ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 4, April 2013

Digital Crime Investigation using Various Logs and Fuzzy Rules: A Review

Deepak Meena[1], Hitesh Gupta[2]

Research Scholar, CSE, Patel Institute of Technology, Bhopal [1]

HOD, CSE, Patel Institute of Technology, Bhopal[2]

Abstract: Computer crime has increased a lot now days in the form of hacking. It may harms for the computer system as well as the human being. The computer crimes take place all over the world by any one. It is very hard to investigate such type of case in short span of time. In order to perform the digital forensic there is a need to use the multi-relation classification. Multi-relational data mining enables pattern mining from multiple tables. This paper is a review of digital forensic. It also gives some introduction related to log files. Here log file is an important factor to investigate the crime scene. This paper also summarized the fuzzy rules.

Keywords: Digital forensic, Cyber Crime, Log Files, Fuzzy Rules

I. INTRODUCTION

Rapid growth of cyber crime takes the interest of researchers in the world of digital forensic. Digital forensic is a science to help the investigator in order to identify & analyze the evidence which are collected from the computers or the networking devices. As far as the digital forensic is concert it works with the multi-relation classification [1,2].

Most of the time it seems to be that the investigation has to perform with huge data set. When data is large there is a need to narrow the search for the criminal. It is a major problem before us [2].

The log file contains the whole information regarding the user's activity. These activity is written in various log files like web log, firewall log, network log etc. there log files have the millions of entries. This needs the large time to investigate the case [3].

Here data mining approaches can apply in order to find the related data from the huge data set. The multi-relation classification works with pattern mining. It means the similar type of malicious activity can find with pattern mining and it is possible to reduce the large size of data [1,3].

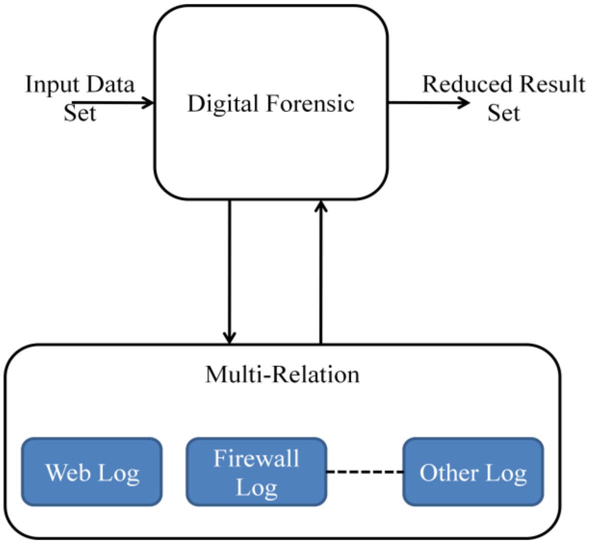


Figure 1: Simple digital investigation approach

This paper is organized into six major parts including first this one. The second section gives the brief introduction of digital forensic. Third section describes the Log file and the various types of Log. It also throws some light on the various attribute of log files. Fifth section gives an idea of fuzzy rules theory and finally the paper was conclude in the section six.

II. DIGITAL FORENSIC

Digital forensic is a field to investigate the computer crime. Normal forensic investigation can use for number of reasons. Criminal activity needs the investigation but as far as the digital forensic techniques are concert it is capable of finding the hacker or the culprit user. Basically this user has its own identity like IP address, DNS name etc [1].

Now days the internet is in access of normal user as well as attacker. Here the forensics investigator should be able to track an attacker on the Internet. The IP address and Domain name tracing is the first step to detect the suspicious user. There are many tools available like nslookup [4].

