International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-10 Issue-1, November 2020

File Log Server and Application in Server System Privacy

Phan Thi Ha, Trinh Thi Van Anh

Abstract. The article research on file log server and application off file log into operation as well as server system privacy. From that, authorities had conducted log analysis system of which ability of updating data according to real time, helping the log serve analysing find to be easier and more intuitive than Kibana, system helps server privacy be easier when detecting errors and warning immidiately, recognizing abnormal signatures help administrator to receive recommendations in the earliest and most precise way. Comparing to handmade or following to troubles, this is absolutle an optimal choice.

I. INTRODUCTION

A. Overview

Nowadays, the Internet thrives strong and there are many moves that have surpassed a positive contribution to economics, society and especially human, The Internet development followed the online website created more and more and more and more of course the web servers were rapidly increasing to the need to create website and also create more profitable environment for other players with other players. The safety of the web servers has always been a very difficult problem. According to the annual security report of year 2019 from the Bulletproof, A DoS attacks or DDoS can cause damage to a company over two million dollars or 120.000 for a small company[1]. Or the DDoS attacks on Wikipedia-- the top encyclopedia website-- makes this website stop for one day[4]. In 2/2019, the Bkav Corporation has issued a warning of a massive attack of foreign hackerson public servers in Vietnam. Hundreds of organization in Vietnam were attacked by hackers, hacked into server, then made a complete encryption of all data on the server[3]. When being attacked by pirates, the server will be affected a lot that can lead to the stagnant operation of the system or loss of data. This will not only cost the time to heal the system but also lead to the economy, or the security. That's why we need an early hit against assaulting to server. To quickliest prevent attacks to the server from a hacker, the administrator needs to know soon the potential dangers which can affect their system. The log files analysing will help the administrator know what's happening to the server and produce rapidly processes [2], this will not only help to prevent but also to know how to attack faster than normal, It also helps minimize human resources in term of managing the server. Log file of the server: It's a simple document that contains all the activity of a specific server in a certain time (for instance: a day).

The Log server is automatically created, maintain and can provide for the administrator's detailed look of the way, the web page time or the application installed on the server. However, not exactly all file are same structured , with each operating system and apps, there are specific identities. Log file will provide the administrator with all of the server's active information, supporting the problems that the server has suffered, as long as administrator know how to analyze, use the received information in order to reconstruct some of the popular server types today:Common Log Format, Common Event Format , JSON Log Format, W3C Extended Log Format

The log file analysis system: normally, the administrator will proceed seasonally or when the system fails, along with the simple analysis as usual, the amount of log data can be analyzed is very small compared to the amount of log generated every day.

This creates many problems like whether the data analyzing log reflect the entire problems of the system or not. The time to be able to process all log data, log data data analyse too many extra details that waste time analyzing,… System of log analysis will develop in order to fullfill existing loopholes by using using principle of gathering, assess the log recorded and point out specify events that can represent for problems or threats.

Besides, some compulsive features of log file analyzing system bring to the administrator can be told as: Storing focusing data, checking system and warning, the ability to display data, enhance the ability to analyze data. With a lot of data supporting to the log data management, the log analysis system is the option of most recent companies and the tendency of small companies.

In order to solve the log server's accounts and automatically alert when there's been a lot of technological solution, some solutions that include the fee paying and open sources that can be mentioned as well such as Splunk[8], Graylog[9] or ELK stack[10]. Outstanding features of monitoring solution and log date analysing contmporary time have to mentioned are strongly looking, conducting monitor according to real time, report, limitation warning, history date analyzing, tracking,… In spite of that, after considering, authorities decide to choose ELK stack[10] technic platform as a solution for problems relating to log data controlling and analysis carried out by ELK focusing some strengths like:

Revised Manuscript Received on November 30, 2020.
***** Correspondence Author

Phan Thi Ha*, PTIT-Department Of Information Technology Posts and Telecommunications Institute of Technology Hanoi, Vietnam.
Trinh Thi Van Anh, FPT-Department Of Information Technology Posts and Telecommunications Institute of Technology Hanoi, Vietnam.

