International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-4, April 2019

Detecting and Analyzing the Malicious Linux Events using Filebeat and ELK Stack

J Bhuvanesh Babu, Srinivas Prasad, Gudapati Syam Prasad

Abstract: ***If we look at the current day scenario almost every individual and businesses are moving their ways of data storage from traditional ways (i.e., paper and files) to digital ways (i.e., cloud storages), which provides a platform to store and maintain data in an accurate, reliable and secure way. But, if the system is not configured securely it leads to data breaches and results in confidential data of an individual or a business being landed in the hands of bad guys, which results in huge financial and reputation loss and may even lead to life loss in some major cases. 0Although Linux is considered as the most secured operating system compared to other competition, but in recent times attackers started exploiting the vulnerabilities present in the Linux operating system and is becoming the next big target for the cyber criminals Now, the major provocation for any business or IT companies is to train its internal employees and maintain log analysis and monitoring domain, which is time consuming and requires expensive resources and knowledge. We have multiple log analysis commercial tools available in market which are expensive for small scale businesses and start-ups. So, in this paper I am going to propose a profitable way of implementing log monitoring and analysis infrastructure using open source tools like ELK stack and Moloch. ELK Stack is an open-source tool which is a combination of three open source tools Elasticsearch Logstash and Kibana which is used for monitoring and analyzing logs, here we are using ELK stack and Filebeat, Auditbeat which is light weight data shipper used to push Linux events to remote server, to build a profitable log monitoring and analysis infrastructure which can also be used for establishing a small scale Security Operations Center (SOC) services.***

Index Terms: ELK Stack, Filebeat, Auditbeat Moloch, Wireshark, Log Monitoring, Malware Analysis.

I. INTRODUCTION

Almost of the software companies and business organizations are using Linux operating system to setup a server for cloud and host their webpages, as it is an open source operating system and supports a lot of customization options as per their business requirements, although it has a lot of advantages compared to other operating systems available in the market, if any security misconfiguration occurs during the process of installation or configuration, it may lead to huge financial or reputational and in some cases it may lead to both along with harm to life. Once upon time Linux operating system is considered as secured one for the features it offered but now the hackers are exploiting the vulnerabilities present in those security features and compromising the systems so it is causing some sleepless nights to the users who relies on Linux as operating system for hosting their sites or cloud services. Linux operating systems are prone to various attacks like viruses, trojans, backdoors, ransomware, etc. The ransomware attack is considered as a serious concern as it completely encrypts the system and the data present in it, and makes it inaccessible for the users until he pays the demanded ransom amount which is also not guaranteed whether he decrypts after paying the amount. Apart from ransomware attacks other forms of online attacks like rootkits, trojans, malwares, etc., which are having various persistence methods are considered as severe threats to the Linux operating systems. As, there is threat posing from all kinds of malware attacks, there is a serious need for the IT companies and business organizations to establish a dedicated and effective security architecture to defend the current and upcoming cyber attacks by continuously monitoring the systems and the networks, in order to process that we need to monitor and analyze the critical logs to detect and eliminate the malicious activities on the systems and networks. So, in order to maintain a defensive mechanism for an organization here I am proposing a model which can be used to monitor and detect the Linux system and network logs by using ELK stack as log collecting, storing and visualizing tool and Filebeat and Auditbeat as log shipping tool and Moloch as packet capturing tools used to capture the data present in network packets and Wireshark for analyzing the collected packets and extracting the anomaly and then performing malware analysis on the extracted malicious file for build up a threat intelligence database to detect and block the future attacks of similar kind.

Logs:

Logs plays an important role in system administration as they consist of every minute detail about the actions performed by the person using the system or operations performed by the malicious files present in the system. Linux maintains a centralized log repository located at the following path **/var/log/**. These logs contain information about the server messages, kernel operations and all the services and operations running on the system. So, by analyzing these logs we can identify the way used by the attacker to gain access to the system and also the operations performed by the attacker after entering into the system.

Linux logs are classified into four types. They are:

1. *Service logs*
2. *System logs*
3. *Application logs*
4. *Event logs*

Manuscript published on 30 April 2019.
***** Correspondence Author (s)
J Bhuvanesh Babu, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, A.P, India.
Dr Srinivas Prasad, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, A.P, India.
Dr Gudapati Syam Prasad, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, A.P, India.

In Linux, the most critical logs paths that we should consider for monitoring are as follows.