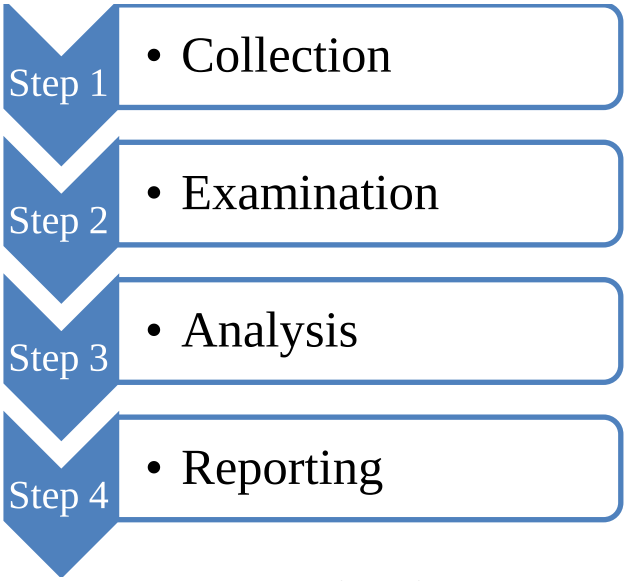


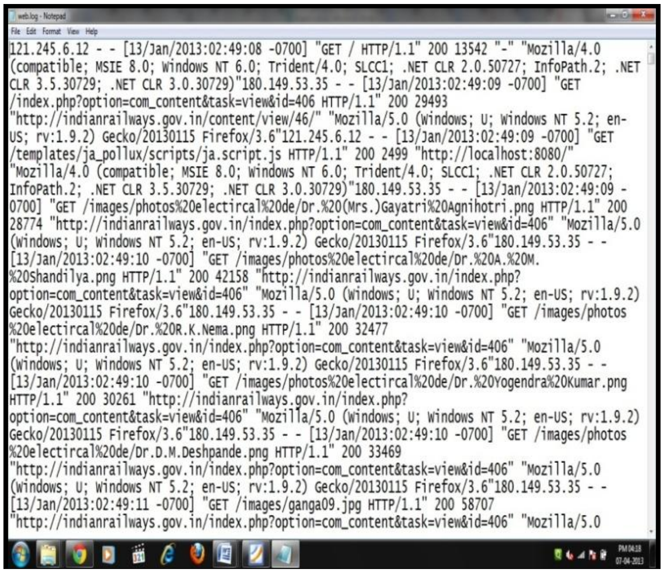
Figure 1: Process of Digital forensic

To perform the digital forensic there are four major steps have to use. First of all there is a need to collect the evidence from the computer system or the crime site. This evidence or data collection can do by the various discussions related to crime. These digital evidences may be CD, Floppy, Hard Disk, Pen Drive, other memories. These collected evidences will have to examine. This examination pay attention to various factors related to the crime. Examiner try to find the answer of "Why" and "Whom". After this evidence will take place into analyses phase. After collecting various proof and resigns finally report will generate that show investigation result [5, 6].

III. LOG FILES

Log file that have all the entry related to incoming user and outgoing user. These file are generated by the process of installation. It can maintain by server machine, firewall, web servers, and routers etc. Generally the log files are in the text format can be read by notepad or simple text editor. Due to the plain text the size of log file will also reduces [7].

Log file don't have the standard format. There are many log formats which was used in randomly by many techniques. Due to availability of make own log format the complexity was increased. To make them understandable everyone some consortiums come forward and decide to make the standards for log files in order to efficient management of log. There are two very popular log formats available first one is web star format and other one is w3c extended format [8].



```
121.245.6.12 - - [13/Jan/2013:02:49:08 -0700] "GET / HTTP/1.1" 200 13542 "-" "Mozilla/4.0
(compatible; MSIE 8.0; Windows NT 6.0; Trident/4.0; SLCC1; .NET CLR 2.0.50727; InfoPath.2; .NET
CLR 3.5.30729; .NET CLR 3.0.30729)"180.149.53.35 - - [13/Jan/2013:02:49:09 -0700] "GET
/index.php?option=com_content&task=view&id=406 HTTP/1.1" 200 29493
"http://indianrailways.gov.in/content/view/46/" "Mozilla/5.0 (Windows; U; Windows NT 5.2; en-
US; rv:1.9.2) Gecko/20130115 Firefox/3.6"121.245.6.12 - - [13/Jan/2013:02:49:09 -0700] "GET
/templates/ja_pollux/scripts/ja.script.js HTTP/1.1" 200 2499 "http://localhost:8080/"
"Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 6.0; Trident/4.0; SLCC1; .NET CLR 2.0.50727;
InfoPath.2; .NET CLR 3.5.30729; .NET CLR 3.0.30729)"180.149.53.35 - - [13/Jan/2013:02:49:09 -
0700] "GET /images/photos%20electircal%20de/Dr.%20(Mrs.)Gayatri%20Agnihotri.png HTTP/1.1" 200
28774 "http://indianrailways.gov.in/index.php?option=com_content&task=view&id=406" "Mozilla/5.0
(Windows; U; Windows NT 5.2; en-US; rv:1.9.2) Gecko/20130115 Firefox/3.6"180.149.53.35 - -
[13/Jan/2013:02:49:10 -0700] "GET /images/photos%20electircal%20de/Dr.%20A.%20M.
%20Shandilya.png HTTP/1.1" 200 42158 "http://indianrailways.gov.in/index.php?
option=com_content&task=view&id=406" "Mozilla/5.0 (Windows; U; Windows NT 5.2; en-US; rv:1.9.2)
Gecko/20130115 Firefox/3.6"180.149.53.35 - - [13/Jan/2013:02:49:10 -0700] "GET /images/photos
%20electircal%20de/Dr.%20R.K.Nema.png HTTP/1.1" 200 32477
"http://indianrailways.gov.in/index.php?option=com_content&task=view&id=406" "Mozilla/5.0
(Windows; U; Windows NT 5.2; en-US; rv:1.9.2) Gecko/20130115 Firefox/3.6"180.149.53.35 - -
[13/Jan/2013:02:49:10 -0700] "GET /images/photos%20electircal%20de/Dr.%20Yogendra%20Kumar.png
HTTP/1.1" 200 30261 "http://indianrailways.gov.in/index.php?
option=com_content&task=view&id=406" "Mozilla/5.0 (Windows; U; Windows NT 5.2; en-US; rv:1.9.2)
Gecko/20130115 Firefox/3.6"180.149.53.35 - - [13/Jan/2013:02:49:10 -0700] "GET /images/photos
%20electircal%20de/Dr.D.M.Deshpande.png HTTP/1.1" 200 33469
"http://indianrailways.gov.in/index.php?option=com_content&task=view&id=406" "Mozilla/5.0
(Windows; U; Windows NT 5.2; en-US; rv:1.9.2) Gecko/20130115 Firefox/3.6"180.149.53.35 - -
[13/Jan/2013:02:49:11 -0700] "GET /images/ganga09.jpg HTTP/1.1" 200 58707
"http://indianrailways.gov.in/index.php?option=com_content&task=view&id=406" "Mozilla/5.0
```

