A Self Adaptive Learning Approach for Optimum Path Evaluation of Process for Forensic use to Finding Uniqueness

[1]Pallavi Ramteke pallavi.scjp@gmail.com
(P.G.Scholar)
Department of Computer Science. LNCT, Indore(M.P.) 453331
[2]Yashpal Kadam yashpal.kadam@rediffmail.com(Asst.Prof.)
Department of Computer Science. LNCT, Indore(M.P.) 453331
[3]SanketGupta
sanket.jec@gmail.com
(P.G. Scholar)
Department of Computer Science ,SATI, Vidisha(M.P.) 464001

Abstract
One approach to protected systems is from first to last the examination of audit trails or logs. 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 specific individual or system element. In the optimum path evaluation of a process working with the audit log of the system generated by our data format, a different kinds of processes accessed during the different user session .our proposed work is based on the concept of forensic activities and the process mining, where we introduce a new process analysis method by which we discover what the users next application.

Keywords - Audit Logs, reference graph building,forensic,optimum path evaluation, process mining.

I. INTRODUCTION

In auditing log file investigations use of Process mining stands for a set of techniques to analyze process models and logs. However, the extent to which it can be used for security auditing that has not been investigated. the data examined may not contain the contents but just discloses activities that were electronically conducted by the criminals. In this paper, our primary concerns mostly address the auditing log files, that is, we are only concerned with users' activities and determine the meaning of the sequence of the activities indirectly. Indeed, only retrieving the user's malicious activities from the auditing log files is very challenging. 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 .Mining and analysis of this system access log can be helpful for extracting different kind of security applications, forensic applications or performance matrix application. there are a large amount of efforts and applications based on data mining is found in different domain of knowledge processing but with process mining that is too rare.

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), and
(iv)Events have a timestamp and are totally ordered.

In the proposed work we are working with the audit log of the system which is generated by our data format different kinds of processes accessed during the different user session, our proposed work is based on the concept of forensic activities and the process mining, where we introduce a new process analysis method by which we discover what the users next application.

The goal of mining this perspective is to and a good characterization of all possible paths.

The rest of this paper is organized as follows: Section II reviews auditing log files and Background work; Section III proposes work ; Section IV explains how to construct a Directed Graph and working of logs according to events and shows similarity function; Section V proposed system architecture; finally, we conclude the paper in Section VI.

I. BACKGROUND WORK

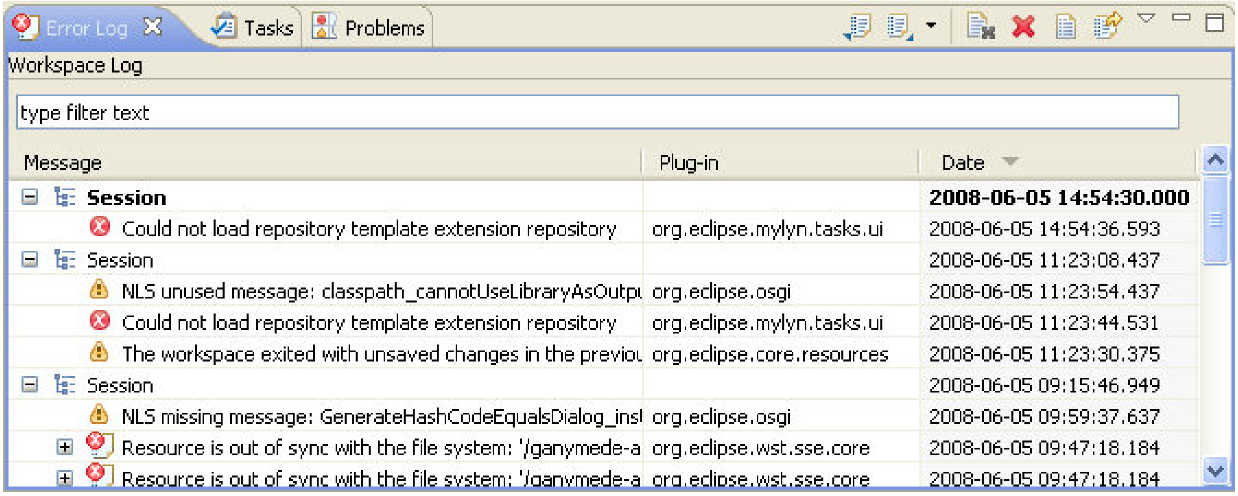
A. Auditing Events

1. 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.

