ISSN : 2278 – 1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 1, Issue 4, June 2012

Handling End User through Markov Model

Mr.Sujeet Singh, Mr.Ganesh Wayal, Mr.Niresh Sharma

Department of Computer Science and Engineering, RKDF, Bhopal, India

Abstract:
In the cyber crime investigation, log files are an important source of evidence. The importance of event logs, as a source of information in systems and network management cannot be over accentuated. With the ever-increasing size and complexity of today's event logs, the task of analyzing event logs has become cumbrous to carry out manually. Now these days recent research has focused on the automatic analysis of these logs files in order to identify suspicious user. This paper through some light on those techniques by which it will easy to retrieve the suspicious user using log files. This paper also shows about web mining and Markov model.

Keywords: Cyber Forensic, Event Correlation, log file, web mining, Markov Model

I. INTRODUCTION

Event logging and event logs play a crucial role in modern world of computer systems. Today, many applications, operating systems, network devices, and other system components are able to log their events to a local or remote log server. Therefore, the event logs are an excellent source for determining the health status of the system. On the other hand, quite many essential event-processing tasks involve event correlation – a conceptual interpretation procedure where new meaning is assigned to a set of events that happen within a predefined time interval. Event correlation is one of the most prominent real-time event processing techniques today. It has received a lot of attention in the context of network fault management and event log monitoring over the past decade [1]. A log file is used to track the operation performed by any user simply by storing messages generated by an application, service, or an operating system [2]. Overall, the log file can help in order to locate the suspicious user or activity for the respective system. this paper through some light on such types of activity and help to find the solution.

This paper is organized into eight sections including this the session. First section defines the introduction of the paper and provides the brief description of topic. The second section through some light on data mining and web mining. The third section have the information regarding log files. The Markov Model has defined in fourth section. Finally paper is concluded into fifth section.

II. OVERVIEW OF WEB MINING

A method in order to discover pattern from the web known as web mining which can be done in the three different way.

1) Content-based web mining 2) Usage Based Web Mining and 3) Structure based Web Mining. But all these are work in a same manner. There are some steps which are define below always take place by web mining

To clarify the confusion to determine what forms Web mining. Kosala and Blockeel [3] had suggested a decomposition of Web mining in the following steps:

1. Resource finding: the task of retrieving intended Web documents.
2. Information selection and pre-processing: automatically selecting and pre-processing specific information from retrieved Web resources.
3. Generalization: automatically discovers general patterns at individual Web sites as well as across multiple sites.
4. Analysis: validation and/or interpretation of the mined patterns should be apply in this stage.

A. CHALLENGES IN WEB MINING

The challenges involved in web usage mining could be divided in three phases [8]:

1. Pre-processing. The data available tend to be noisy, incomplete and inconsistent. In this phase, the data available should be treated according to the requirements of the next phase. It includes data cleaning, data integration, data transformation and data reduction.
2. Pattern discovery. Several different methods and algorithms such as statistics, data mining, machine learning and pattern recognition could be applied to identify user patterns.
3. Pattern Analysis. This process targets to understand, visualize and give interpretation to these patterns. . Web usage mining depends on the collaboration of the user to allow the access of the Web log records. Due to this dependence, privacy is becoming a new issue to Web usage mining, since users should be made aware about privacy policies before they make the decision to reveal their Personal data. For further information about Web usage mining please refer to [3,8,9]. We should note

ISSN : 2278 – 1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 1, Issue 4, June 2012