Open source code saving system conducting fee which can be gathered log from many sources; support to multi-platform, multi-format log which has platform being able to strongly search(ElasticSearch), assits to data figure gathered (Analytic), managing focusing log, searching and annoucing errors automatically. Basically, ELK contain 3 parts:

Elastic Search[5]: Basically, this is a tool for searching and analyse text. Elastic search take unstructure data from many other source and store it at high-optimized format for searching relying on language. ElasticSearch performs the data in the form of JSON structured-documents. Users can use ElasticSearch through API RESTful of developers or PHP, Python and Ruby languages for storing, searching, analyzing huge block of data quickly and efficiently. It is specially useful in processing data which is natural language.

Logstash[6]: LogStash is software gathering source code data written on Java platform with the ability to obtain real time data. LogStash have function to obtain log from many different sources then reshape it and send it to other data base. Besides, LogStash is used to filtered serving for analysis problem and visualization data

Kibana[7]: Kibana is a source code software used to make data abstract, conducting charts, reports, monitors controlling and analyzing real time data from data source in ElasticSearch. Kibana provide display allowing users to execute the query on the whole system of retrieving information called ElasticSearch, conducting charts, control monitors from indexing data on ElasticSearch quickly.

In scope of this article, the authorities focus on researching on file log serve, through which we can build abnormal detecting system and warning the manager in order to have solutions against serve assaulting. The article also research on data in file log of a web server, from which we can find the problem affecting this server.

With that purpose, content of article can be split into following 3 parts:

Log 1: Overall about file log server; **Log 2**: Researching and designing system of server analysis; **Log 3:** Applying experiment of practical receiving file log server.

II. CONDUCTING MODEL AND CONFIGURATION OF THE LOG SERVER ANALYSIS SYSTEM

2.1 Building file log server analyzing models

Relying on the structure of ELK stack, in this article, authorities conduct in detail a log analysis system Figure 2.1, according to approach of optimizing log obtaining and announcing to the administrators, limiting in the best way for overlapping cause heavy load on the system. Data from log server is extracted by Filebeat and send to LogStash, then LogStash filter the log contend and put it in the log marking on ElasticSearch or take it to mail server if there are any errors. Data is indexed in Elastic Search is displayed on Kibana for administrators to monitor:

Log server text is extracted by Filebeat and send to LogStash in frequency configured previously. Filebeat operates as a service helping to reduce the work load for CPU but still ensure the data is transfer in the fastest way. In here, Filebeat play a role as service taking log server from files located and take to LogStash for processing.

Data after being transferred by FileBeat to LogStash will be taken into “Filter Plugin”. In here, data will be standardized and filtered as orders from administrators in order to output the essential data. Output for this Filter stage is JSON documents contain contend of log message.

Log server after being filtered will be taken to ElasticSearch or put in the mail servers if there are any errors

-JSON data when being taken to ElasticSearch will be log marked which serve for searching problems and visualizing data, conducting reports and analyzing log data.

-Data after being log marked in ElasticSearch will be visualized, conduct conducting reports and control monitors on Kibana.

Figure[2.1] illustrate in detail log analysis system to which authorities build according to orientation of taking log data as well as announcement to administrators , limit in the best way to the overlapping causing heavy load on the system.