1. */var/log/messages:* This is the first log to be considered by the system administrator if any malfunction occurs in the system, as it maintains a record of all the error messages related to booting which is not related to kernel, application service errors, and startup messages.
2. */var/log/auth.log:* This log should be considered when a security breach or unauthorized access takes place on a system, as it maintains a record of all the messages related to all successful and failed login attempts on a system.
3. */var/log/boot.log:* This log should be monitored when the system behaves in a strange manner like unexpected or uneven shutdown or restart occurs, as this log maintains a record of system down time and uptime.
4. */var/log/dmesg:* This log should be considered when there are any issues related to the hardware of the system i.e., let us consider there is a connectivity issue related to the network device like LAN card, by observing this log messages we can identify the issue for connectivity.
5. */var/log/mail.log:* This log helps in tracking all the incoming and outgoing mails from a system and failed delivery mails, which helps in tracing the spam mails generated from a system by the malicious software present in the system to perform mail bombing attack.
6. */var/log/httpd:* This log helps in identifying all the HTTP requests and responses generated from a system, which helps in tracing the malicious traffic generated by the malware present in the system and the IP address or the domain it is trying to communicate with and the file it is trying to download, generally this is helpful in case of backdoors where the backdoor file requests and downloads the main malware from a Command and Control (C&C) server.

7. */var/log/mysqld.log:* This log helps in identifying the queries made and the time taken to respond to that particular query, by monitoring these logs we can trace all the logins made to the server and also trace any unauthorized attempts made using various attack methods like SQL-injections.

Syslog-NG:

Syslog-NG is an open source tool used for collecting and sorting logs based on the operation which generated that particular log and store them at /var/log location. It is helpful in parsing the huge logs generated by the system into their own path and separate the which makes it easy for the identifying the process which generated that particular log. For example, let us consider that an organization is using a firewall, the logs generated by the firewall are forwarded to a dedicated server and at server side we use Syslog-NG to store these forwarded logs of firewall to save in a separate file which makes I easy to identify that those logs belong to firewall.

Moloch:

Moloch is also an open source tools which is used for collecting and parsing entire network packets and is used as plug-in for Elasticsearch, which it uses as database for storing the collected network packets. It also provides a web interface for viewing the captured packets.

Wireshark:

Wireshark is also a network packet capturing tool and it can also be used for extracting and merging the files present in packets, which can be helpful in collecting the malicious sample for malware analysis, we are using Moloch instead of Wireshark because if we run Wireshark continuously it will occupy the entire RAM capacity and the system will hang. So, in order to overcome this we are using Moloch.

II. Related Work

Here, we are performing log analysis using the open source tool called ELK stack for log storing, parsing and viewing and Filebeat for log data shipping and Moloch for packet capturing and Wireshark for extracting and merging the files in collected pcap's and then performing manual malware analysis on the collected samples.

III. COMPONENTS OF ELK STACK

ELK is an integration of three tools called Elasticsearch, Logstash and Kibana. ELK Stack is used as IT log management solution, Database, Search engine and many other operations for organizations who want to establish a Security Operations Center in a cost-effective way.

Elasticsearch:

Elasticsearch is an opensource tool written in Java and provides a HTTP web interface for searching the data stored in it. It can be used as database, search engine, analytics and uses Apache Lucene for querying.

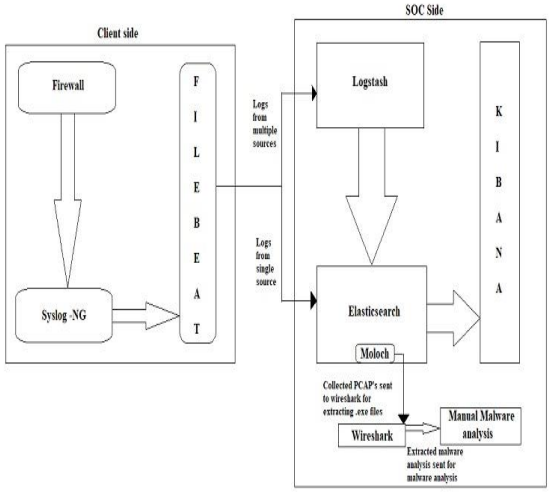
Logstash:

Logstash is used for parsing the huge unorganized data coming from beats installed on different source systems into organized manner and acts as pipeline between source and Elasticsearch to transfer data

Kibana:

Kibana is just a web interface used to browse through the data stored in Elasticsearch and analyze the logs present in it and search the data present in it using Lucene query syntax.

IV. WORKFLOW



Above workflow describes the monitoring of the logs generated in Linux system and network logs by configuring firewall and syslog-ng along with filebeat.

