Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 6, Issue, 04, pp.6209-6212, April, 2014

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

ISSN: 0975-833X

RESEARCH ARTICLE

INTRUSION DETECTION SYSTEM IN DATABASE USING LOG MINING APPROACH

*[1]Kaveri Sable, [1]Jyoti Mache, [2]ManjushriSumbe and [1]SonaliShirsath

[1]SIT, Lonavala, Maharashtra, India
[2]SIT, Pune, Maharashtra, India

ARTICLE INFO

Article History:
Received 26th January, 2013
Received in revised form
24th February, 2014
Accepted 10th March, 2014
Published online 23rd April, 2014

Key words:
Database Security,
Intrusion Detection,
Data mining.

ABSTRACT

A considerable effort has been recently devoted to the development of database management systems (DBMS) which guarantee high assurance security and privacy. Organizations spend a significant amount of resources securing their servers and network perimeters. An important component of any strong security solution is represented by intrusion detection systems, able to detect anomalous behavior by applications and users. To date, however, there have been very few intrusion detection mechanisms specifically tailored to database systems. We have proposed a novel solution called Log mining approach. The approach we propose to intrusion detection is based on mining database traces stored in log files. In this Paper, we present a new technique for identifying malicious database transactions, are ideal for profiling data correlations for identifying malicious database activities. The result of the mining process is used to form user profiles that can model normal behavior and identify intruders.

Copyright © 2014 Kaveri Sable et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Data mining (the analysis step of the "Knowledge Discovery in Databases" process), a field at the intersection of computer science and statistics, is the process that attempts to discover patterns in large data sets. It utilizes methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Aside from the raw analysis step, it involves database and data management aspects, data preprocessing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating. With the 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. The OW ASP Top 10 Project identified SQL injection as the second most serious vulnerability in web applications. Also, employees in most organizations are given much more access than what they actually need to do their jobs. Disgruntled insiders may sabotage databases by abusing privileges or masquerading as other users. 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 at databases.

Corresponding author: Kaveri Sable
SIT, Lonavala, Maharashtra, India.

The actual data mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection) and dependencies (association rule mining). This usually involves using database techniques such as spatial indexes. These patterns can then be seen as a kind of summary of the input data, and may be used in further analysis or, for example, in machine learning and predictive analytics. For example, the data mining step might identify multiple groups in the data, which can then be used to obtain more accurate prediction results by a decision support system. Neither the data collection, data preparation, nor result interpretation and reporting arepart of the data mining step, but do belong to the overall KDD process as additional steps.

The related terms data dredging, data fishing, and data snooping refer to the use of data mining methods to sample parts of a larger population data set that are (or may be) too small for reliable statistical inferences to be made about the validity of any patterns discovered. These methods can, however, be used in creating new hypotheses to test against the larger data populations. Log files are generated by system processes to record activities for subsequent analysis. They can be useful tools for troubleshooting system problems and also to check for inappropriate activity. The UNIX releases are preconfigured to record certain information in log files, but configuration settings are available to increase the amount of information recorded. A server log is a log file (or several files) automatically created and maintained by a server of activity

performed by it. Log files can be very useful resources for security incident investigations. They can also be essential for prosecution of criminal activity. For these reasons log files should be periodically backed up to separate media, and precautions need to be taken to prevent tampering with the log files. It is expected that an unauthorized intruder into a computing system will attempt to remove any trace of their activities from the system log files.

RELATED WORK

Data mining is widely used to identify interesting, potentially useful and understandable patterns from a large data repository [Abhinav Srivastava *et al.*, With many organizations focusing on web based on-line transactions; the threat of security violations has also increased. Since database stores valuable information of an application, its security has started getting attention. An intrusion detection system (IDS) is used to detect potential violations in database security. In every database, some of the attributes are considered more sensitive to malicious modifications compared to others. A. Shrivastvaet. al. proposed an algorithm for finding dependencies among important data items in a relational database management system [Abhinav Srivastava *et al*. Any transaction that does not follow these dependency rules are identified as malicious. Ahors have shown that this algorithm can detect modification of sensitive attributes quite accurately. Authors also suggest an extension to the Entity-Relationship (E-R) model to syntactically capture the sensitivity levels of the attributes.

