International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 2 Issue 5, May - 2013

Fuzzy-Temporal Log Correlation Approach To Understand End User Charactersitcs

Sourabh Jain
Computer Science Dept .Radharaman Institute of Technology & Science, Bhopal (M.P.) India

Assistant Prof. Susheel Jain
Computer Science Dept .Radharaman Institute of Technology & Science, Bhopal (M.P.) India

Prof. Anurag Jain
Computer Science Dept .Radharaman Institute of Technology & Science, Bhopal (M.P.) India

Abstract— **Cyber fraud analysis in the client server scenario is a tedious job. Log files have an significant duty in recent cyber criminal investigation which is capture in different logs like firewall log file at client side, network log file at gateway and web log file at server side when end user interact with each other in web environment. But single log file is not to be considering as a overall source of information cyber fraud analysis. Whereas for efficient investigation log files of different source of same time are need to be correlated. This paper presents a novel methodology i.e. based on fuzzy and temporal classification and responsible to generate association rule on the basis of chain of evidence and used to preprocess the real generated data from logs and classify the end user.**

Keywords- cyber forensic; log file; correlation; fuzzy assoication rule;temporal correlation ;chain of evidence;cyber crime.

I. INTRODUCTION

With worldwide computerization of an organization lead to attract computer intruder or extruder to make a criminal offence over computer .This criminal offence need a cyber forensic tools to carry out investigations into digital crimes or incidents. The aim of such cyber investigations is narrow the overall investigation process and expose nearby suspicious user, which often leads to prosecution and conviction.

Main problem in cyber investigation is authentication of cyber forensic tools ie how it preserved, collect, validate, identified, analyzes, interpret ate and present the cyber evidence that will be accept in court of law. Gradually enhancement of cyber crimes committed to the development of a whole slew of cyber forensic tools that ensure cyber evidence is acquired and preserved properly and that accuracy of results regarding the processing of digital evidence is maintained [1].

Now these day Researchers that might be focus in cyber forensic area for centralized analysis to log files that's lead easily survey able and detection of correlating events with better rationalization in time consumption of log file examination/cost effectiveness and quicker response time in order to increased security[1,3].

In this paper focuses on the field of cyber forensic, log files, role of log file in cyber forensic, evidence gathering through log file, and various log files management issue and present an framework based on temporal and fuzzy correlation of log file to build the chain of evidence classify the suspicious user.

II. RELATED WORK

More and more researchers pay attentions on the data mining with taking time factor into account, generally speaking, data are coordinated by time, instead of analyzing the data globally, the research focus on the information hidden within certain time slot/period/transaction. Along with that Fuzzy set theory has been used more and more frequently applied in intelligent systems because of its simplicity and similarity to human reasoning. Several frameworks can detect the popular relevant topics specific to a main keyword on certain periods [4]. Three different kinds of relevant topics could be selected by this work which are non-periodical topic, periodical topic and burst topic respectively. By considering the power of the time, it is possible to extract different relevant topics to specific keyword on different time. Guillame-Bert and James L. Crowley [5] introduced a model based on temporal pattern called as Temporal Tree Associative Rule a data mining approach. That can used both uncertainty and temporal inaccuracy of temporal events to express as Symbolic Time Sequences. Maragatham. G and Lakshmi. M [6] proposed an algorithm that able to mine temporal association rules based on utilities by adapting the support with relevant to the time periods and utility [6].

In spite these studies we analyze more research papers. Here we describe some more paper in which the temporal mining can apply.

In [7] an algorithm had been proposed in order to get fuzzy temporal association rules from a publication database. Where algorithm calculates the life of an item

set with the help of publication database. It counts the publication period of an item sets. Also, an item set table structure is designed to effectively keep and efficiently obtain information of item sets for mining. Whereas digitization plays an important role so Shusaku Tsumoto, Shoji Hirano and Hidenao Abe and Yuko Tsumoto [8] proposed the use of temporal mining in the hospital information system. It provides the efficient hospital management and services with evidence and proposed a beginning approach to service improvement in a hospital by using data mining in which temporal behavior of global hospital activities are visualized. The outcomes of the proposed work show the improvement of the hospital services. It also provides the efficient tool to reuse the hospital data in order to better hospital management. As far as the MANET is concert, it is a collection of node in order to communicate without any cartelized device. In [9] temporal mining can apply in MANET the resultant packet can choose the path randomly. But there are some similarity between the packets which are independent from each other, there is a challenge to find there hidden patterns of relationship. . The proposed work applies association rules method in time domain to achieve these hidden relations. Whereas for knowledge discovery concern, Knowledge discovery in databases is an important task of data mining technique there is need to enhance this method. Tim Schluter and Stefan Conrad [10] proposed two types of tree structure. First one is EP tree and other one is ET tree. Both tree is improved methodology of basic association rile mining and applied this method on the market analysis for collecting the information about the market and apply the temporal data mining with the proposed algorithms.