As observed in the above workflow firewall is configured to forward all the collected network logs to the syslog-ng which is being operated on port 514.

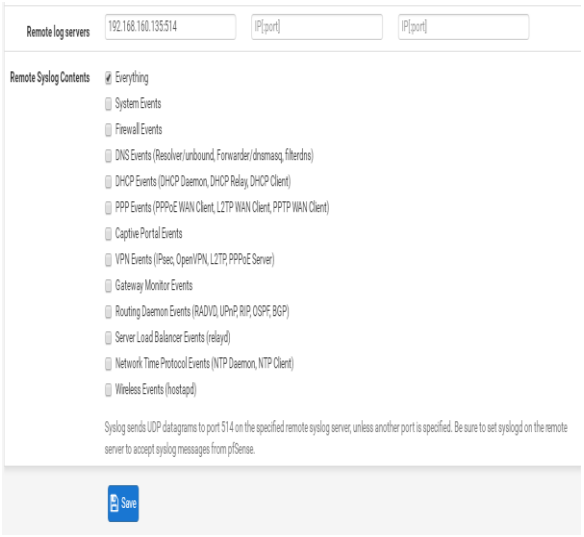


Figure: pfSense configuration page

Syslog-ng is then configured to collect all the firewall forwarded logs and saves them at "/var/log/<filename>.log" as configured in the syslog-ng configuration file.

```
 GNU nano 2.5.3            File: syslog-ng.conf

@version: 3.5
@include "scl.conf"
@include "`scl-root`/system/tty10.conf"
    options {
        time-reap(30);
        mark-freq(10);
        keep-hostname(yes);
        };
    source s_local { system(); internal(); };
    source s_network { udp(ip(192.168.160.135) port(514)); };

    destination d_local {
    file("/var/log/syslog-ng/messages_${HOST}"); };
    destination d_logs {
        file(
            "/var/log/pfsense.log"
            owner("root")
            group("root")
            perm(0777)
```

Figure: Syslog-ng custom configuration

All the collected logs are then shipped to the Elasticsearch or Logstash using filebeat which is configured to forward all the collected data to the dedicated Elasticsearch server directly or through Logstash to Elasticsearch server and also we can configure filebeat for which log file to be forwarded.

```
#========================== Filebeat prospectors =============================

filebeat.prospectors:

# Each - is a prospector. Most options can be set at the prospector level, so
# you can use different prospectors for various configurations.
# Below are the prospector specific configurations.

- type: log

  # Change to true to enable this prospector configuration.
  enabled: true

  # Paths that should be crawled and fetched. Glob based paths.
  paths:
    - /var/log/*.log
```

```
#------------------------- Elasticsearch output ------------------------------
output.elasticsearch:
  # Array of hosts to connect to.
  hosts: 192.168.160.135:9200

  # Optional protocol and basic auth credentials.
  #protocol: "https"
  #username: "elastic"
  #password: "changeme"

#----------------------------- Logstash output --------------------------------
#output.logstash:
  # The Logstash hosts
  #hosts: ["192.168.160.135:5044"]

  # Optional SSL. By default is off.
  # List of root certificates for HTTPS server verifications
  #ssl.certificate_authorities: ["/etc/pki/root/ca.pem"]
```

Figure: Filebeat configuration file

Elasticsearch operates on port 9200, Logstash on port 5044 and kibana on port 5601for these service to operate correctly we have to configure all the configurations of the three tools correctly and use 0.0.0.0 which is a non-loopback IP for network address of elasticsearch and kibana so that even if the IP address of the server changes it automatically binds it to the server.


```
{
  "name" : "Bhuvanesh",
  "cluster_name" : "Bhuvanesh",
  "cluster_uuid" : "8Cb35JqYQBGxVfVkKOcsBA",
  "version" : {
    "number" : "6.2.4",
    "build_hash" : "ccec39f",
    "build_date" : "2018-04-12T20:37:28.497551Z",
    "build_snapshot" : false,
    "lucene_version" : "7.2.1",
    "minimum_wire_compatibility_version" : "5.6.0",
    "minimum_index_compatibility_version" : "5.0.0"
  },
  "tagline" : "You Know, for Search"
}
```