The Internet and computer networks are exposed to an increasing number of security threats. With new types of attacks appearing continually, developing flexible and adaptive security oriented approaches is a severe challenge [MueenUddin *et al.*, 2009]. Intrusions detection systems (IDSs) are systems that try to detect attacks as they occur or after the attacks took place. IDSs collect network traffic information from some point on the network or computer system and then use this information to secure the network. In this context, signature-based network intrusion detection techniques are a valuable technology to protect target systems and networks against malicious activities. Signature-based detection is the most extensively used threat detection technique for (IDSs). One of the foremost challenges for signature-based IDSs is how to keep up with large volume of incoming traffic when each packet needs to be compared with every signature in the database. When an IDS cannot keep up with the traffic flood, all it can do is to drop packets, therefore, may miss potential attacks. Uddin *et al.* proposed a new model called Signature-based Multi-Layer IDS using mobile agents, which can detect imminent threats with extremely high success rate by dynamically and automatically creating and using small and efficient multiple databases, and at the same time, provide mechanism to update these small signature databases at regular intervals.

Sequential pattern mining deals with data represented as sequences (a sequence contains sorted sets of items). Compared to the association rule problem, a study of such data provides “inter-transaction” analysis (Agrawal and Srikant,). Applications for sequential pattern extraction are numerous and the problem definition has been slightly modified in different ways. Associated to elegant solutions, these problems can match with real-life time stamped data (when association rules fail) and provide useful results. For a large database of customer transactions. Each transaction consisting of items purchased by a customer in a visit. Rakesh Agrawal *et al.* present an efficient algorithm that generates all significant association rules between items in the database [RakeshAgrawal *et al*. An algorithm incorporates buffer management and novel estimation and pruning techniques. Authors also present results of applying this algorithm to sales data obtained from a large retailing company, which shows the effectiveness of the algorithm

PROPOSED WORK

The proposed database intrusion detection system consists of log mining mechanism and an intrusion detection mechanism. In this we are using vector concept for detection of intrusion. Vector is array list with extended properties which follows dynamic and automatic addition of data at run time. So it reduces the computations. Inthis work we are mining log file for comparison purpose to detect intrusion. Initially, system copies the contents from log file into temporary file as no one can perform operations on log file directly. Then with original database the comparison is carried out. And intrusion is detected if any and report is generated which gives field where the intrusion is occurred and also gives date & time.

So the system follows

1. To allow Users to perform transactions.
2. Provides the facility to read log file and collect all transaction data.
3. Provides a facility to collect all infected data from Master DB.
4. Provides a facility to detect tamper detection.
5. Provides a facility to perform forensic analysis on tampered data

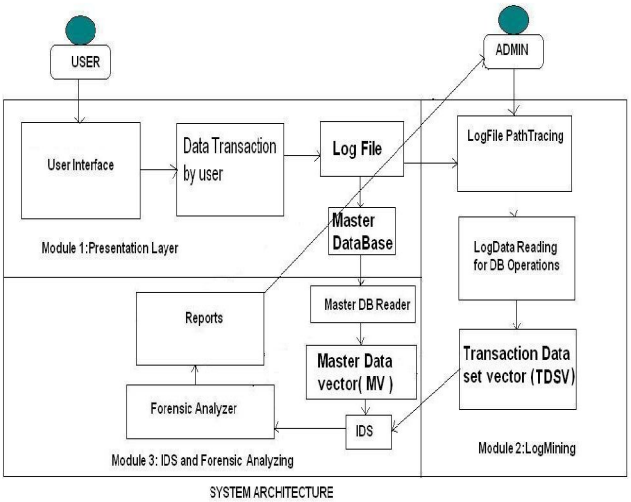

Figure 1. System Architecture

SYSTEM ARCHITECTURE

The working flow of the intrusion detection on data mining using log mining approach. The user can be a administrator, an auditor or a third party user (employee). The user will authenticate giving his legal identity. Depending on the type of user the access to the data will be granted. If the user is an administrator, the access to the original database is granted. If the user is an auditor or an employee, a copy of database is accessed not the original one. All the modifications and updating by the auditor and employee are done only on the copy of database.

For every change or access to the data, whether on original or on the copy, a default log file is created, which will have all the details of what changes where done, at what time and by whom. Only the administrator has the access to the log files. Using the Log Mining algorithm, the administrator can extract the specific log files, which will help to compare the original database and the modified copy of database. If the changes made are legal or the same as expected then there's no intrusion in the database. If the changes made are illegal then intrusion is detected and action can betaken. The effective log mining approach in the above case helps in extracting the specific log files required for a particular set of changes done in a specific period of time. This can be done by using a time stamp with the log mining approach. The administrator won't have to waste time in cross checking all the log files and then comparing it. Which will be very time consuming and inefficient. As the access to the log files are only with the administrator, the log files can't be harmed and hence illegal intrusion can be easily traced out. Our approach presented uses time signatures in discovering database intrusions.