III. FUZZY ASSOCIATION RULES

Fuzzy association rules deal with inaccuracies in physical measurements and better handle unnatural boundaries found in crisp partitions. They provide a linguistic interpretation of numerical values for interfacing with experts. Evolving fuzzy association rules [9] enhances the interpretability of quantitative association rules. A emerging methodology based on classical association rule mining is Fuzzy Association Rules mining. Fuzzy set theory has been used more and more frequently in intelligent systems because of its simplicity and similarity to human reasoning [5, 7, 10]. Several fuzzy data mining algorithms for inducing rules from given sets of data have been designed and used to good effect with specific domains [3]. As to fuzzy temporal data mining, since fuzzy calendar algebra could help users describe temporal requirements in fuzzy temporal calendars easily, Lee proposed two temporal patterns that were fuzzy temporal association rules and fuzzy periodic association rules based on fuzzy calendar algebra [11]. Based on Lee's approach, Zhuo et al. introduced a relativity based interest measure value for mining fuzzy calendar-based temporal association rules. However, those fuzzy data mining approaches didn't take item lifespan into consideration. Although Lee proposed two algorithms for discovering fuzzy temporal association rules and fuzzy periodic association rules by using fuzzy calendar algebra [9], lifespan of each item still didn't be considered. Also, the life spans of items in publication stores are different from those in general stores because of their continuous exhibition periods.

IV. TEMPORAL ASSOCIATION RULES

Extraction of temporal patterns from large data sets is the main concern of temporal data mining. The usefulness of temporal data mining continues to grow as increasing amounts of temporal data about everyday activities become available. Three classes of temporal patterns exist:

Temporal Association Patterns: - Associative relations between times sampled observations. For example, 'if the door bell rings at any instance, someone probably enter in the house with in some minutes'

Temporally Constrained:- Represent, progressive pattern of time. For example, 'Between 7am and 9am, people who buy bread in any parlor, also buy milk with a 90% chance'.

Temporally Evolving Association Patterns:- Represent the successive pattern over time. For example, 'If I leave my house now, I will be in shopping mall after t minutes where t depending on traffic of the day.

V. PROPOSED FRAMEWORK

Proposed framework for fuzzy temporal log correlation present two phase architecture, where phase first encompasses three layer's architecture where duty of each and every layer is exclusive but depends upon previous layer output. In this architecture first preprocess the real generated past data (record of event perfume by each and every client) from logs then passing its via whole layer and generate training data set and fuzzy temporal association rule for classification. Furthermore second phase classify the user on behalf of that classification rule..

The Proposed framework can be defined by following stages:

Centralization of Log files: In this step, log files maintained by the web server and firewall are extracted to be to be centralized on the centralization server. The files are converted in such a format that analysis work can be done easily. The centralization server may have any compatible database as a backend, which can store numerous entries as they are.