2. 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(System Log) 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 non existent directory, in a world-write able 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.

3. Error Logs : The Error Log view captures all the warnings and errors logged by plug-ins. The underlying log file is a *.log* file stored in the *.metadata* subdirectory of the workspace.

Example of error log :



Workspace Log

type filter text

Message	Plug-in	Date
⊟ **Session**		**2008-06-05 14:54:30.000**
Could not load repository template extension repository	org.eclipse.mylyn.tasks.ui	2008-06-05 14:54:36.593
⊟ Session		2008-06-05 11:23:08.437
NLS unused message: classpath_cannotUseLibraryAsOutpu	org.eclipse.osgi	2008-06-05 11:23:54.437
Could not load repository template extension repository	org.eclipse.mylyn.tasks.ui	2008-06-05 11:23:44.531
The workspace exited with unsaved changes in the previou	org.eclipse.core.resources	2008-06-05 11:23:30.375
⊟ Session		2008-06-05 09:15:46.949
NLS missing message: GenerateHashCodeEqualsDialog_insl	org.eclipse.osgi	2008-06-05 09:59:37.637
⊞ Resource is out of sync with the file system: '/ganymede-a	org.eclipse.wst.sse.core	2008-06-05 09:47:18.184
⊞ Resource is out of sync with the file system: '/ganymede-a	org.eclipse.wst.sse.core	2008-06-05 09:47:18.184

Fig.1 Error LogsEvents in the log view can be sorted by Message, Plug-in ID or Date in ascending or descending order. Simply click on the column header that you want the sorting to be based on. In most of the cases auditing log files are ordered in a time sequential manner, two consecutive logs or actions are likely to be irrelevant or not related to each other.

B. Process Mining in Forensics

The main use of audit logs was to monitor performance and to detect intrusions originating from an external source. With the passage of time however, the term “intrusion” has begun to express a wider meaning closely related to security policy.

Forensic Data Mining (or Forensic Mining) which originated from Forensic Science can be described as the application of data mining techniques and other scientific tools to investigative process for good evidence. In the same vein, the investigative process of auditing called forensic auditing could be defined as the application of auditing skills to situations that have legal consequences.

In most of the cases auditing log files are ordered in a time sequential manner, two consecutive logs or actions are likely to be irrelevant or not related to each other. This can be especially true for the network auditing log files. Moreover, in the network auditing log files, which are captured by the aforementioned network monitoring tools, even one action produces many packets.

The use of process mining for (internal and external) audits has been gaining in momentum. However, to the best of our knowledge there is no previous academic work addressing the intersection of security audits and process mining. The idea of process mining is to discover, monitor and improve real processes (i.e., not assumed processes) by extracting knowledge from event logs readily available in today’s (information) systems. Process mining includes (automated) process discovery (i.e., extracting process models from an event log), conformance checking (i.e., monitoring deviations by comparing model and log), social network/organizational mining, automated construction of simulation models, model extension, model repair, case prediction,and history-based recommendations.

III. PROPOSED WORK

In the proposed working with the audit log of the system which is generated by our data format, different kinds of processes accessed during the different user session, our proposed work is based on the concept of forensic activities and the process mining, where we introduce a new process analysis method by which we discover what the users next application.

The complete audit log contains the following information:

1. User type
2. Process Name
3. Access Time
4. Remote Process
5. Defined Process
6. New Process

This information is represented using a weighted and directed graph.

Suppose the complete log contains only three processes than the process can be known as A, B, C and the weighed graph is designed as the below and all other session which use this process in defined as the sub graph of the reference graph, this weighted graph is used as weighted matrix to evaluate the next user session using random walk theory and traversing probability. which node is required to explore.