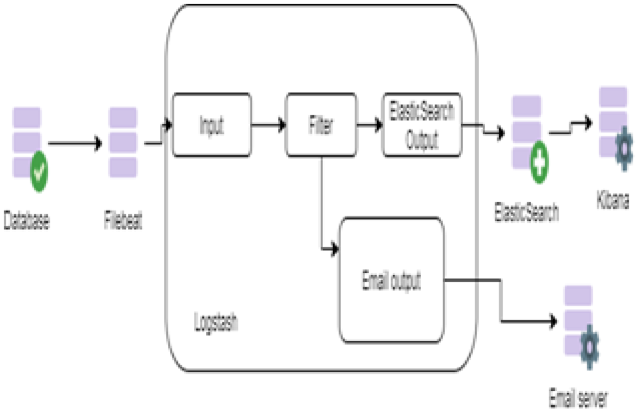


Figure 2.1. Log analysis system

2.2 Configuration of log analysis system by ELK stack

File beat have duty to transfer log data from files containing log to LogStash. In here, FIlebeat will receive all the file log lying on the file and push it to port 5044 for LogStash to receive the log data, thus, first work to do is configuring the Filebeat setting.

Then, setting configuration of LogStash when running LogStash acquire one config file for operate instruction. Thus, it needs a config file contain all input, filter and output.

After establishing LogStash setting, data is transferred to indexing at ElasticSearch and then display through Kibana. With basic model, basic available settings in ElasticSearch is enough to meet the need for the log analysis on one server. In order to display data indexed in ElasticSearch, authorities configure port running both Kibana and ElasticSearch.

III. ANALYZING DATA OBTAINED FROM LOG SERVER

From received data after setting ELK stack log server analysis system, authorities can rely on these data to check whether the system face to dilemmas or not.

Analyzing DDOS assaulting from log data

DDoS is attacking method. Though being familiar to most of us but it is difficult to completely prevent. Normally, assaulting will connect server to revolting order, thus, monitoring 404 response is able to benefit the DDoS assaulting analysis. Therefore, authorities will create chart checking response from client server.

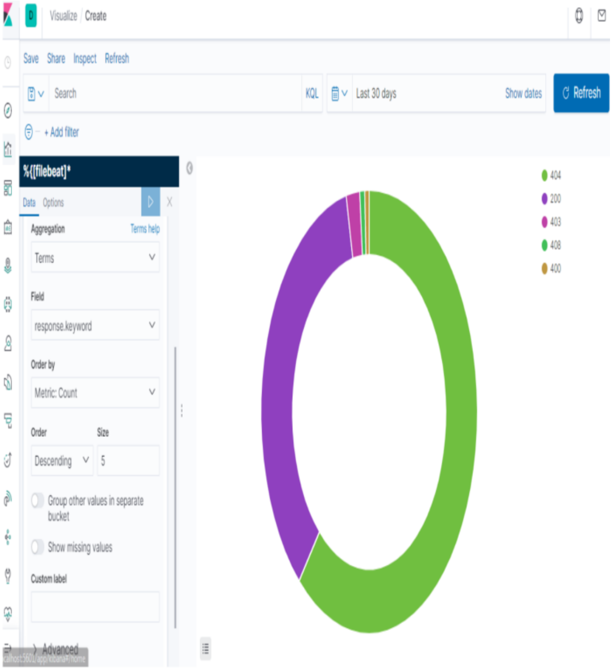


Figure 3.1. Client server response chart

The chart rely on log data as the chart Figure 3.1 to reveal the number of request come from server, 404 responses have the much more many compare to the other response. This can be a loophole when there are too many accesses to server but to a page that is not available on the system. Administrators need to utilize which time system facing up to increasing number of 404 responses.

Through chart Figure 3.2filtering 404 as the time fly, we can easily recognize that on March 30th, the number of 404 response abnormally increase, continue to click on the 30th to have more information about this time, Figure 3.3.