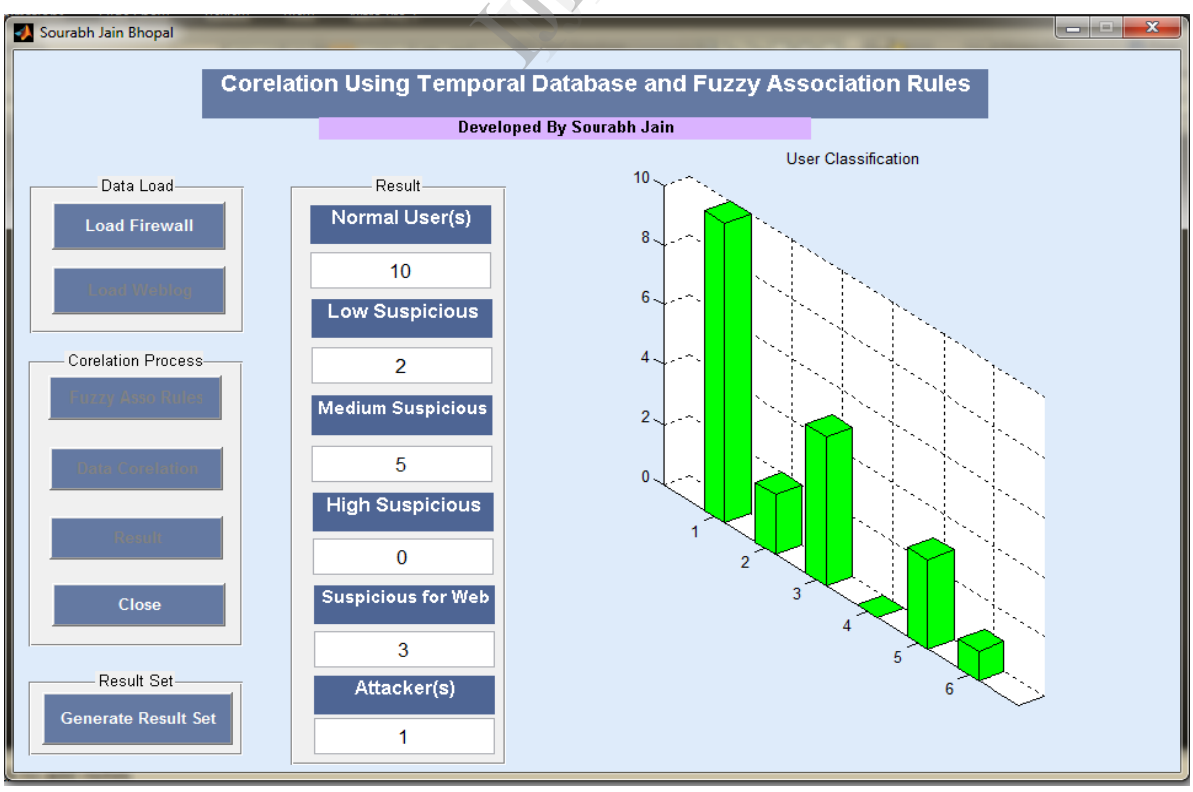
A. F–T Correlation: F T Correlation(fuzzy temporal coreelation) is the process of analysing and determining a set of related events, based on a set of rules that are used to interpret the data contained in the events [4]. There are several types' correlations; some of them are as follows,

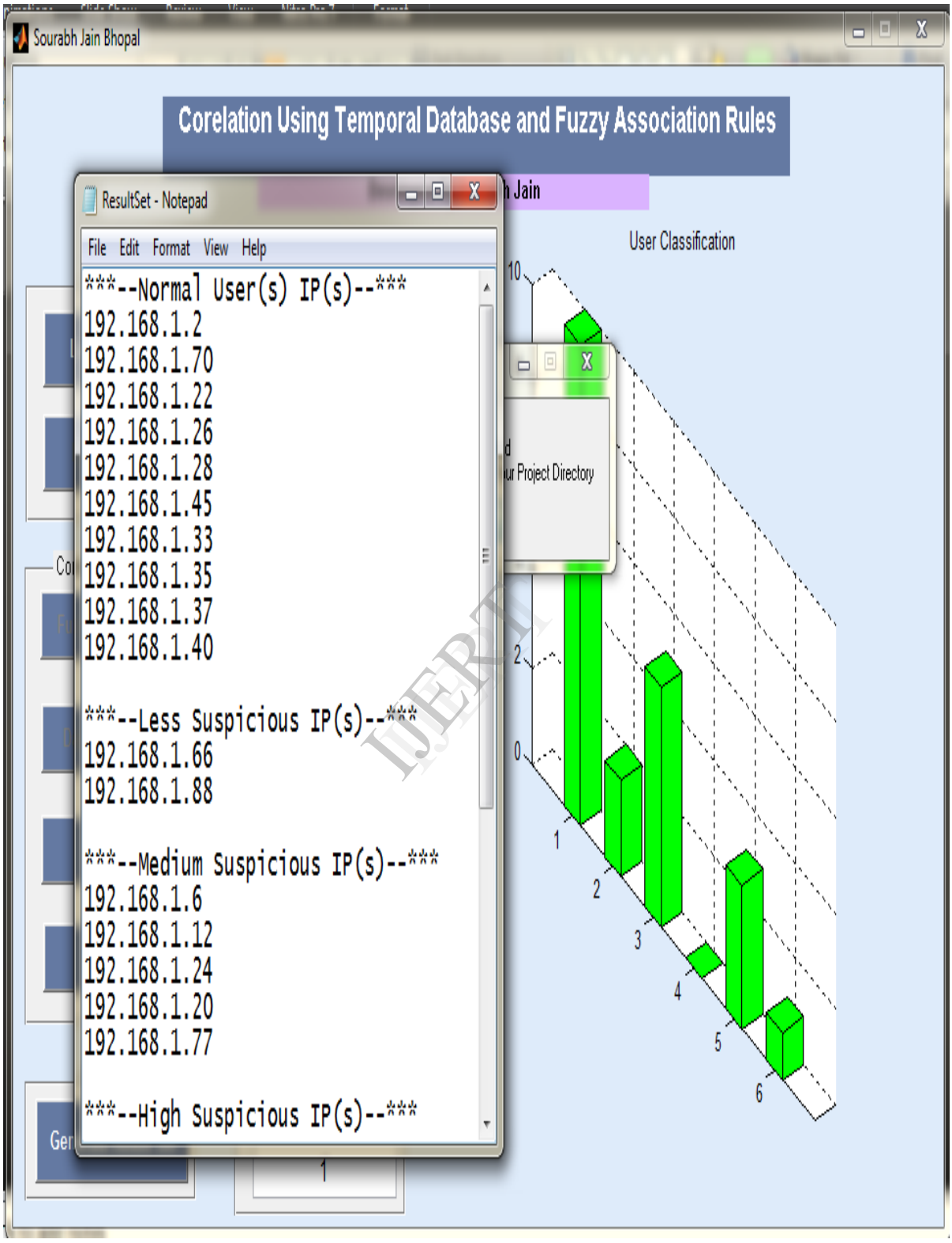
- Sequential correlation: The ability to sort the events in a log by various fields contained in the events (for example, time stamp)
- Associative correlation: The ability to filter or group the events displayed in a log by the values in various fields contained in the events (for example, grouping by thread ID). The correlation types can be used together to provide a complete picture. For example, when grouping a set of events together you typically also order the events in the group.

