e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:03/Issue:01/January-2021 Impact Factor- 5.354 www.irjmets.com

IOT DYNAMIC LOG FILE ANALYSIS: SECURITY APPROACH FOR ANOMALY DETECTION IN MULTI SENSOR ENVIRONMENT

Dr. Monika Saxena*1

*1Computer Science & Engineering, Banasthali Vidyapith, India.

ABSTRACT

In the Multisensory environment, the valuable information like: how users interact with smart IoT devices & web servers, are stored in log files. Here web access logs play important role in terms of security and privacy of users. Fetching of needed information, mining of logs, modeling of task and prevention from unauthorized access is main aim to this research. There are already some approaches like security management and network optimization web mining of normal fixed and Ad-hoc networks. But in case of multi sensor, researcher need to manage more secure accessing of users, so researcher is proposing security framework and processing for the same. Main aim of this framework is modeling and examines the HTTP requests that access the sensors. Researchers propose a methodology to analysis of IoT users IP, TCP with aggregated HTTP requests. To evaluate results of the proposed methodology, researcher implement an algorithm for user requests identification. In this research paper researcher are discussing about IoT log files and proposed framework that enhance the security and privacy of users by detection of anomaly user requests.

Keyword: IoT web-based mining, security, intrusion detection, anomaly detection, Multi sensor environment.

I. INTRODUCTION

Now a days all the companies is having online working scenario. Maximum of the industries are using the world wide web to communicate their data and their services acceptability is increasing day by day, in short, the accessing of internet bandwidth is increasing rapidly. on the web many of the intruders and unwanted users are available and they possibly increasing the number of attacks and threats. If we talk about IoT users, then maximum of data sharing is also done on web and they have been increasing. Threat and Attacks can be come from different heterogeneous sources. Intruders can be internal intruders as well as external intruders. Internal Intruders are those have permission to access the internal network with Limitation or we can say some restrictions [1,8]. Maximum time to secure data access, security is applied in form of Users' authentication and data encryption. In order to detect the unknown user in network researcher can analysis the log details. The log details can fetch from TCP and HTTP Protocols. In this research work researcher's analysis intersections of different log files of IoT web server to detect anomaly user's IP address and HTTP requests.

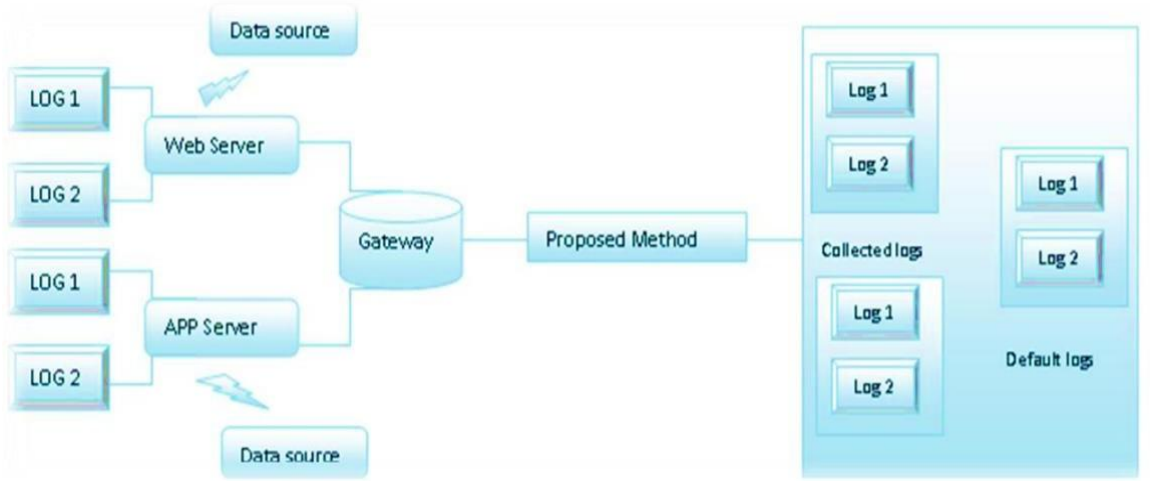


Fig.-1.1: Proposed Framework of IoT Log data analyses

Figure 1.1 Shows the complete framework of the process that researchers applied, the data is collected from web server because the web server is having the details of all logs. In order to achieve result, the complete log data analytics is performed with the proposed method.