Figure 2: Example of Web Log

The Webster server's log format is highly configurable. Webster supports using Common Log Format (CLF) logs, Extended Log Format (ExLF) logs and Webster Log Format (WLF). Summary requires the following tokens in Webster format logs [9]:

Table 1: Web Star Log's Major Attributes

DATE	Date of request.
URL	The requested item. Same as **CS-URI** and **CS-URI-STEM**.
HOSTNAME	Name or IP address of the requesting computer.
TIME	Time of request.
BYTES	Bytes sent, same as **BYTES_SENT**. Required for all of the Bytes related columns in various reports.
SC-STATUS	It is same as a STATUS field.
REFERRER	Site and page that referred the visitor to your site

Table 2 W3C Extended Log's Major Attributes

DATE	Date of the request.
CS-URI	The requested item
C-IP	Client IP addresses.
TIME	Time of request.
BYTES	Bytes sent.
SC-STATUS	Result code.
CS(REFERER)	Site and page that referred the visitor to your site.

IV. TYPES OF LOG FILES

There are many types of log file used in the digital world. Some of them has discussed below [7,8,9]:

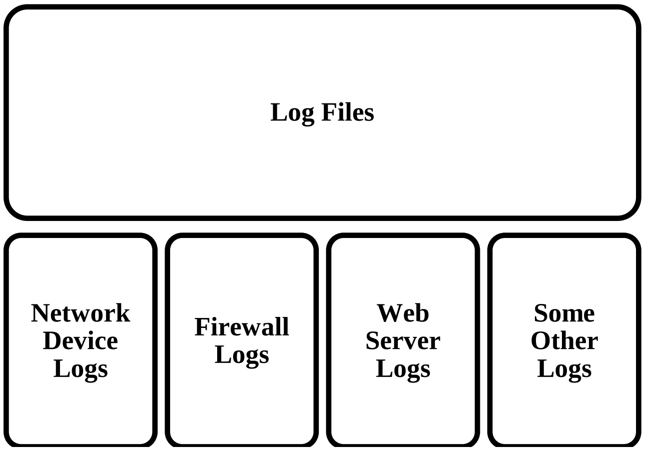


Figure 3: Classification of Log files

Network Device Logs: To perform communication in the network there is a need of various internetworking device like router, hub, switch etc. Most of the devices maintain their own log file. These log files called the networking device logs. There logs enclose information of network traffic which is meeting an application rule or a packet filter rule.

Firewall Logs: Firewall is software or hardware or combination of both is called the firewall. It install in a network in order to monitor the network's incoming and outgoing traffic. It allows or denies the user activity either incoming or outgoing on the basis of predefined rules. All traffic will pass through the firewall so the firewall logs keep all the information about this activity.

Web Server Logs: Web server log is a result given by software program called web server. There are many Web server available like tomcat apache, IIS, Nginx web server etc. each of them maintain access log and reports the number of visitors, views, hits, most frequently visited pages, and so forth.

Some Other Logs: there are many other systems are available those use the log files. IDS Log, Database log, System Log are the example of these log files.

V. FUZZY RULES

Group of items having similar sort of properly is known as set. These set items are the elements of set. This is a traditional approach to represent the set theory[10].
There are some problems with the traditional approach of set theory. The set theory is used for the specific data set. It always gives the answer in Yes or No. some time it seems to be that it is not practically suitable. To remove such types of complexity fuzzy set theory comes in existence.