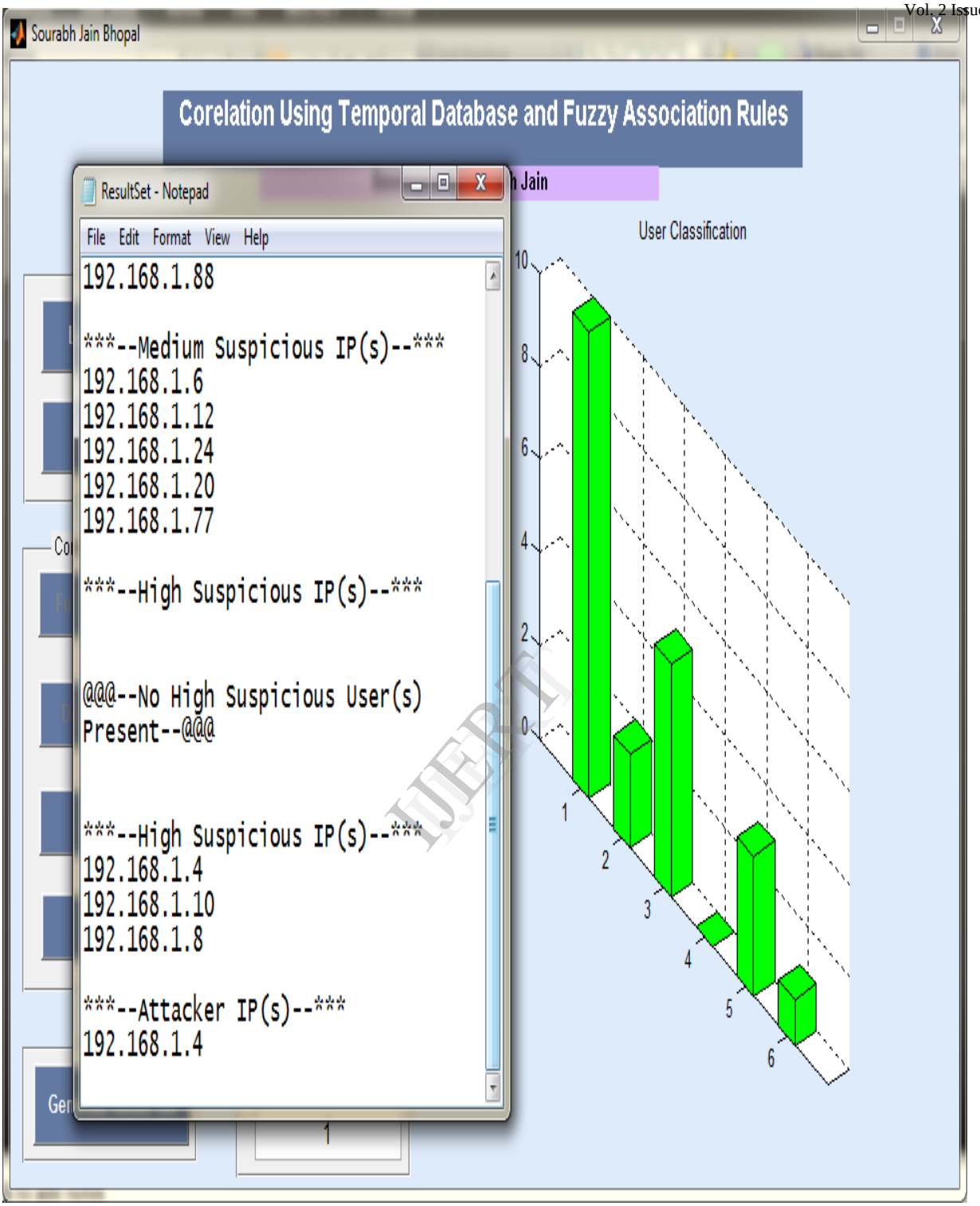
B. Classification of end user-: Proposed Frame work having two different phase where second phase responsible for classification of end user on the basis of its correlation charcertisics with the help of F-T assosiaction rule.