1.1 Logging

Logging is the method of recording events/functions on a computer. Log messages in LINUX OS will store in the /var/log directory. For example:

Formats of log:

1. frame_number$frame_time$frame_len_src$length_dst$ip_srcip_dstip_protIDip_lentcp_len$tcp_srcport$tcp_dstport

Example_log:

202012:37:22.749684991IST$54$50:02:91:69:66:71$98:22:ef:d5:cc:2f$192.168.0.35$192.168.0.121$6$40$0$56521$80$56521 â†' 80 [ACK] Seq=1 Ack=1 Win=2144 Len=0"

2. Device_name/Track_ID/Module_Name/Content_Keyword/Time_Range

Example_log:

ABC02/22.34.45.106/Node1/LogLevel/22:30-23:00

Search condition	Description
Device name	Enter a device name. You can search for the logs of a device by specifying a device name.
Tracing ID	Enter a tracing ID to search for the logs of series modules.
Module name	Enter a module name to search for the logs that are generated by the module.
Content keyword	Enter a keyword to search for the logs that contain the keyword. Valid keywords include the values of the following parameters: **ProductKey**, **DeviceName**, **Code**, **LogLevel**, **Module**, **IotId**, **LogContent**, **TraceContext**, and **TraceId**.
Time range	Select a time range.

Syslog-ng is used in IoT and it contains all the log information about gateways and sensors. There are many benefits of using syslog-ng log file in an IoT environment. It delivers reliable log collection and simplifies the IoT system architecture. Using this log file users can easily maintain application logs and metrics of collected data. Parsing and filtering is also giving very high performance and decrease processing load on IoT Gateway.

1.1.1 syslog-ng:

Syslog-ng file has four important functions like collection of log data, preprocessing of log data, filtering of log data, and storing of log data.

Collection of log data: As researcher know, there are wide variety of heterogeneous data is available on network. The data from all sensors collected on central device like IoT gateways, that central device plays role as central log collector and it contains the file syslog-ng. As normal network, IoT network is also support and works on UDP, TCP and other encrypted protocols.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:03/Issue:01/January-2021 Impact Factor- 5.354 www.irjmets.com

Preprocessing of log data: Here, researcher can classify, filter and normalize structure or unstructured log messages with built-in parsers. User defined parsing is also available, researcher can simplify own parsing rules.

Filtering of log data: In this phase, log data is filtered in to required data. Researchers can fetch useful information from log file, that is used to detect anomaly user access, like wise IP addresses, protocol details etc.

Storing of log data: The log files stored locally on central syslog servers or researcher can use dynamic databases like SQL databases, HDFS, MongoDB, etc.

1.1.2 Types of Logs:

In multi sensor environment basically logs are defined into two categories like Time series and Non-time series logs.

Time series Log

This is very commonly used log. In this log information stored in log file with time stamp values, using this log researcher can easily identify at which time anomaly user's login to sensors or gateways. Also, it is available in very simple format, researcher don't need to clean this much and this is already available in structured format. some examples are.

a) Standard Event Log

Mar 08 22:12:44.117 2019 sjc04-b01-cvc02 SJC04-B01-CVC02: message repeated 13 times.