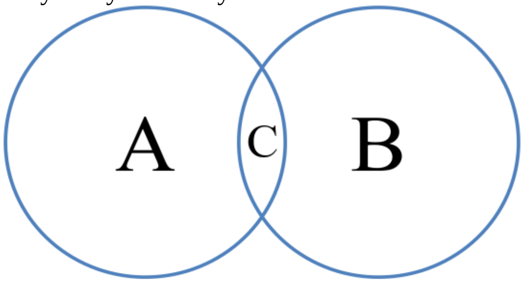


Figure 4: Simple sets

As shown in figure A and B are two different Sets. Each has their own property but set C have the common properties.

Fuzzy set is an extended form of binary set theory. Fuzzy Association Rules mining is an approach which is relied on the traditional association rule. There is very famous problem called sharp boundary problem faced in classical set approach. There are three basic approaches to solve the sharp boundary problem. Quantitative approach, Fuzzy Taxonomic Structures and Approximate Item set Approach are the famous methodology [11].

VI. FUZZY OPERATIONS

The fuzzy operations are same as the traditional approach. Here some major operation and their properties are shown [10, 11].

1) Union

This operator is use to find the elements which are in both the sets. Suppose we have two sets A and B then it will represent by $A \cup B$
Here it is support some other properties

$$A \cup B = B \cup A$$
$$A \cup (B \cup C) = (A \cup B) \cup C$$

2) Intersection

This operator is use to get those items which has the common property related to both the set.

$$A \cap B = B \cap A$$
$$A \cap (B \cap C) = (A \cap B) \cap C$$

3) Complement

It is a set-theoretic difference between sets. If there is a set A then set A' refers the items which are not in A. we can say that items outside the A. symbolically Some time it also represented by $A - B$

Here $(A - B) \neq (B - A)$

VII. Conclusion

Cyber world is a rapid accessible unit all over the world. Having the access of large number of people the crime rate has also increased. Here Digital forensic comes in existence to investigate the crime. In this study it seems to be that the digital forensic is efficiently works with the multi-relation classification using frequent pattern mining. Here log file plays an important role but it has a huge amount of data. To reduce this large search space there is a need to apply pattern mining. This pattern mining can also associate with the fuzzy association rules which give small search space with some connections. This paper is review of digital forensic, log files and the fuzzy rules and operations in order to analyze the malicious activity and the attacker.

Acknowledgement

The research presented in this paper would not have been possible without our college, at Patel Institute of Technology, 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.

References

[1]. Rogers, M.K. and K. Seigfried, The future of computer forensics: A needs analysis survey. 2004. p. 12-16.

[2]. Xue-Gang Hu, Xie-Fei Hu, De-Xing Wang, Dong-Yan Zhang and Chun-Ling Hu "A classification algorithm based on multi-relation domain knowledge", IEEE 2005, pp 2067-2072.

[3]. D.S. Sisodia and S. Verma, "Web usage pattern analysis through web logs: A review", IEEE 2012, 49-53.

[4]. Hamzah, Z., E-Security Law & Strategy. ISBN 967-962-632-6. 2005, Kelana Jaya Malaysia, KL: Malayan Law Journal Sdn Bhd, 47301. 122.

[5]. Ning, L.Z.a.W., Developing a Computer Forensics Program in Police Higher Education, in Computer Science & Education. ICCSE '09. 4th International Conference. 2009. p. pp. 1431-1436.

[6]. Tamas Abraham and Olivier de Vel "Investigative Profilling with Computer Forensic Log Data and Association"IEEE,2002

[7]. Chu-Hsing Lin, Jung-Chun Liu and Ching-Ru Chen, "Access Log Generator for Analyzing Malicious Website Browsing Behaviors", IEEE 2009, pp 126-129.

[8]. Chu-Hsing Lin, Jung-Chun Liu and Ching-Ru Chen, "Access Log Generator for Analyzing Malicious Website Browsing Behaviors" IEEE 2009, pp 126-129.

[9]. Log file format from "http://www.w3.org" accessed on 14/03/2013.

[10]. D. Ramot, M. Friedman, G. Langholz and A. Kandel, "Complex fuzzy logic", IEEE 2003, pp 450-461.

[11]. A. Katbab, "Fuzzy logic and controller design-a review" IEEE 1995, pp 443-449

BIOGRAPHY

Deepak Meena has completed his graduation from VNS Institute of Technology Bhopal & Pursuing M-Tech Program from Patel Institute of Technology Bhopal

Prof. Hitesh Gupta is working in Patel college of science & Technology, Bhopal. He is a Head of Department of Computer science & Engineering Branch.