International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 – 6435(Online), Volume 1, Issue 1, Oct 2013)

Audit Trail Based on Process Mining and Log

Kundan Kumar Mishra[1], Rahul Kaul[2]
[1]CSE, BMCT, Indore
[2]CSE, SDBCT, Indore

***Abstract*-- An audit trail is an evidence of all procedures that take place in a system and across a network; it provides an outline of user/system events so that safety measures events can be associated to the actions of a specie individual or system element. Audit trails can be inspected for the existence or nonexistence of confident patterns. Audit trails can be used for measuring security issues, access patterns and also used for managing the performance of any employee in software industries. In this research paper, we will propose a process mining based technique to evaluate audit trails for security measures. This paper is inspiration of the work based on alpha algorithm to support security efforts at various levels ranging from low-level intrusion detection to high-level fraud prevention.**

***Keywords*-- Audit Trails, Process mining, Pattern Discovery, Log, Data Mining**

I. Introduction

The basic idea for analysis of business processes is to extract knowledge from event logs recorded by an information system. Process mining aims at improving this by providing techniques and tools for discovering process, control, data, organizational, and social structures from event logs. The goal of process mining is to extract information about processes from transaction logs.

The three different perspectives of Process mining are:

(1) The process perspective,
(2) The organizational perspective and
(3) The case perspective.

The process perspective focuses on the control flow i.e., the ordering of activities. The goal of mining this perspective is to find a good characterization of all possible paths. Process mining is useful for at least two reasons. First of all, it could be used as a tool to find out how people and/or procedures really work. Consider for example processes supported by an ERP system like SAP (e.g., aprocurement process). Such a system logs all transactions but in many cases does not enforce a specific way of working. In such an environment, process mining could be used to gain insight in the actual process. Second, process mining could be used for Delta analysis, i.e., comparing the actual process with some predefined process. Note that in many situations there is a descriptive or prescriptive process model, such a model specifies how people and organizations are assumed/expected to work.

By comparing the descriptive or prescriptive process model with the discovered model, discrepancies between both can be detected and used to improve the process.

Clearly, both aspects (discovery and delta analysis) are relevant for computer security and auditing. For example, in an approach for intrusion detection is presented. This method inspect audit trails and uses fixed-length patterns to distinguish self (i.e., normal process execution) from other (i.e. a potential security violation). Thus, we explore the concept of process mining and one algorithm in particular (the α-algorithm) in the context of security.

II. Problem Definition

In the domain of knowledge extraction and data mining, data analysis is performed. in the medical, engineering, academics and other domain take advantages from the data mining and knowledge processing. in all the above given fields a large amount of work found in last some years but too few effort is made for the domain of processing and its uses.

In process mining for extracting knowledge the running processes are evaluated and system access patterns and logs are analyzed. Additionally running processes in any system indicates the user behavior analyses, and the security.

Data consumption and using the process mining previously used as security infrastructure, but the performance of this architecture is suspicious due to variable length of processes in different machines. Required to enhance data model and the techniques by which system analyse the process for security effectively and provide the high accurate results.

III. Solution Domain

In the process mining the various methods and various sources of data is available such as to evaluate the user behaviour we can use the audit log, system log and other but to make an effective security architecture required to evaluate the current running process list in the different machines.

To provide effective systems the sub components are most helpful for developing a suitable data model. In the previously developed techniques the authors are used decision trees and self-designed algorithms for model training and evaluation. The decision trees are working with a fix length instances and generate the data model.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 – 6435(Online), Volume 1, Issue 1, Oct 2013)

To resolve the problem of variable length of processes sequence is not properly defined by the decision trees, thus required to find such kind of classification utility by which we analysis the incoming data effectively.

Following is the list of assumptions and dependencies:

- It can run on all windows based operating systems only.
- The end user who is going use this application should have dot net framework installed in his machine; otherwise he can't able to run this application on his machine.
- Preparation of a centralize log
- Mining logs to discover user working pattern.