b) Multi-line Event Log

This is the example of java exception file messages:

```
016-08-18 00:00:12,759 ERROR [commonScheduleQueue.SchedulingThread]
com.ptc.wa.propertiesFile.GlassbeamPropertiesPlugin Administrator - Error writing Property files to storage
java.lang.NullPointerException: Local JMX URL string cannot be null here.
at wt.util.jmx.JmxConnectUtil.getLocalConnector(JmxConnectUtil.java:138)
at wt.util.jmx.JmxConnectUtil.getLocalConnector(JmxConnectUtil.java:98)
at wt.util.jmx.JmxConnectUtil.getDefaultServerManagerLocalConnector(JmxConnectUtil.java:86)
at com.ptc.wa.propertiesFile.WindchillPropertiesHelper.collectClusterProperties(WindchillPropertiesHelper.java:70)
at com.ptc.wa.propertiesFile.WindchillPropertiesPlugin.collect(WindchillPropertiesPlugin.java:89)
at com.ptc.wa.create.WAReportCreationManager.generateReport(WAReportCreationManager.java:226)
at com.ptc.wa.create.WAReportCreationManager.generateReportFromQueue(WAReportCreationManager.java:128)
```

Event data from a sensor (IoT)

In this scenario, two Air conditioners are in one room. To maintain this all one IoT application is there. The IoT application is calculating speed(rpm) & Flow of air of Air conditioner. The state of log data is following:

```
room_1,AC1Monitor,speed_rpm,245.2552546338,Fri mar 14 12:14:10 IST 2018,analog
room_1,AC2Monitor,flow_air,153.197123094758,Fri mar 14 22:19:07 IST 2018,analog
```

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:03/Issue:01/January-2021 Impact Factor- 5.354 www.irjmets.com

II. PROBLEM STATEMENTS

In a Multisensory scenario, an internetwork-based anomaly model would be developing to secure specific network environments from attackers and intruders. In this phase, researcher required some training data that were previously collected from old network traffic and can be detect anomalous behaviors of users or intruders. Then such model should be implemented and tested on real traffic. As in today's scenario researcher are having different network patterns.

Therefore, in this research paper researcher are addressing this problem of investigating to the evaluated old network traffic and comparing the results with the real traffic on proposed model.

The new challenge for the IoT log analysis is following:

1. Analysis of both non time series and time series logs.
2. Filter complex formats and define a structure format.
3. Detection of anomaly users and intruders.
4. Provide a secure multisensory environment.

III. RESEARCH METHODOLOGIES

In this research, following methodology is followed by researchers:

1. Unstructured heterogeneous data collected by different web sources log files, especially system log files and Gateway log files.
2. Unstructured data of log files is parsed on structured data.
3. Data is preprocessed; mainly removal of unused data and collection of useful data will be done in this phase.
4. The implementation is done in JAVA programming language as well as on MATLAB.
5. And the results in form of different number of IP addresses of intruders.