VI. Implementation Details

Proposed frame work has been implemented by using a real time scenario of client server architecture having 20 clients and 1 server. Which capture firewall log and the web log file at client side and server side respectively of same time and use it as input and generates a fuzzy –temporal association Rule that analysis client behavior whether client is normal user, suspicious user, suspicious web user or an attacker?




Sourabh Jain Bhopal
Corelation Using Temporal Database and Fuzzy Association Rules
ResultSet - Notepad
File Edit Format View Help
--Normal User(s) IP(s)--
192.168.1.2
192.168.1.70
192.168.1.22
192.168.1.26
192.168.1.28
192.168.1.45
192.168.1.33
192.168.1.35
192.168.1.37
192.168.1.40
--Less Suspicious IP(s)--
192.168.1.66
192.168.1.88
--Medium Suspicious IP(s)--
192.168.1.6
192.168.1.12
192.168.1.24
192.168.1.20
192.168.1.77
--High Suspicious IP(s)--
User Classification
10
2
0
1
2
3
4
5
6
ur Project Directory
IJERT


Sourabh Jain Bhopal
Corelation Using Temporal Database and Fuzzy Association Rules
ResultSet - Notepad
File Edit Format View Help
192.168.1.88
--Medium Suspicious IP(s)--
192.168.1.6
192.168.1.12
192.168.1.24
192.168.1.20
192.168.1.77
--High Suspicious IP(s)--
@@@--No High Suspicious User(s)
Present--@@@
--High Suspicious IP(s)--
192.168.1.4
192.168.1.10
192.168.1.8
--Attacker IP(s)--
192.168.1.4
User Classification
10
8
6
4
2
0
1
2
3
4
5
6

VII. EXTRACTING OF FUZZY TEMPORAL ASSOISATION CLASSIFICATION RULES BASED:

$$R1:\ F^{[0,10]}_{\text{ip having number of port scaning is geater than 200}} + W^{[05,15]}_{\text{IP having web entry}} + W_{\text{IP having restricted zone web entry}} \longrightarrow IP_{Attacker}$$

$$R2: F^{[0,10]}_{\text{ip having number of port scaning is geater than 200}} + {}_{not} W^{[05,15]}_{\text{IP having web entry}} + {}_{Not}\ W_{\text{IP having restricted zone web entry}} \longrightarrow IP_{\text{Web Suspicious user}}$$

$$R3: F^{[0,10]}_{\text{ip having number of port scaning is greater than 550}} + W^{[05,15]}_{\text{IP having web entry}} + {}_{Not}\ W_{\text{IP having restricted zone web entry}} \longrightarrow IP_{\text{High Suspicious user}}$$

$$R4: F^{[0,10]}_{\text{ip having number of port scaning is bw 350 \& 550}} + W^{[05,15]}_{\text{IP having web entry}} + {}_{Not}\ W_{\text{IP having restricted zone web entry}} \longrightarrow IP_{\text{Medium Suspicious user}}$$

$$R5: F^{[0,10]}_{\text{ip having number of port scaning is bw 150 \& 450}} + W^{[05,15]}_{\text{IP having web entry}} + {}_{Not}\ W_{\text{IP having restricted zone web entry}} \longrightarrow IP_{\text{Less Suspicious user}}$$

$$R6: F_{\text{ip having number of port scaning is less than 200}} \longrightarrow IP_{\text{Normal user}}$$