IV. Terminology Used

Audit trails: A record showing who has accessed a computer system and what operations he or she has performed during a given period of time. Audit trails are useful both for maintaining security and for recovering lost transactions. Most accounting systems and database management systems include an audit trail component. In addition, there are separate audit trail software products that enable network administrators to monitor use of network resources.

Process mining: Process mining is a process management technique that allows for the analysis of business processes based on event logs. The basic idea is to extract knowledge from event logs recorded by an information system. Process mining aims at improving this by providing techniques and tools for discovering process, control, data, organizational, and social structures from event logs.

Data Mining: Mining is used to detect fraud for its effectiveness. Data mining is a process that uses a variety of data analysis tools to discover patterns and relationships in data that may be used to make a valid prediction. The six basic steps of data mining process are defining the problem, preparing data, exploring data, building models, exploring and validating models, deploying and updating models.

Data mining is a field at the intersection of computer science and statistics, is the process that attempts to discover patterns in large sets. 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.

V. Client Server Architecture

In this system mainly focus on client's activity and measuring their performance at different time and detect which activity is not according to the prediction, for the proposed system. Log management system there are five types of logs like Database log, Audit log, System log, Device log and Application log. In the proposed system use three type of log. They are Audit log, System log, Application log and activity of each log in the system.

System Log: The system log file contains events that are logged by the operating system components. These events are often predetermined by the operating system itself. System log files may contain information about device changes, device drivers, system changes, events, operations and more.

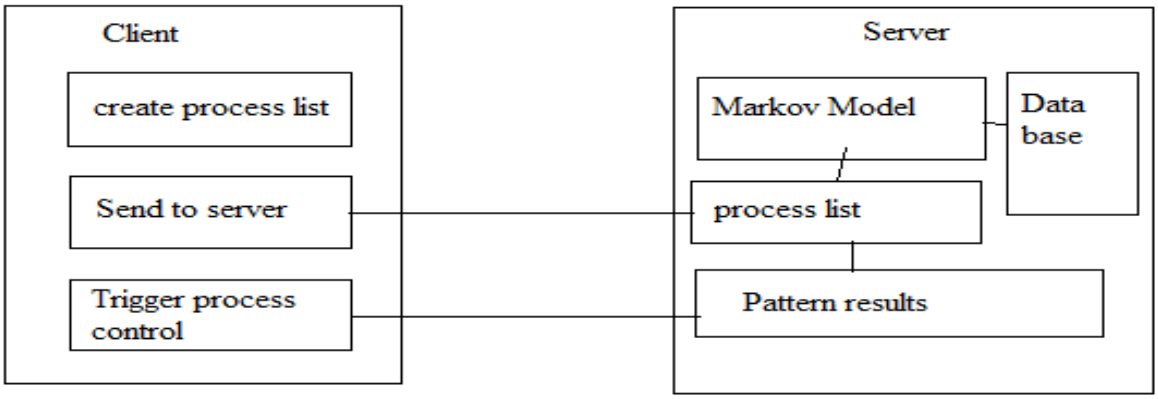
Application log: the application log file contains events that are logged by the applications used on a computer system. Events that are written to the application log are determined by the developers of the software program, not the operating system.

Viewing report server events:

1) Report server
2) Report manager
3) Scheduling and delivery process

Audit Log: It is security relevant chronological record, set of records, source and destination of records showing who has an accessed a computer system and what operation he or she performed during a given period of time. Audit logs/trails are useful both for maintaining security for recovering lost transaction such as financial transaction , scientific research and health care data transaction or communication by individual people, system or other activity.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 – 6435(Online), Volume 1, Issue 1, Oct 2013)



VI. Technology & Algorithm Used

C 4.5

C 4.5 decision tree algorithm

INPUT: Experimental data set D which is showed by discrete value attributes.

OUTPUT: A decision tree T which is created by giving experimental dataset.

