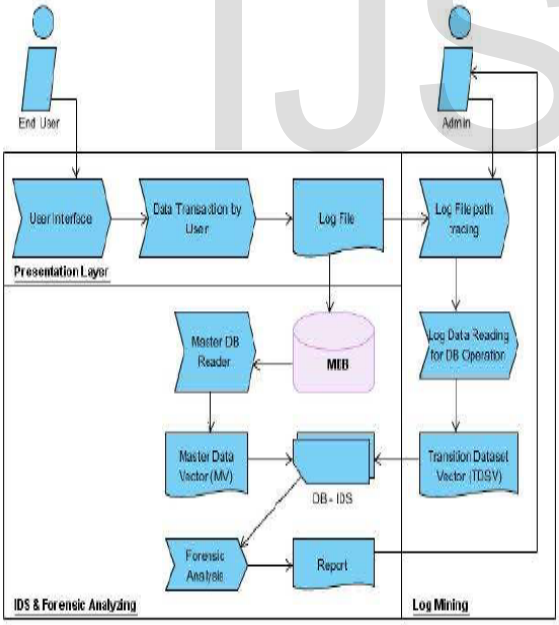

Fig. System Architecture

4.1 IMPLEMENTATION DETAILS

Presentation Layer Module: This module basically consists of the GUI. Here the purchaser can make the purchase of the desired product, submit its information, make online payments. Each purchase made by the consumer would be stored in the form of a transaction, which would be stored in the log file and also in the master database at the server-side. On the other hand, the owner here can login and check the administrator account, change password or the email id.

Log-mining: On choosing for the option of database intrusion detection by the owner, necessary transactions made by purchasers would be traced and copied into a Transaction Dataset Vector (TDSV).

IDS and Forensic Analysis Module: For database intrusion detection, comparision of MDSV obtained from master database and TDSV would be done and sent for further forensic analysis. Final reports of detected intrusions would be mailed to owner.

The web-application user (customer) has following facilities:

1. The user does not require any authentication to access the web application.
2. End user can make a hassle free purchase through the web application.
3. The UI provided is simple and easy to use.
4. Customer gets an order confirmation message via mail for the products purchased.

The system owner has following facilities:

1. The application is runnable on any computer having Windows operating system.
2. Authentication would be done each time the owner will login.
3. Owner can check for intrusion according to his/her will and convenience.
4. In case of intrusion, original database gets restored automatically.
5. The complete forensic report of data tampering is sent to the owner of the database through mail.

5 THEORETICAL FRAMEWORK

The basic idea of the project lies in mining the log files generated by the application server. The application server such as – Tomcat- contains a logs folder where all the log files generated during application processing reside. These log files are dynamic in nature as well as they can't be edited. The last log file of this folder contains detailed application logs along with exceptions that occur during the run time of a particular application. Further processing with the help of these log files needs these log files to be copied to a text file, so that they can be converted to a static and editable format. Whenever the owner of the application wishes to opt for intrusion detection the database will be checked for any tampering and if any difference is found, the detailed report will be mailed to the owner. The report will contain the information about what data have been tampered, when it had been tampered and who was the server administrator during that period.

In addition to this, after sending the report to the owner, the intruded fields of the database will also be restored with their original values so as to avoid further damage of the owner. The proposed system is feasible for any E-Commerce and E-banking application.

6 ALGORITHM

a. Start
b. Scan Server logs folder for latest log file.
c. Get the path of log file.
d. Read the contents of the log file
e. Copy the content to a new text file log-copy.txt with a specified path
f. Read the content and store it in a string object.
g. Get log string object from step f.
h. Check the transaction trace in log object and put it in log data i.e.TDSV.
i. Store the database content in master object vector i.e MV.
j. Get TDSV size as n
k. For i=0 to n-1
l. If valid cahges=true go to next step else go to step n.
m. Update changes into mater database.
n. Search for TDSV(i) in MV with tid
o. If transaction id(tid) found then
p. Compare TDSV(i) with MV object
q. If equal then no intrusion
r. Else alert intrusion and go for forensic analysis.
s. Compare TDSV (i) with databasedata with each data field to get what has been tampered.
t. Replace tampered field with original datafields from logdata i.e. TDSV.
u. Fire a query using DMV to get when (date and time) the intrusion was done and who (by the user credentials) did the intrusion.
v. Prepare a report in text file and mail to the owner.
w. End.

7 SYSTEM FEATURES

1. Hassle Free Transaction: The system allows customer to purchase a product with ease. Customer receives an order confirmation message as soon as his/her transaction is completed successfully.
2. Enhanced security: The main aim of the system is to provide security to the database of the owner. This security is provided by finding out the illegal changes done on

the database.

3. Forensic analysis: The system finds out 'who' did the tampering with the help of forensic anslysis to catch the culprit.
4. Implicit and Explicit options: Owner of the database gets both implicit and explicit option of detecting intrusion. In case of implicit mode, system automatically runs the intrusion engine after a given time interval to check if any database tampering has been performed. In case of explicit mode, the owner clicks on a specific button to start the intrusion engine.
5. Forensic report to owner sent via mail: The complete forensic report of data tampering is sent to the owner of the database through mail.
6. Database restoration: System restores the original database and does not reflect the illegal changes done by the culprit. This makes the original database safe and secure.

CONCLUSION

In this paper, log mining approach for detecting malicious database transactions is presented. The system is designed to identify attacks launched by malicious transactions submitted to relational database systems. We formally introduced a series of concepts related to profiling legitimate database access patterns for identifying malicious transactions. As part of our future work, we plan to study how we can optimize the performance of the intrusion detection process.

ACKNOWLEDGMENT

We take this opportunity to thank our project guide and head of the department Prof. Vaishali S. Nandedkar for her valuable guidance and for providing all the necessary facilities, which were indispensable in the completion of the project. We are also thankful to all the staff members of the Department of Information Technology of Padmabhooshan Vasantdada Patil Institute of Technology, Pune for their valuable time, support, comments, suggestions and persuasion. Finally we offer our great thanks to the institute for providing the required facilities, Internet access and important books.

REFERENCES

[1] An Effective Log Mining Approach for Database Intrusion Detection* Yi Ru, Alina Campan, James Walden, Irina Vorobyeva, Justin Shelton 978-1-4244-6588-0/10/$25.00 ©2010 IEEE.

[2] Development of Host Based Intrusion Detection System for Log Files. Firkhan Ali Bin Hamid Ali, Yee Yong Len 978-1-4577-1549-5/11/$26.00 ©2011 IEEE.

[3] The Tiled Bitmap Forensic Analysis Algorithm, Kyriacos E. Pavlou and Richard T. Snodgrass, Senior Member, IEEE transaction on knowledge and data engineering Vol 22, pp no. 590-601, April 2010.

[4] Web Log File Data Clustering Using K-Means and Decision Tree, Supinder Singh, Sukhpreet Kaur.

[5] Forensic Analysis of Database Tampering Kyriacos Pavlou, Richard T. Snodgrass.