VIII. Result Analysis

In this paper, the MATLAB (7.1.4) simulated experiments are performed to verify the accuracy of proposed model. Log format synchronization is one of great challenge in log management issue, recently researcher focus on that problem. Proposed model in [6,7] is log format dependent where as Proposed model is not format dependent

SN	Year	Log Format	Time	Size	Fuzzy based
1	IEEE,2012 [13]	D	D	ND	N
2	IEEE,2011 [14]	D	ND	ND	N
3	IEEE,2011 [15]	ND	D	D	N
Proposed Methodology		ND	D	ND	Y

.Along with that log rotation (size of log file) and clock synchronization is another most challenging issue in log management .Proposed model is time dependent and result set is independent form size of log file.

IX. Conclusions

Proposed model extracts the evidence from log file and correlate these generated logs on the basis of fuzzy –temporal association classification rule generated by fuzzy temporal correlation of firewall and web log file. Proposed frame work encourages the web investigator to navigate the end user behavior and assist to enforce the effective security policy.
The future work will cover the issues related to log consistency, log integrity and other log management issue.

X. Acknowledgments

The research presented in this paper would not have been possible without our college, at RITS Bhopal. We wish to express our gratitude to all the people who helped turn the World-Wide Web into the useful and popular distributed hypertext it is. We also wish to thank the anonymous reviewers for their valuable suggestions.

XI. References

[1]. Risto Vaarandi "Tools and Techniques for Event Log Analysis", Faculty of Information Technology, Department of Computer Engineering, Chair of System Programming, Tallinn University of technology,2005

[2]. Muhammad Kamran Ahmed, Mukhtar Hussain and Asad Raza "An Automated User Transparent Approach to log Web URLs for Forensic Analysis" Fifth International Conference on IT Security Incident Management and IT Forensics 2009.

[3]. Raymond Kosala, Hendrik Blockeel, Web Mining Research: A Survey,ACM SIGKDD Explorations Newsletter, June 2000, Volume 2 Issue 1.

[4]. Tim Schluter and Stefan Conrad "About the Analysis of Time Series with Temporal Association Rule Mining" IEEE 2011.

[5]. Mathieu Guillame-Bert, James L. Crowley, "New Approach on Temporal Data Mining for Symbolic Time Sequences: Temporal Tree Associate Rules," ictai, pp.748-752, 2011 IEEE 23rd International Conference on Tools with Artificial Intelligence, 2011

[6]. Maragatham. G and Lakshmi. M "A Strategy for Mining Utility based Temporal Association Rules", IEEE 2010, pp 38-41.

[7]. Guo-Cheng Lan, Chun-Hao Chen, Tzung-Pei Hong and Shih-Bin Lin, "A Fuzzy Approach for Mining General Temporal Association Rules in a Publication Database", IEEE 2011, pp 611-615

[8]. Shusaku Tsumoto, Shoji Hirano and Hidenao Abe and Yuko Tsumoto "Temporal Data Mining in History Data of Hospital Information Systems", IEEE 2011, pp 2350-2356

[9]. Ahmad Jabas, Dr. Rama M. Garimella, Prof. S. Ramachandram, "MANET Mining: Mining Temporal Association Rules", IEEE 2008, pp 765-770.

[10]. Tim Schluter and Stefan Conrad "Mining Several Kinds of Temporal Association Rules Enhanced by Tree Structures" IEEE 2010, pp 86-93

[11]. L. Baum et. al. A maximization technique occuring in the statistical analysis of probablistic functions of markov chains. Annals of Mathematical Statistics, 41:164–171, 1970.

[12]. Katherine A.Heller,YEE Whye The And Dilan , "Infinite Hierarchical Hidden Markov Models" in Proceedding Of the12th International conference on AISTATS,2009

[13]. Stefan Hommes, Radu State, Thomas Engel, "A Distance-Based Method to Detect Anomalous Attributes in Log Files" in IEEE ,2012

[14]. P.W.D.C. Jayathilake, "A Novel Mind Map Based Approach for Log Data Extraction " in 6TH international conference on industrial and information system,IEEE,,2011

[15]. Thomas Reidemeister, Miao Jiang and Paul A.S. Ward, "Mining Unstructured Log Files for Recurrent Fault Diagnosis" in IEEE,2011