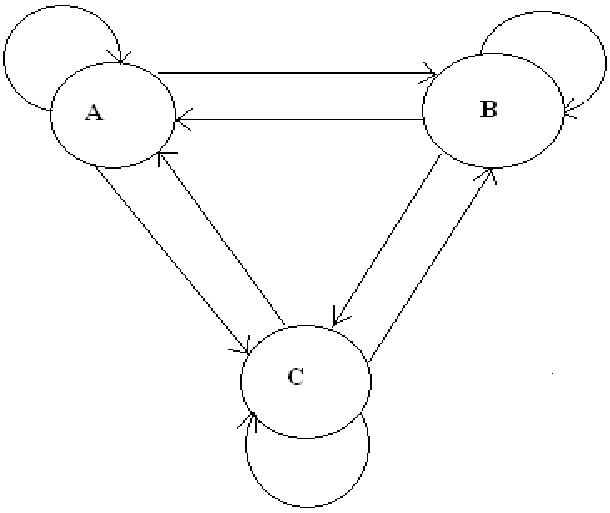


Fig.2 Complete Graph of Process

IV. SYSTEM ARCHITECTURE

In the system architecture data is collected form n no. of systems which are connected in network. This data is mainly contains the Application Logs , System Logs , Audits Logs and Error .All the Logs having different functionally and purpose in which first is

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.

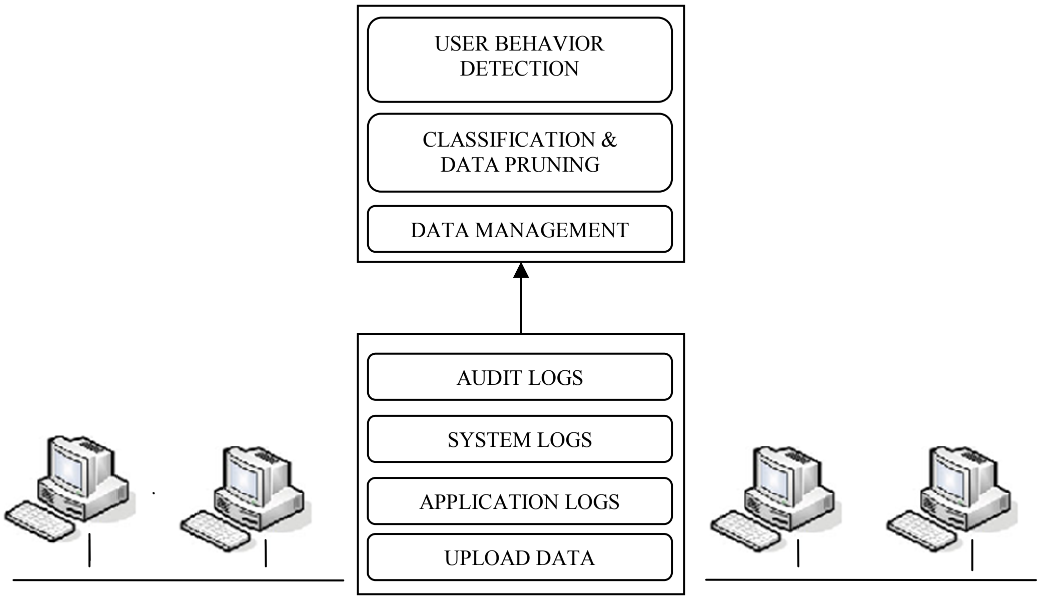


Fig.3 System Architecture

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.
Audit Logs : An audit trail (or audit log) is a security-relevant chronological record, set of records, or destination and source of records that provide documentary evidence of the sequence of activities that have affected at any time a specific operation, procedure, or event.
Error Logs : The Error Log view captures all the warnings and errors logged by plug-ins.

All this Logs are now come to the data management section in which data are managed by using the classification algorithms. After this data are pruned with the help of algorithm, final data come for the detection process. For the users behavior detection we use random walk theory.
Random Walk Theory : The random walk theory is based on stock price change. An investment theory which claims that market prices follow a random path up and down, without any influence by past price movements, making it impossible to predict with any accuracy which direction the market will move at any point. Investors who believe in the random walk theory feel that it is impossible to outperform the market without taking on additional risk, and believe that neither fundamental analysis nor technical analysis have any validity.