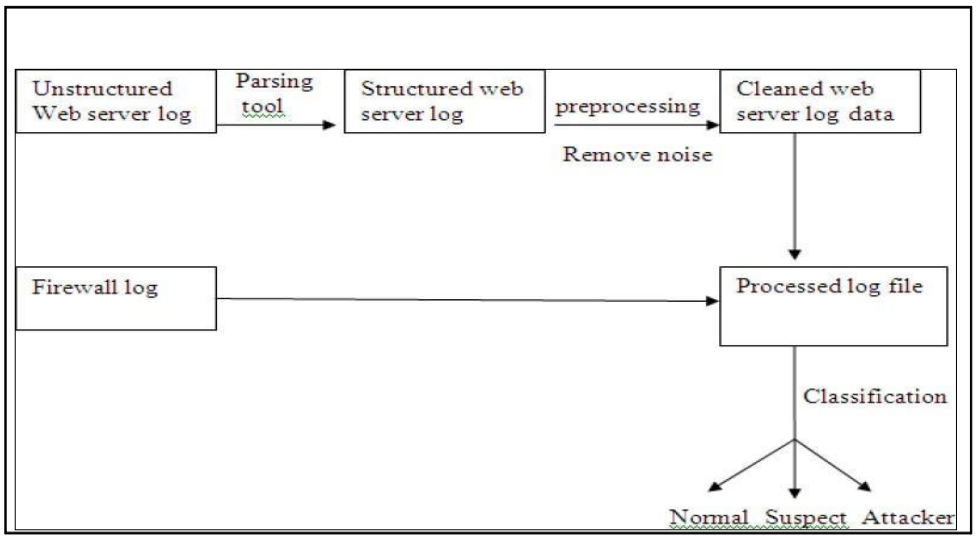
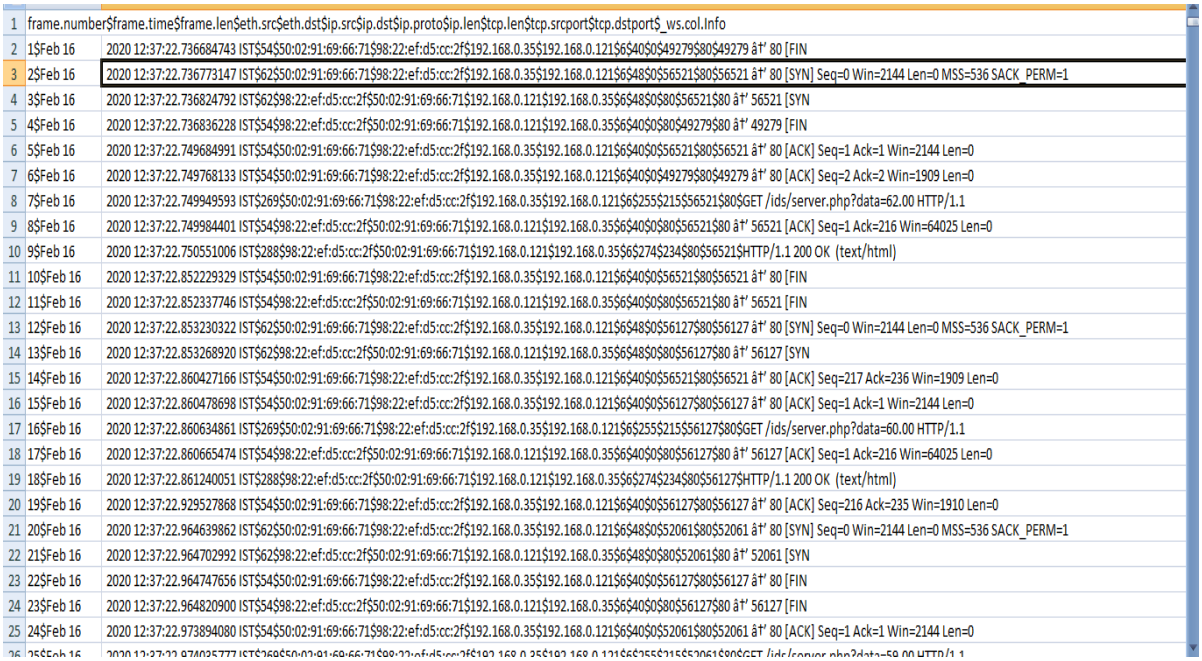


Fig.-3.1: Proposed Research Methodology

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:03/Issue:01/January-2021 Impact Factor- 5.354 www.irjmets.com