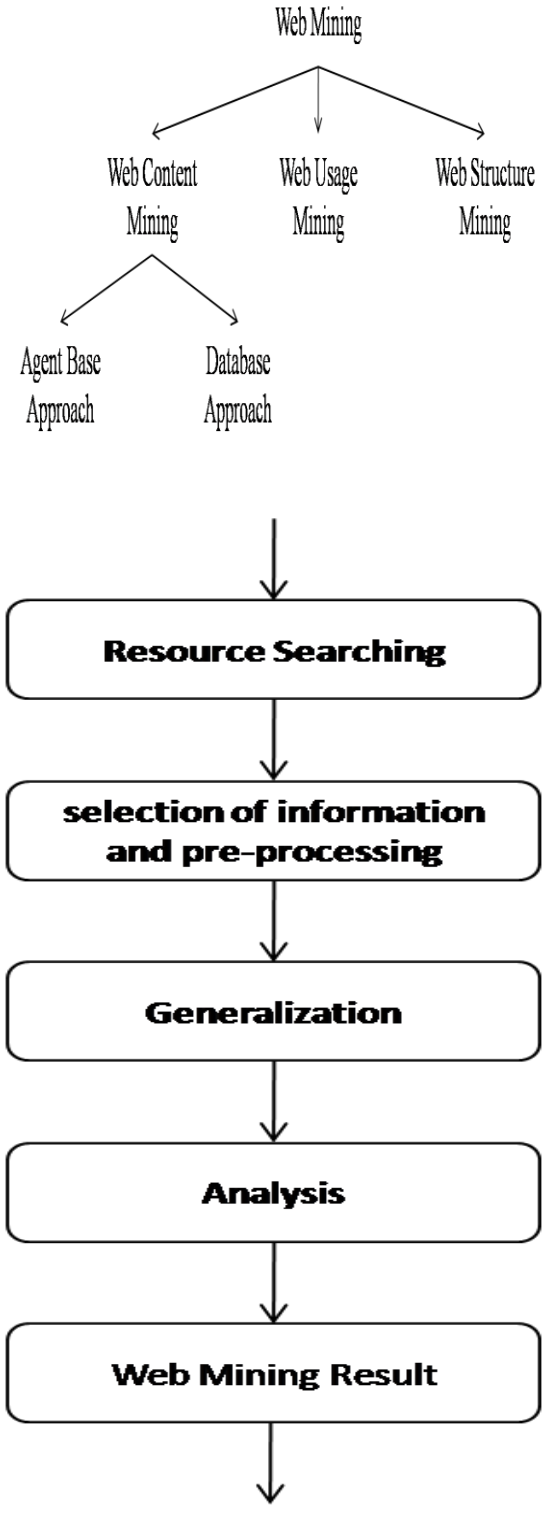

Figure 1: Web Mining Steps

III. LOG FILES

Assuming that a log source offers configuration options, it is generally prudent to be conservative when selecting initial logging settings. A single setting could cause an enormous number of log entries to be recorded, or far too much information to be logged for each event. Excessive logging can cause loss of log data, as well as operational problems such as system slowdowns or even denial of service conditions.

3.1 VARIOUS TYPES OF LOG

Various kinds of log files are available in our computer system; depending on their characteristics several of them are used for different purposes like security, data retrieval, analysing, Authentication & etc. Several of them are following [7].

3.1.1. Network Device Logs

Contain information on network traffic which is meeting an application rule or a packet filter rule. Traffic is not logged unless the Log communication to network log option is enabled. There are several network device logs.

3.1.2. Firewall Logs

Firewall logs provide useful information about the inbound and outbound packets, Information about particular servers e.g. Web Server, Packets which have been dropped, Alerts to the SA, Probing the system.

3.1.3. Ids Logs

IDS logs Provides the information about Alerts on suspicious packet types, Attack statistics (Host / Network based). Helps in determining the probes and generating new attack signatures.

3.1.4. Web Server Logs

A file that records every request and important information about the requests to web server made by users. For example, every time a browser requests a page, an entry is automatically made in this log by the web server, containing information such as the address of the computer on which the browser was running, the time at which the access was made, and the transfer time of the page, etc.

3.2 EVIDENCE GATHERING THROUGH LOG FILE

Web server logs are a significant resource for evidence gathering in cyber forensics. Web servers maintain log files that records user navigation activities on web site. If the

ISSN : 2278 – 1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 1, Issue 4, June 2012

suspicious end user exploits web form as an access point for input attacks like cross-site scripting, SQL injection and buffer overflow attack on web application, it may be detected using the log file [6]. An interesting question is raised, why event data should be logged on a given system. Essentially there are four categories of reasons.

Accountability: Log file data can be used to identify which type of accounts are associated with certain events and that information can be used to emphasize where training and/or disciplinary actions are needed.