Create the node N;

If instance is belong to the same class

Then return node N as the leaf node and marked with CLASS C;

IF attribute List is null, THEN

Return the node N as the leaf node and signed with the most common CLASS;

Selecting the attribute with highest information gain in the attribute List, and signing the test_attribute;

Signing the node N as the test_attribute;

FOR the known value of each test_attribute to divide the samples;

Generating a new branch which is fit for the test_attribute= ai from node N;

Suppose that Ci is the set of test_attribute = ai in the samples;

IF Ci is null THEN

Adding a leaf node and signed with the most common CLASS;

ELSE we will add a leaf node return by the Generate_decision_tree.

SLIQ: SLIQ is a decision tree classifier that can handle both numeric and categorical attributes. SLIQ use pre-sorting techniques in the tree growth phase to reduce the cost of evaluating numeric attributes.

This sorting procedure is integrated with a breadth-first tree growing strategy to enable SLIQ to classify disk-resident database .SLIQ is also use a new tree pruning algorithm based on the minimum description length principle . This algorithm is inexpensive and result in compact and accurate tree. The combination of these techniques enables SLIQ to scale for large data sets and classify data sets with a large no. of classes and attributes.

VII. Background Work

In this era, mostly the work is done on Linux/Unix OS and for windows system there is a lake of such kind of system, thus required to build a system for windows operating system to mine process for recognizing the pattern of the user. As classification algorithm and research paper are concerned in the comparative study are very less in this work area. So in our proposed system to sort out this type of problem using SLIQ algorithm and C4.5 algorithm for the use of data classification and data pruning. So I use SLIQ algorithm and C4.5 algorithm and their comparisons for data classification and data pruning. SLIQ stands for supervised learning in Quest, where quest is a data mining project at IBM Alma den Research Center. It is a novel technique that improves learning time for classifier without loss in accuracy. At the same time this technique allows classification to be performed on large disk resident training data. SLIQ exhibits the same accuracy characteristics but executes faster and produces small trees, however, SLIQ impose no restrictions on the amount of training data or the no of attributes in the examples. SLIQ is also use a new tree pruning algorithm based on the minimum description length principle. This algorithm is inexpensive and result in compact and accurate tree. The combination of these techniques enables SLIQ to scale for large data sets and classify data sets with a large no. of classes and attributes.

Performance is measured in five parameters:

a) Accuracy
b) Error Rate
c) Memory used
d) Built Time
e) Search Time

VIII. CONCLUSION

A new improved α algorithm is used by which we recognize a particular user who is login right now on the given system. This is done by mining log files created by the system. This log contains information related to the user access pattern. Using data mining approach. We will mine the log file and extract the user. To mine the log file will improve α algorithm to get better accuracy.

REFERENCES

[1] S., Forrest; A.S., Perelson; L., Allen; R., Cherukuri; "Self-Nonself-Discrimination in a Computer. Symposiumon Research in Security and Privacy", IEEE Computer SocietyPress, Los Alamitos, California, 1994 pages 202–212.

[2] A., Wespi; M.,Dacier; H.,Debar; "Intrusion Detection Using Variable-LengthAudit Trail Patterns. In H. Debar, L. Me, and S.F. Wu, editors, RecentAdvances in Intrusion Detection" Springer-Verlag, Berlin, 2000, pages 110–129.

[3] A,Yavuz; P, Ning; M, Reiter; "Efficient,Compromise Resilient and Append-only CryptographicSchemes for Secure Audit Logging", Financial Cryptography and Data Security, Springer 2012

[4] "Implementation of A Algorithm for Process MiningIn Software Industry", Certified International Journal of Engineering and Innovative Technology (IJEIT) June 2012, ISSN: 2277-3754 ISO 9001:2008 Volume 1, Issue 6,

[5] Process Mining and Security: DetectingAnomalous Process Executions and CheckingProcess Conformance,The final version will be published inElectronic Notes in Theoretical Computer ScienceURL:www.elsevier.nl/locate/entcs

[6] "Creating User-Relationship-Graph in Use of FlowNetand Log Files for Computer and NetworkAccountability and Forensics",The 2010 Military Communications Conference - Unclassified Program - Cyber Security and Network Management, 2010 IEEE 978-1-4244-8179-8/10/$26.00