Figure:Elasticsearch page

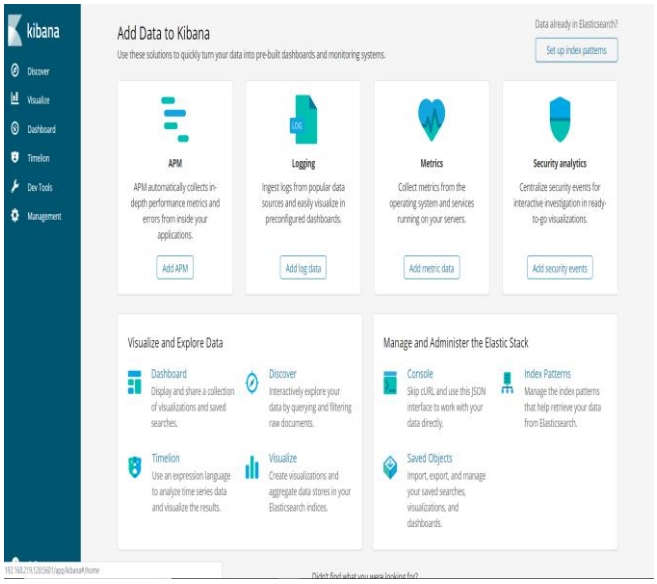


Figure: Kibana Homepage

Using kibana the logs can be monitored as shown below and it contains the details about the reason for which the log has been generated.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

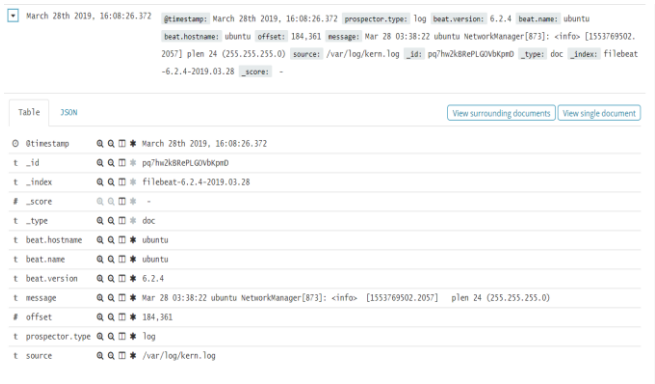
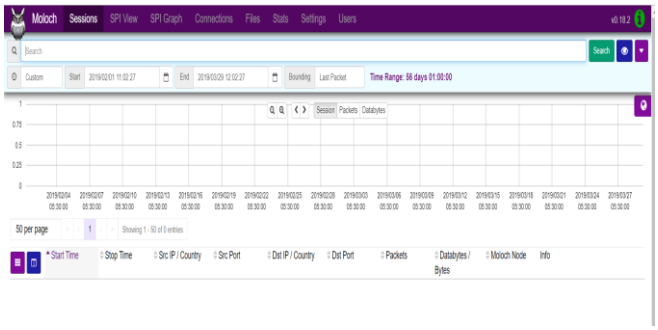


Figure: Log monitoring using kibana

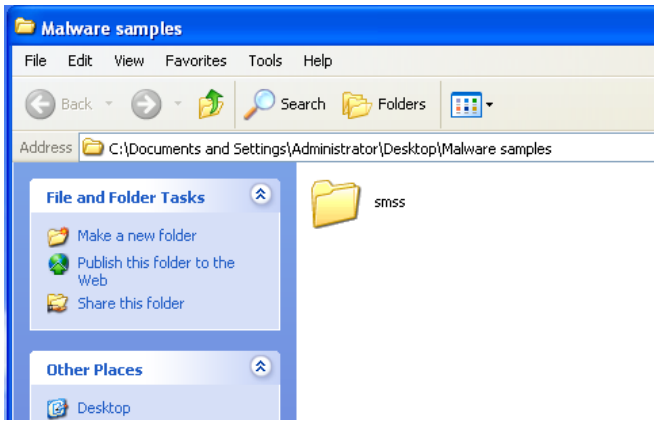
Moloch is installed on the Elasticsearch server which is used to collect all the network packets being transmitted on the server side and store them in Elasticsearch server as Elasticsearch also acts as a database and the Moloch home page appears as shown below.



By clicking on the right top corner drop box we can select various options related to PCAP's and the collected PCAP's can be downloaded from here which can be used for extracting malicious files (if any) present in the network packets.

The collected sample is then analysed in a controlled and isolated environment to determine its behaviour and the actions performed by it on the infected system.