Reconstruction: What was happening before and during an event can be reviewed chronologically by using log file data. For this it should be ensured that the clocks are regularly synchronized to a central source to ensure that the date/time stamps are in synchronization. Accuracy and coordination of system clocks is a critical task, that's why reconstruction of events not an easy task.

Intrusion Detection: Log data can be reviewed for detecting unusual or unauthorized events, assuming that the correct data is being logged and reviewed. But variation of unusual activities is a main problem i.e. login attempts outside of designated Schedules, failed login attempts, port sweeps, locked accounts, network activity levels, memory utilization, key file/data access, etc.

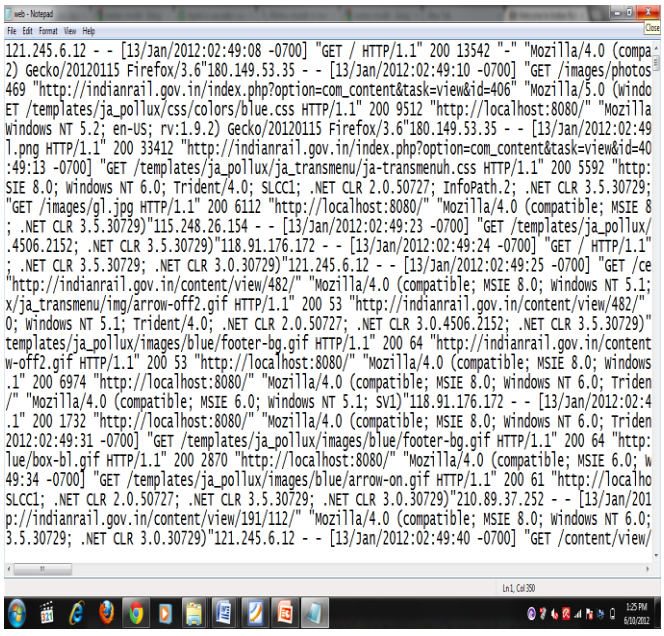

Figure 2: Web Log Example

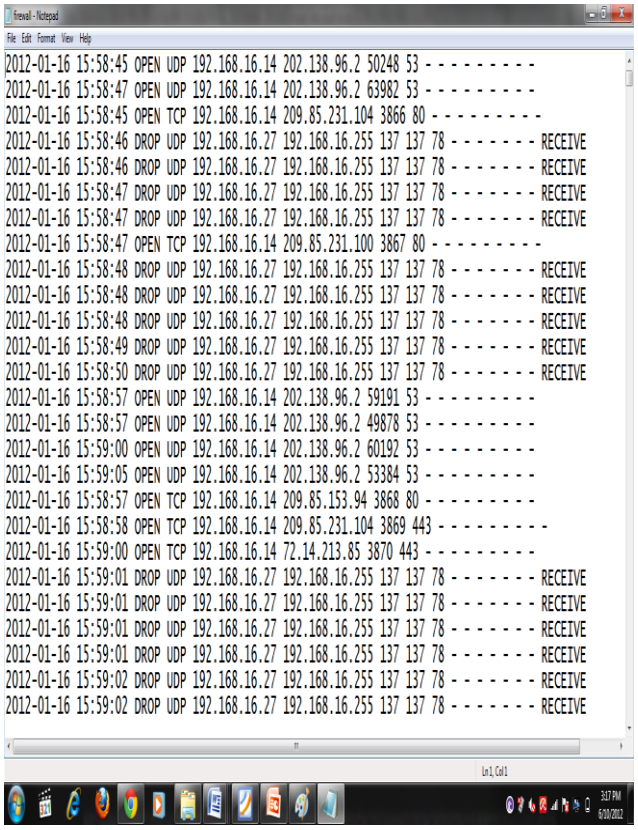

Figure 3: Firewall Log

IV. MARKOV MODEL

The Markov Model is a powerful statistical tool in order to modelling generative sequences that can be characterized by an underlying process generating an evident sequence. In other words A Probabilistic process over a finite set is known as Markov model. These final set called its states for a instance {S1,S2,….Sk}. [5]