	frame.number$frame.time$frame.len$eth.src$eth.dst$ip.src$ip.dst$ip.proto$ip.len$tcp.len$tcp.srcport$tcp.dstport$_ws.col.Info
2	1$Feb 16 2020 12:37:22.736684743 IST$54$50:02:91:69:66:71$98:22:ef:d5:cc:2f$192.168.0.35$192.168.0.121$6$40$0$49279$80$49279 â†' 80 [FIN
3	2$Feb 16 2020 12:37:22.736773147 IST$62$50:02:91:69:66:71$98:22:ef:d5:cc:2f$192.168.0.35$192.168.0.121$6$48$0$56521$80$56521 â†' 80 [SYN] Seq=0 Win=2144 Len=0 MSS=536 SACK_PERM=1
4	3$Feb 16 2020 12:37:22.736824792 IST$62$98:22:ef:d5:cc:2f$50:02:91:69:66:71$192.168.0.121$192.168.0.35$6$48$0$80$56521$80 â†' 56521 [SYN
5	4$Feb 16 2020 12:37:22.736836228 IST$54$98:22:ef:d5:cc:2f$50:02:91:69:66:71$192.168.0.121$192.168.0.35$6$40$0$80$49279$80 â†' 49279 [FIN
6	5$Feb 16 2020 12:37:22.749684991 IST$54$50:02:91:69:66:71$98:22:ef:d5:cc:2f$192.168.0.35$192.168.0.121$6$40$0$56521$80$56521 â†' 80 [ACK] Seq=1 Ack=1 Win=2144 Len=0
7	6$Feb 16 2020 12:37:22.749768133 IST$54$50:02:91:69:66:71$98:22:ef:d5:cc:2f$192.168.0.35$192.168.0.121$6$40$0$49279$80$49279 â†' 80 [ACK] Seq=2 Ack=2 Win=1909 Len=0
8	7$Feb 16 2020 12:37:22.749949593 IST$269$50:02:91:69:66:71$98:22:ef:d5:cc:2f$192.168.0.35$192.168.0.121$6$255$215$56521$80$GET /ids/server.php?data=62.00 HTTP/1.1
9	8$Feb 16 2020 12:37:22.749984401 IST$54$98:22:ef:d5:cc:2f$50:02:91:69:66:71$192.168.0.121$192.168.0.35$6$40$0$80$56521$80 â†' 56521 [ACK] Seq=1 Ack=216 Win=64025 Len=0
10	9$Feb 16 2020 12:37:22.750551006 IST$288$98:22:ef:d5:cc:2f$50:02:91:69:66:71$192.168.0.121$192.168.0.35$6$274$234$80$56521$HTTP/1.1 200 OK (text/html)
11	10$Feb 16 2020 12:37:22.852229329 IST$54$50:02:91:69:66:71$98:22:ef:d5:cc:2f$192.168.0.35$192.168.0.121$6$40$0$56521$80$56521 â†' 80 [FIN
12	11$Feb 16 2020 12:37:22.852337746 IST$54$98:22:ef:d5:cc:2f$50:02:91:69:66:71$192.168.0.121$192.168.0.35$6$40$0$80$56521$80 â†' 56521 [FIN
13	12$Feb 16 2020 12:37:22.853230322 IST$62$50:02:91:69:66:71$98:22:ef:d5:cc:2f$192.168.0.35$192.168.0.121$6$48$0$56127$80$56127 â†' 80 [SYN] Seq=0 Win=2144 Len=0 MSS=536 SACK_PERM=1
14	13$Feb 16 2020 12:37:22.853268920 IST$62$98:22:ef:d5:cc:2f$50:02:91:69:66:71$192.168.0.121$192.168.0.35$6$48$0$80$56127$80 â†' 56127 [SYN
15	14$Feb 16 2020 12:37:22.860427166 IST$54$50:02:91:69:66:71$98:22:ef:d5:cc:2f$192.168.0.35$192.168.0.121$6$40$0$56521$80$56521 â†' 80 [ACK] Seq=217 Ack=236 Win=1909 Len=0
16	15$Feb 16 2020 12:37:22.860478698 IST$54$50:02:91:69:66:71$98:22:ef:d5:cc:2f$192.168.0.35$192.168.0.121$6$40$0$56127$80$56127 â†' 80 [ACK] Seq=1 Ack=1 Win=2144 Len=0
17	16$Feb 16 2020 12:37:22.860634861 IST$269$50:02:91:69:66:71$98:22:ef:d5:cc:2f$192.168.0.35$192.168.0.121$6$255$215$56127$80$GET /ids/server.php?data=60.00 HTTP/1.1
18	17$Feb 16 2020 12:37:22.860665474 IST$54$98:22:ef:d5:cc:2f$50:02:91:69:66:71$192.168.0.121$192.168.0.35$6$40$0$80$56127$80 â†' 56127 [ACK] Seq=1 Ack=216 Win=64025 Len=0
19	18$Feb 16 2020 12:37:22.861240051 IST$288$98:22:ef:d5:cc:2f$50:02:91:69:66:71$192.168.0.121$192.168.0.35$6$274$234$80$56127$HTTP/1.1 200 OK (text/html)
20	19$Feb 16 2020 12:37:22.929527868 IST$54$50:02:91:69:66:71$98:22:ef:d5:cc:2f$192.168.0.35$192.168.0.121$6$40$0$56127$80$56127 â†' 80 [ACK] Seq=216 Ack=235 Win=1910 Len=0
21	20$Feb 16 2020 12:37:22.964639862 IST$62$50:02:91:69:66:71$98:22:ef:d5:cc:2f$192.168.0.35$192.168.0.121$6$48$0$52061$80$52061 â†' 80 [SYN] Seq=0 Win=2144 Len=0 MSS=536 SACK_PERM=1
22	21$Feb 16 2020 12:37:22.964702992 IST$62$98:22:ef:d5:cc:2f$50:02:91:69:66:71$192.168.0.121$192.168.0.35$6$48$0$80$52061$80 â†' 52061 [SYN
23	22$Feb 16 2020 12:37:22.964747656 IST$54$50:02:91:69:66:71$98:22:ef:d5:cc:2f$192.168.0.35$192.168.0.121$6$40$0$56127$80$56127 â†' 80 [FIN
24	23$Feb 16 2020 12:37:22.964820900 IST$54$98:22:ef:d5:cc:2f$50:02:91:69:66:71$192.168.0.121$192.168.0.35$6$40$0$80$56127$80 â†' 56127 [FIN
25	24$Feb 16 2020 12:37:22.973894080 IST$54$50:02:91:69:66:71$98:22:ef:d5:cc:2f$192.168.0.35$192.168.0.121$6$40$0$52061$80$52061 â†' 80 [ACK] Seq=1 Ack=1 Win=2144 Len=0