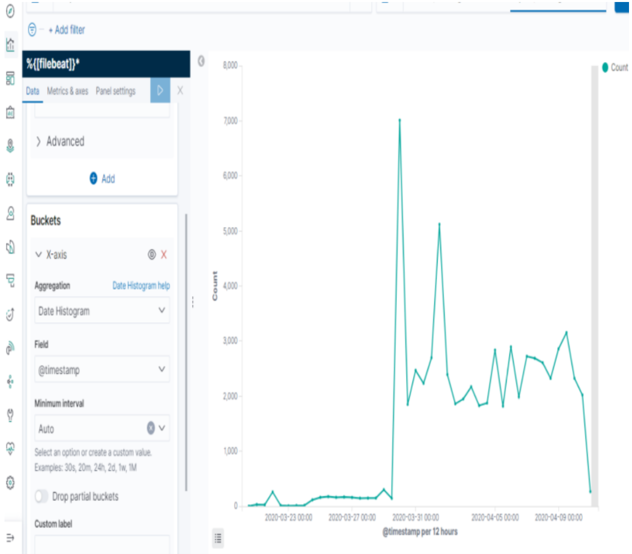


Figure 3.2. 404 response from server

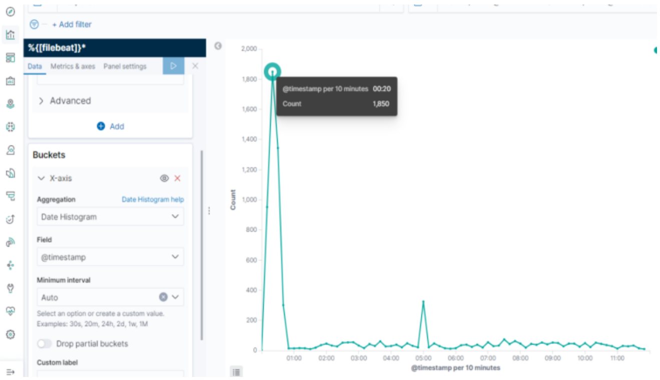


Figure3.3: Detail of number of 404 responses in March 30th

Then, filter the IP sending many request to server in the checking point of time.

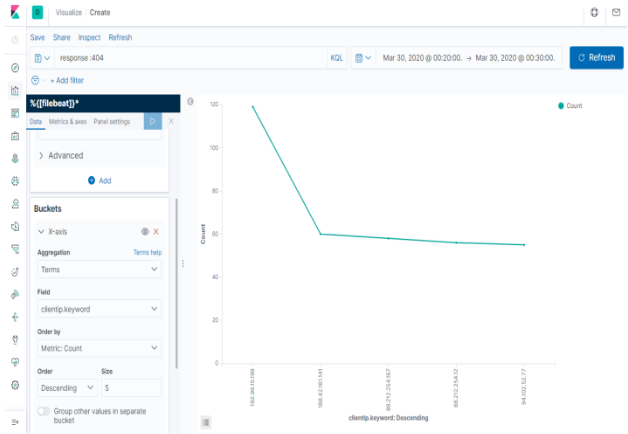
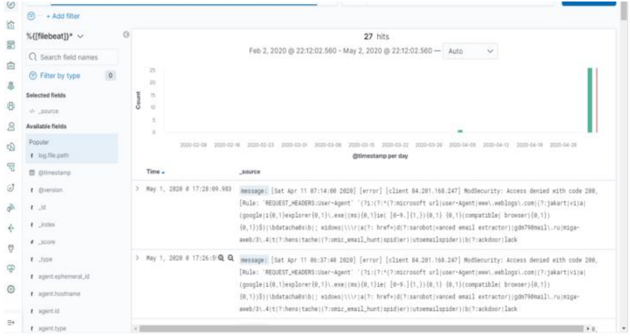

Visualize / Create
Save Share Inspect Refresh
response :404
KQL
Mar 30, 2020 @ 00:20:00. → Mar 30, 2020 @ 00:30:00.
Refresh
+ Add filter
%{[filebeat]}*
Data Metrics & axes Panel settings
Advanced
Add
Buckets
X-axis
Aggregation: Terms
Field: clientip.keyword
Order by: Metric: Count
Order: Descending Size: 5
Group other values in separate bucket
Count
120
100
80
60
40
20
0
192.99.15.199
clientip.keyword: Descending

Figure 3.4: StatisticalIPs receiving 404 responses highly on March 30th

Through the above chart ,**Figure 3.4**, we can see that IP 192.99.15.199 haves highest 404 responses. In detail, in 10 minutes it had sent 120 requests to server, thus, this IP can be put to the doubtful list and prevent it if high consecutive times do not stop.