For example if we observe the below figure, we can find a folder named "smss" but when we unhide the extensions in folder options we can find .exe is extension linked to it, which means that it is a malicious file which appears to be a folder and if the user clicks on it considering it as a folder the malware gets executed and starts its operation on the system



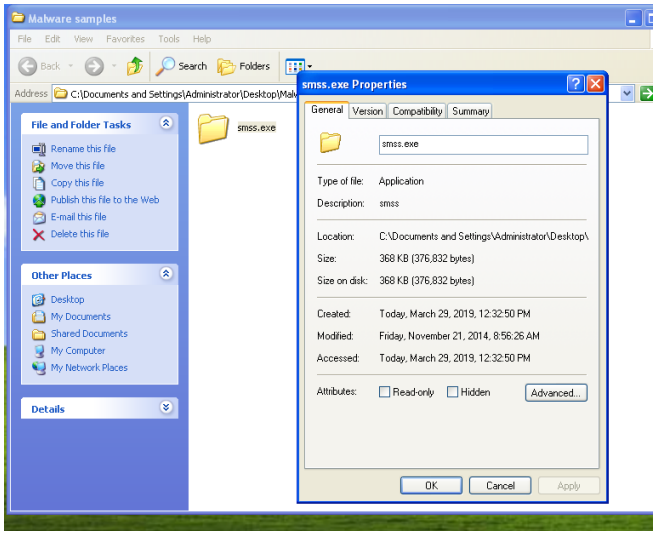


Figure: collected malware sample

Also if we observe the properties of the file it is showing as Application which means that when executed it infects the system by running as a legitimate service since smss is a legitimate service which comes built-in with the windows system. In order to further confirm whether it is malicious or not we can use HashCalc to calculate the hash and submit it in virustotal.com website which gives score from almost 70 antivirus engines by cross checking the hash with their data base.

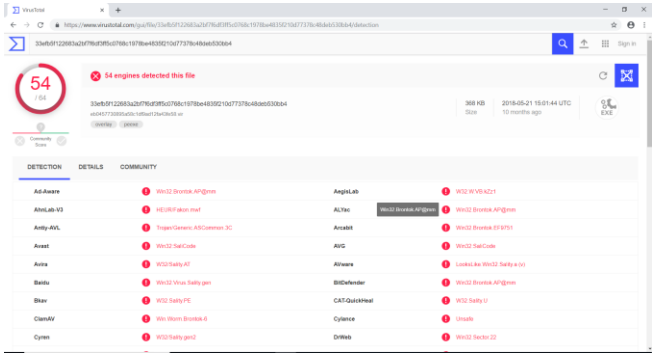


Figure: Virus Total file details of a file uploaded

We can further perform various types of malware analysis on the sample and determine its behaviour and generate a signature and use it in future.

V. FUTURE SCOPE

In this proposed work we collected logs and data network packets and collected signature of the samples. By collected all the signatures of the available samples and the samples from wild we can create a threat intelligent database which can be used to detect the anomaly data being transmitted by integrating Artificial Intelligence we can feed all the anomaly signatures to this database. s files.

VI. CONCLUSION

By using this model we can analyze the Linux events by configuring syslog-ng, filebeat and ELK stack and detect the malicious activities being performed on the system.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

REFERENCES

1. Ibrahim Yahya Mohammed AL-Mahbashi, Dr. M. B. Potdar, Mr. Prashant Chauhan. (2017) "Network Security Enhancement through Effective Log Analysis Using ELK" International Conference on Computing Methodologies and Communication (ICCMC), 978-1-5090-4890-8/17
2. Kwon, "Performance of ELK Stack and Commercial System in Security Log Analysis"
3. Online "Open Source Search & Analytics" elastic
4. Online "SwiftOnSecurity/sysmon-config" github
5. Online "AutorunsToWinEventLog" github
6. Online "Sysmon - Windows Sysinternals" docs.microsoft
7. María del Carmen Prudente Tixteco, Lidia Prudente Tixteco, Gabriel Sánchez Pérez, Linda Karina Toscano. "Intrusion Detection Using Indicators of Compromise Based on Best Practices and Windows Event Logs" International Conference on Internet Monitoring and Protection (ICIMP 2016), 978-1-61208-475-6
8. Online "Virus Total"
9. Onilne "ELK Stack: Elasticsearch, Logstash, Kibana" elastic
10. Online "Winlogbeat: Analyze Windows Event Logs" elastic

AUTHORS PROFILE

Dr. G. Syam Prasad is Currently working as Professor in CSE at KLEF(Deemed to be University), Vaddeshwaram, Guntur . Andhra Pradesh, India.He received the B.Tech and M.Tech degrees from the Department of computer Science and Engineering , Acharya Nagarjuna University, Guntur, India in 1999 and 2004 respectively, and Ph.D. degree from the Department of Computer Science and Systems Engineering , Andhra University at Visakhapatnam, India , in 2015. His research interests include network Security, cryptography, security and privacy, image processing, Data Mining, compilers and algorithms.