The formal definition of a Markov Model is as follows:

$$\lambda = (A, B, \pi) \text{…….(1)}$$

S is our state alphabet set, and V is the observation alphabet set:

$$S = (s1, s2, \cdots, sN) \text{ ……. (2)}$$

$$V = (v1, v2, \cdots, vM) \text{ ……. (3)}$$

We define Q to be a fixed state sequence of length T, and corresponding observations O:

$$Q = q1, q2, \cdots, qT \text{ ……. (4)}$$

$$O = o1, o2, \cdots, oT \text{ ……. (5)}$$

A is a transition array, storing the probability of state j following state i . Note the state transition probabilities are independent of time:

$$A = [aij\ ]\ , aij = P(qt = sj\ |qt{-}1 = si). \text{ ……. (6)}$$

ISSN : 2278 – 1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 1, Issue 4, June 2012

B is the observation array, storing the probability of observation k being produced from the state j, independent of t:

$$B = [b_i(k)] , b_i(k) = P(x_t = v_k | q_t = s_i). \quad \ldots\ldots\ldots (7)$$

π is the initial probability array:

$$\pi = [\pi_i] , \pi_i = P(q_1 = s_i). \quad \ldots\ldots\ldots (8)$$

Markov Model have focus on application in many eras interested in digital signal processing, and in particular speech processing. Andrei Markov gave his name to the mathematical theory of Markov processes in the early twentieth century[4], but it was Baum and his colleagues that developed the theory of Markov Model s in the 1960s[5].

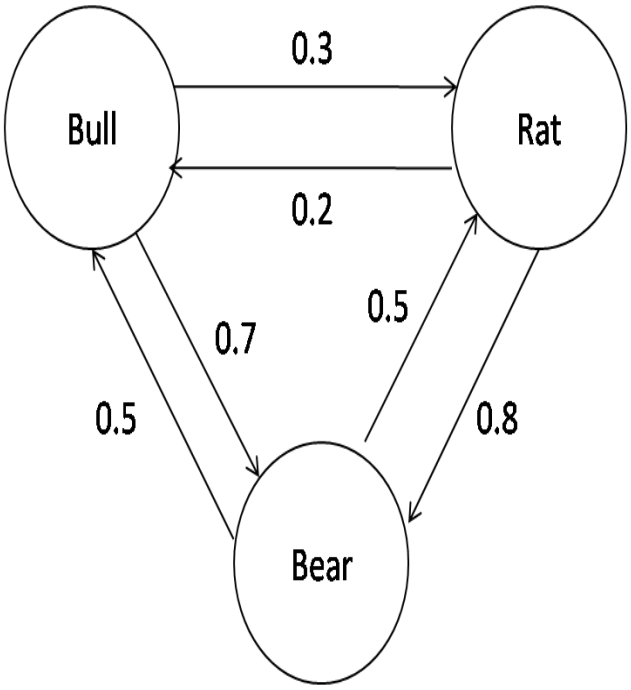

Figure 4: Simple Markov Model

V. CONCLUSION

Web server logs are commonly captured the behaviour of machine not the behaviour of end user. Log file provide troubleshooting, security and pro-active system administration that provide significant help in caching suspicious end user and in process of cyber forensic. This paper is trying to extracts the study of methods by which it is easy to get evidence from different sources, relates generated logs based on Markov Model. This paper also shows that it is possible to apply Markov model in order to implement web mining in various log files.

VI. ACKNOWLEDGEMENT

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

VII. 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] *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.*

[5] *A. Markov. An example of statistical investigation in the text of eugene onyegin, illustrating coupling of tests in chains. Proceedings of the Academy of Sciences of St. Petersburg,1913.*

[6] *Carrier, B.D., Spafford, E.H "Defining Digital Crime Scene Event Reconstruction" Journal of Forensic Sciences, 49(6). Paper ID JFS2004127,2004*

[7] *M. Bishop, "A Standard Audit Trail Format", National Information Systems Security Conference, Baltimore, MD , 1995*

[8] *[11] Han, J., Kamber, M. Kamber. Data mining: concepts and techniques. Morgan Kaufmann Publishers, 2000.*

[9] *[13] Cooley, R.; Mobasher, B.; Srivastava, J.; Web mining: information andpattern discovery on the World Wide Web. Tools with Artificial Intelligence,1997. Proceedings., Ninth IEEE International Conference. Page(s):558 – 567 -3-8 Nov. 1997.*