Fig.-3.2: Raw data collected from multi sensors

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	frame.nur	frame.time	frame.len	eth.src	eth.dst	ip.src	ip.dst	ip.proto	ip.len	tcp.len	tcp.srcpor	tcp.dstpo	Value	normality		
2	1	1.23723E+14	54	8.8E+13	1.67E+14	192168035	1921680121	6	40	0	49279	80	-99	0		
3	2	1.23723E+14	62	8.8E+13	1.67E+14	192168035	1921680121	6	48	0	56521	80	-99	0		
4	3	1.23723E+14	62	1.67E+14	8.8E+13	1921680121	192168035	6	48	0	80	56521	-99	0		
5	4	1.23723E+14	54	1.67E+14	8.8E+13	1921680121	192168035	6	40	0	80	49279	-99	0		
6	5	1.23723E+14	54	8.8E+13	1.67E+14	192168035	1921680121	6	40	0	56521	80	-99	0		
7	6	1.23723E+14	54	8.8E+13	1.67E+14	192168035	1921680121	6	40	0	49279	80	-99	0		
8	7	1.23723E+14	269	8.8E+13	1.67E+14	192168035	1921680121	6	255	215	56521	80	62	1		
9	8	1.23723E+14	54	1.67E+14	8.8E+13	1921680121	192168035	6	40	0	80	56521	-99	0		
10	9	1.23723E+14	288	1.67E+14	8.8E+13	1921680121	192168035	6	274	234	80	56521	-99	0		
11	10	1.23723E+14	54	8.8E+13	1.67E+14	192168035	1921680121	6	40	0	56521	80	-99	0		
12	11	1.23723E+14	54	1.67E+14	8.8E+13	1921680121	192168035	6	40	0	80	56521	-99	0		
13	12	1.23723E+14	62	8.8E+13	1.67E+14	192168035	1921680121	6	48	0	56127	80	-99	0		
14	13	1.23723E+14	62	1.67E+14	8.8E+13	1921680121	192168035	6	48	0	80	56127	-99	0		
15	14	1.23723E+14	54	8.8E+13	1.67E+14	192168035	1921680121	6	40	0	56521	80	-99	0		
16	15	1.23723E+14	54	8.8E+13	1.67E+14	192168035	1921680121	6	40	0	56127	80	-99	0		
17	16	1.23723E+14	269	8.8E+13	1.67E+14	192168035	1921680121	6	255	215	56127	80	60	1		
18	17	1.23723E+14	54	1.67E+14	8.8E+13	1921680121	192168035	6	40	0	80	56127	-99	0		
19	18	1.23723E+14	288	1.67E+14	8.8E+13	1921680121	192168035	6	274	234	80	56127	-99	0		

