ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 1, Issue 6, June 2012

Implementation of A Algorithm for Process Mining In Software Industry

Pallavi Ramteke, Dr. Shikha Vishwkarma

Abstract- One approach to protected systems is from first to last the examination of audit trails. An audit trail is a evidence of all procedures that take place in a system and across a network, it provides a 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. In this project I will propose a process mining technique to evaluate audit trails for security measures. This project 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: **Process Mining, Security, Audit Trails, Pattern Discovery, Data Mining.**

I. INTRODUCTION

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. The goal of process mining is to extract information about processes from transaction logs. We assume that it is possible to record events such that

i) Each event refers to an activity (i.e., a well-denned step in the process)

ii) Each event refers to a case (i.e., a process instance)

iii) Each event can have a performer also referred to as originator (the actor executing or initiating the activity),

iv) Events have a timestamp and are totally ordered.

Some event logs contain more information on the case itself, i.e., data elements referring to properties of the case. For example, the case handling system flow logs every medication of some data element. We distinguish three different perspectives:

i) The process perspective

ii) The organizational perspective and

iii) 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 and a good characterization of all possible paths.

II.PROBLEM DOMAIN

Due to the security measures of data there are a need of such kind of system by which we know which user is online. According to the pattern of work of the system. This type of pattern is derived using maintaining log for individuals. we found in study of previous works in such domain maximum work found for Linux and UNIX operating systems. 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. Performance is measured in five parameters-securities, error rate, memory used, built time, search time. In sliq technique that improves learning time for the classifier without loss in accuracy. At the same time these techniques allows classification to be performed on large disk-resident training data. Sliq imposed no restrictions on the amount of training data or on of attributes.

III.SOLUTION DOMAIN

To resolve the above described problem I will propose the solution in the following steps:

i) Preparation of a centralize log

ii) Mining logs to discover user working pattern

iii) System is designed for windows operating system.

IV.SYSTEM ARCHITECTURE

In the system architecture there are server is connected to the N number of clients and each client has performed different activity for a particular time. 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.

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.

System Log: The system log file contains events that are often predetermined by operating system itself. The system log directive display profit use of the syslog mechanism and instead redirect all logging output to the specified filename. The filename argument should contain an absolute path and should not be to a file in a

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 1, Issue 6, June 2012

non existent directory, in a world-writeable directory or be a symbolic link use of this directive overrides any facility set by the syslog facility directives. System log files may contains information about device change, device derivers, system change, events operations.

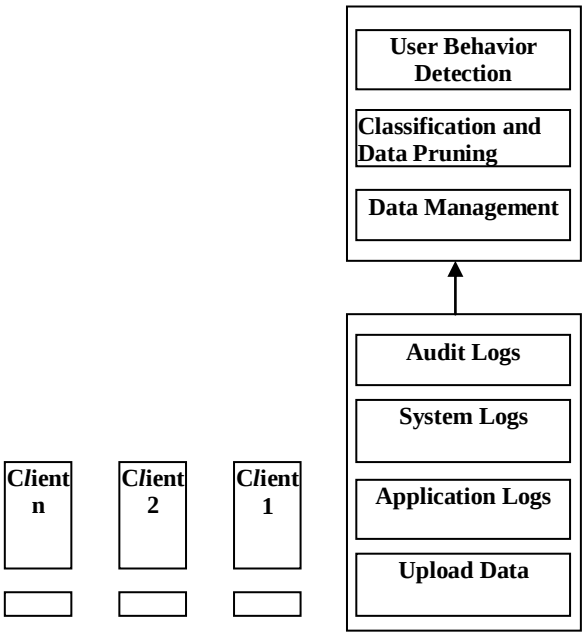

Fig 1. System Architecture

Application Log: Reporting services write event message to the windows application log. You can use the message information written to the application log to find out about events that are generated by the report server application running on the local system.

Viewing report server events:

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

V.BACKGROUND WORK

Classification is an important data mining problem and can be described as follows. The input data also called the training set consist of multiple examples (Records), each having multiple attributes or features. In the classification algorithm and research papers examine there are comparative study is very less. 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 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.

Performance is measured in five parameters:

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

VI.CONCLUSION

I am proposing a new improved α algorithm 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. I will mine the log file and extract the user. To mine the log file I will improve α algorithm to get better accuracy.

REFRENCES

[1] W.M.P. van der Aalst. The Application of Petri Nets toWorkow Management. The Journal of Circuits, Systems and Computers, 8(1):21{66, 1998.

[2] W.M.P. Vander Aalst and B.F. van Dongen. Discovering Workflow Performance Models from Timed Logs. In Y. Han, S. Tai, and D. Wikarski, editors, International Conference on Engineering and Deployment of Cooperative Information Systems (EDCIS 2002), volume 2480 of Lecture Notes in Computer Science, pages 45{63. Springer-Verlag, Berlin, 2002.

[3] W.M.P. van der Aalst and K.M. van Hee. Work ow Management: Models, Methods, and Systems. MIT press, Cambridge, MA, 2002.

[4] W.M.P. Vander Aalst and M. Song. Mining Social Networks: Uncovering interaction patterns in business processes. In M. Weske, B. Pernici, and J. Desel, editors, International Conference on Business Process Management (BPM 2004), Lecture Notes in Computer Science, Springer-Verlag, Berlin, 2004.

[5] W.M.P. van der Aalst, B.F. van Dongen, J. Herbst, L. Maruster, G. Schimm, and A.J.M.M. Weijters. Workow Mining: A Survey of Issues and Approaches. Data and Knowledge Engineering, 47(2):237{267, 2003.

[6] W.M.P. Vander Aalst and A.J.M.M. Weijters, editors. Process Mining, Special Issue of Computers in Industry, Volume 53, Number 3. Elsevier Science Publishers, Amsterdam, 2004.

[7] W.M.P. Vander Aalst, A.J.M.M. Weijters, and L. Maruster. Workow Mining: Discovering Process Models from Event

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 1, Issue 6, June 2012

Logs. QUT Technical report, FIT-TR- 2003-03, Queensland University of Technology, Brisbane, 2003. (Accepted for publication in IEEE Transactions on Knowledge and Data Engineering.).

[8] R. Agrawal, D. Gunopulos, and F. Leymann. Mining Process Models from Workow Logs. In Sixth International Conference on Extending Database Technology, pages 469{483, 1998.