V. IMPLEMENTATION

The implementation of the required model is under java virtual machine, which contains a rich class library for computational ability.
As we discus in the above section we know all the path of the session is a sub-graph of the reference graph.

Reference Graph Building

- count the number of unique attributes in the attribute set.
- find all possible path using the formula
  path = n!

Similarity computation

The computation is based on the most relevance probability match which is estimated using First, we build the Image similarity graph using following cosine function.

$$\mathrm{Sim}(d_p, d_q) = \frac{\mathbf{d}_p \cdot \mathbf{d}_q}{||\mathbf{d}_p||||\mathbf{d}_q||}.$$

Where,

d_p belongs to D and D is the total reference set.
d_q is the nearest graph of d_p

Error correction

here we use the iterative technique to reduce the predictive error in the proposed algorithm where

error = real value – predicted value

that is adjusted in the next predictive value.

VI. EXPECTED OUT COME

After the successfully implementation of the proposed algorithm we found the following improvement over the existing system :

1. Implementation of new predictive technique that is based on the graph theory.
2. A method that promises to provide the high performance results .
3.A comparative study between tradition next path estimation and new path estimation technique.

VII. CONCLUSION

The log files range from a simple system log file to a server side web-based email log file. However, it is easily supposed that our investigations of covert channels are not completed due to the lack of information. The proposed method is use user relationship graph. This method give better investigation because it maintain session by sub graphs. Previous methods use log file for investigation and these method does not cover complete information of overall sessions.

VIII. REFERENCES

[1] Daisuke Takahashi, Yang Xiao, and Ke Meng "Creating User-Relationship-Graph in Use of Flow-Net and Log Files for Computer and Network Accountability and Forensics" The IEEE Military Communications Conference-Cyber Security and Network Management ,pp1941-47,2010

[2] Zhenyu Wang, Qing Yao, Yuqing Sun"The Research of Process Mining Assessment used in Business Intelligence" IEEE/ACIS 11th Inter national Conference on Computer and Information Science,pp179-83,2012

[3] Dina Hadžiosmanovi´c ,Damiano Bolzoni ,Pieter H. Hartel"A log mining approach for process monitoring in SCADA" Int. J. Inf. Secur.pp231–251,2012

[4] Wil M.P. van der Aalst, Schahram Dustdar" Process Mining Put into Context" Published by the IEEE Computer Society,pp82-86,2012

[5] Wil van der Aalst" Process Mining: Making Knowledge Discovery Process Centric" SIGKD D Explorations, Volume 13, Issue 2,pp45 49,2012

[6] Rafael Accorsi ,Thomas Stocker" On the Exploitation of Process Mining for Security Audits:
The Conformance Checking Case" ACM 978-1-4503-0857-1/12/03, March 26-30, 2012

[7] D. Takahashi and Y. Xiao, "Retrieving Knowledge from Auditing Log Files for Computer and Network Forensics and Accountability," (Wiley) Security and Communication Networks. Vol. 1. No. 2, pp. 147 – 160, Feb. 29, 2008

[8] L. Ding and B. Dixon, "Using an edge-dual graph and k-connectivity to identify strong connections in social networks," Proc. of the ACM 46th Annual Southeast Regional Conference, pp. 475-480, 2008.

[9] B. Fu and Y. Xiao, "An Implementation Scheme of Flow-Net and Its Applications on Detecting Attacks in Wireless Networks," Proc. of IEEE GLOBECOM 2010

[10] R. Mayrhofer, K. Nyberg, and T. Kindberg, "Foreword," International Journal of Security and Networks, Vol. 4, Nos. 1/2, pp. 1 - 3, 2009.

[11] S. Ehlert, Y. Rebahi, and T. Magedanz, "Intrusion Detection System for Denial-of-Service flooding attacks in SIP communication networks," International Journal of Security and Networks, Vol. 4, No.3, pp. 189 - 200, 2009.

[12] K. Tsai, C. Hsu, and T. Wu, "Mutual anonymity protocol with integrity protection for mobile peer-to-peer networks," International Journal of Security and Networks, Vol. 5, No.1 pp. 45 - 52, 2010, DOI: 10.1504/IJSN.2010.030722