Fig.-3.3: Cleaned data

IV. IMPLEMENTATION DETAILS

1. The log files collected from IoT Gateway and different sensors try to access in .logformat
2. The web server log and firewall log of both firewall server and desktop system are converted into structured format.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:03/Issue:01/January-2021 Impact Factor- 5.354 www.irjmets.com

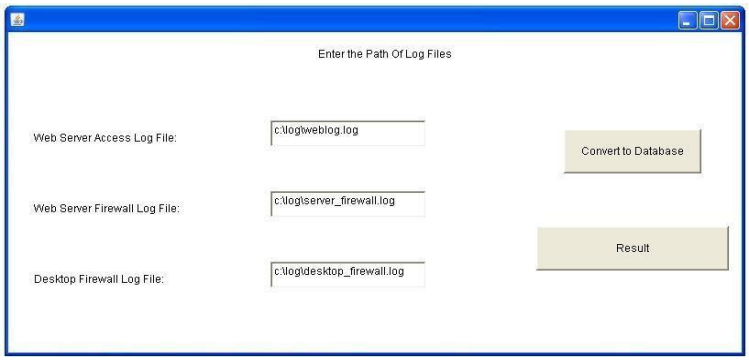


Fig.-4.1: Conversion of unstructured log files into structure format

3. Structured log is inserted into data base after removing noise then the intersection of log files is calculated.

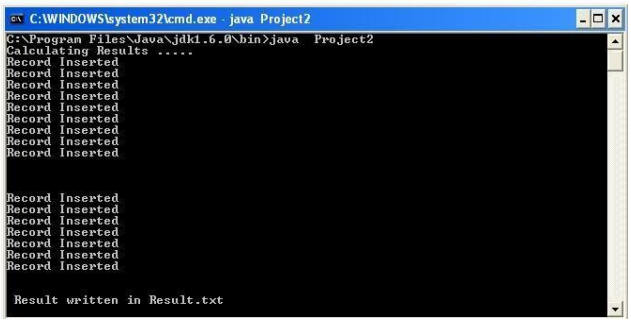


Fig.-4.2: Intersection of log data calculations

4. Intersection of log files is calculated. Finally the result is calculated and written intofile"Result.txt"

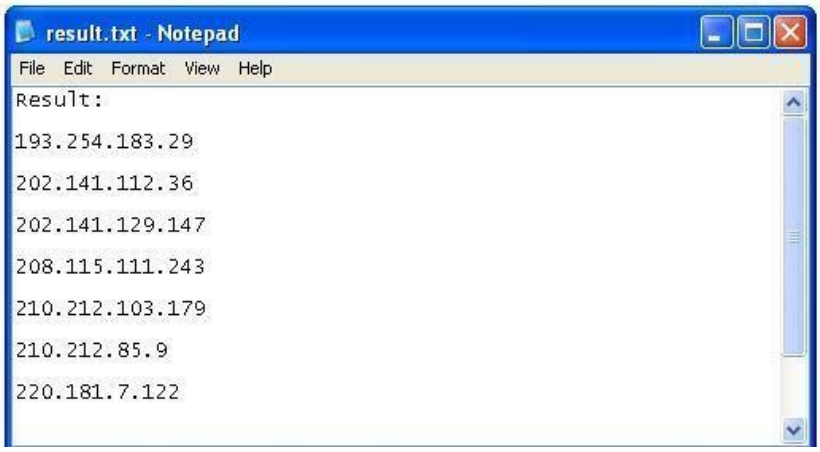


Fig.-4.3: Results: IP addresses of Intruders

5. The IP address of anomaly user is detected and is written in a file result.txt file. Researchers can further secure our system by blocking these IP addresses.