ALGORITHMS

Algorithm for checking database intrusion

1. Start
2. Scan server logs folder for latest log file updated
3. Get the path of the log file
4. Read the content of log file
5. Copy the content into a new logcopy.txt file with specified path
6. Read the content and store in a string object
7. Get log string object from step 6
8. Check the Transaction trace in log object and put it in log data
9. Store the database table content in database data
10. Get log data size as n
11. For i=0 to n-1 do
12. If valid changes=true go to next step else go to step 14
13. Update purchaseinfo table using updateinfo table
14. Search for log_data(i) in database_data with tidoms#
15. If tidoms# found, then
16. Compare log_data(i) and database_data object
17. If equal then no intrusion
18. Else alert intrusion
19. Compare log_data(i) and database_data with each data field as to get what has been tampered
20. Replace tampered fields with original data fields from log data
21. Invoke Investigator to find out date, time and database admin credentials at the instance of intrusion
22. Prepare a detailed report and mail to the owner
23. End

Conclusions and Future Work

Log mining approach for detecting malicious database transactions is presented. As part of our future work, we plan to study how we can optimize the performance of the Intrusion detection process. Determining frequent item set is one of the most important fields of data mining. It is well known that the way candidates are defined has great effect on running time and memory need, and this is the reason for the large number of algorithms. It is also clear that the applied data structure also influences efficiency parameters. However, the same algorithm that uses a certain data structure has a wide variety of implementation. Thus, 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. Our experiments show that the proposed method can achieve desired true and false positive rates when the confidence and support are set up appropriately. As part of our future work, we plan to study how we can incrementally maintain the data dependency rule sets and optimize the performance of the intrusion detection process.

REFERENCES

A Data Mining Approach for Database Intrusion Detection Yi Hu , Brajendra Panda Computer Science and Computer Engineering Department University of Arkansas Fayetteville, AR 72701, USA +1-(479)-973-0849 yhu@uark.edu 2004.

An Adaptive Intrusion Detection System using a Data Mining Approach, Sujaa Rani Mohan, E.K. Park, Yijie Han*University of Missouri, Kansas City,{srmhv7 | ekpark | hanyij}@umkc.edu*

An Effective Log Mining Approach for Database Intrusion Detection*,Yi Ru, AlinaCampan, James Walden, Irina Vorobyeva, Justin Shelton, Computer Science Department Northern Kentucky University, Highland Heights, KY 41099, USA, {huy l, campana 1 , waldenj l, vorobyevai l, sheltonj5}@nku.edu 2010.

Database Intrusion Detection using Weighted Sequence Mining, Abhinav Srivastava1, Shamik Sural1 and A.K. Majumdar2 1 School of Information Technology 2 Department of Computer Science &Engineering ,Indian Institute of Technology, Kharagpur, 721302, India, Email: {abhinavs@sit, shamik@sit, akmj@cse}.iitkgp.ernet.in

Enhancing Intrusion detection Using Layered Approach with PCA B. Ben Sujitha and V. KavithaPonjesly College of Engineering, University college of Engineering, Nagercoil, Tamilnadu, India 2013.

Kamra A, Bertino, E., and Lebanon, G.: Mechanisms for DatabaseIntrusion Detection and Response. In the Proceedings of the 2nd SIGMOD PhD Workshop on Innovative Database Research (2008)

Liu, A, Yuan, Y., Wijesekera, D., and Stavrou, A: SQLProb: A ProxybasedArchitecture towards Preventing SQL Injection Attacks. InProceedings of the 2009 ACM Symposium on Applied Computing (2009)

Log Analysis-Based Intrusion Detection via Unsupervised Learning, *Pingchuan Ma*, Master of Science School of Informatics, University of Edinburgh 2003.

MueenUddin, Azizah Abdul Rehman, NaeemUddin, JamshedMemon, RaedAlsaqour, and SuhailKazi, *"Signature-based Multi-Layer Distributed Intrusion Detection System". International Journal of Network Security*, Vol.15, No.1, PP.79-87, Jan. 2009.

RakeshAgrawal Tomasz Imielinski_ Arun Swami *"Mining Association Rules between Sets of Items in Large Databases"* IBMAlmaden Research Center 650 Harry Road, San Jose, CA 95120.

Srivastava, A, Sural S., and Majumdar, AK.: Database Intrusion Detection Using Weighted Sequence Mining, *Journal of Computers*, vol.1, no. 4 (2006)

****