V. CONCLUSION

It has been concluded that IoT web server logs are collected by the actions and manners of sensors. By log files troubleshooting, security can be done. By analysis of different log files researcher can identify IP addresses of intruders, time stamp value and other details. The IoT data is collected from different log

files and through proposed method perform continuous analysis of log files; in this researcher can use data fusion as well. For that, researcher implemented a proposed framework to collect and analyses log files. As data is rapidly growing so to maintain those log files and to fetch needed data from complete file is a big task to researchers. On the future aspects researcher can use deep learning and data fusion also.

VI. REFERENCES

[1] Malviya, Mahesh, Abhinav Jain, and Nitesh Gupta. "Improving Security By Predicting Anomaly User Through Web Mining: A Review." International Journal of Advances in Engineering & Technology 1.2 (2011): 28.

[2] Saxena, Monika & Jha, Dr. (2019). A New Pattern Mining Algorithm for Analytics of Real Time Internet of Things Data. 10.35940/ijitee.A4506.119119.

[3] Saxena, Monika & Jha, C. (2019). Heterogeneous Big Smart data in Perspective of Smart Cities Development.

[4] Saxena, Monika & Jha, C. (2018). Analytics of IoT Streaming Data using Modified New Pattern Mining Algorithm.

[5] Saxena M., Saurabh P., Verma B. (2012) A New Hashing Scheme to Overcome the Problem of Overloading of Articles in Usenet. In: Wyld D., Zizka J., Nagamalai D. (eds) Advances in Computer Science, Engineering & Applications. Advances in Intelligent and Soft Computing, vol 166. Springer,

[6] Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-30157-5_96

[7] Saxena, M., Saurabh, P., & Verma, B. (2011). USENET: InternetNews Software, Security– Needs and Goals. International Journal of Scientific and Engineering Research, 2(3), 159-164.

[8] Saxena, D., & Saxena, M. (2011). Methods Used to Handle Overloading of Information in Usenet. International journal of scientific and engineering research, 49.

[9] Saxena M. (2020). Novel framework and service model for internet of things in context of smart cities. international research journal of modernization in engineering technology & Science. Volume 2 Issue 9 Pages 1730-1735 e-ISSN: 2582-5208.

[10] Lee, W., Stolfo, S.J., Chan, P.K. "Real Time Data Mining based Intrusion Detection." roc.Second DARPA Information Survivability Conference and Exposition, 2001, Retrieved from http://citeseer.ist.psu.edu/452795.html

[11] Tamas Abraham "Event Sequence Mining to Develop Profiles for Computer Forensic Investigation Purposes" Information Networks Division Defiance Science and Technology Organization, Australia.

[12] Buyer's guide for intrusion prevention systems (IPS).Retrieved June 3, 2004 from http://www.juniper.net/solutions/literature/buryer_guide/710005.pdf

[13] Kazimierz Kowalski, Mohsen Beheshti "Analysis of Log Files Intersections for Security Enhancement.

[14] Bhavani Thuraisingham, Latifur Khan, Mohammad M. Masud, Kevin W. Hamlen "Data Mining for Security Applications" 2008 IEEE/IFIP International Conference on Embedded and Ubiquitous Computing.

[15] Tom Dunigan, Greg Hinkel, "Intrusion Detection and Intrusion Prevention on a Large Network." A case Study Proceedings of the 1st Workshop on Intrusion Detection and Network Monitoring, April 9- 12, 1999, Santa Clara, California.

[16] Heady, R. Luger, G. Maccabe, A. Servilla, M. "The architecture of a network level intrusion detection system". Technical report. Computer Science Department, University of New Mexico, August1990.