-Detecting XSS attacking with log analysis system

XSS assaulting basically can be detected by finding cards like script, frame,.. Through lag data, search the possible card inserted with toxic code.Figure 3.5.



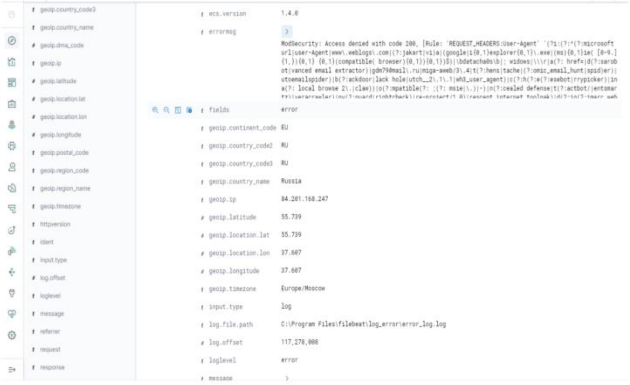
```
May 1, 2020 @ 17:28:09.983  message: [Sat Apr 11 07:14:00 2020] [error] [client 84.201.168.247] ModSecurity: Access denied with code 200,
[Rule: 'REQUEST_HEADERS:User-Agent' '(?i:(?:^(?:microsoft url|user-Agent|www\.weblogs\.com|(?:jakart|vi)a|
(google|i{0,1}explorer{0,1}\.exe|(ms){0,1}ie( [0-9.]{1,}){0,1} {0,1}(compatible( browser){0,1})
{0,1})$)|\bdatacha0s\b|; widows|\\\r|a(?: href=|d(?:sarobot|vanced email extractor)|gdm79@mail\.ru|miga-
aweb/3\.4|t(?:hens|tache|(?:omic_email_hunt|spid)er)|utoemailspider)|b(?:ackdoor|lack

May 1, 2020 @ 17:26:5  message: [Sat Apr 11 06:37:40 2020] [error] [client 84.201.168.247] ModSecurity: Access denied with code 200,
[Rule: 'REQUEST_HEADERS:User-Agent' '(?i:(?:^(?:microsoft url|user-Agent|www\.weblogs\.com|(?:jakart|vi)a|
(google|i{0,1}explorer{0,1}\.exe|(ms){0,1}ie( [0-9.]{1,}){0,1} {0,1}(compatible( browser){0,1})
{0,1})$)|\bdatacha0s\b|; widows|\\\r|a(?: href=|d(?:sarobot|vanced email extractor)|gdm79@mail\.ru|miga-
aweb/3\.4|t(?:hens|tache|(?:omic_email_hunt|spid)er)|utoemailspider)|b(?:ackdoor|lack
```

Figure 3.5: Searching script in message

Through searching, we can realize that a certain number of script are inserted on server, however, they are prevented by Firewall ModSecurity. Click on administrators can see the detail of these script data,Figure 3.6.

Retrieval Number: 100.1/ijitee.A81111110120
DOI: 10.35940/ijitee.A8111.1110120
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication



```
ecs.version              1.4.0
errormsg                 ›
                         ModSecurity: Access denied with code 200, [Rule: 'REQUEST_HEADERS:User-Agent' '(?i:(?:^(?:microsoft
                         url|user-Agent|www\.weblogs\.com|(?:jakart|vi)a|(google|i{0,1}explorer{0,1}\.exe|(ms){0,1}ie( [0-9.]
                         {1,}){0,1} {0,1}(compatible( browser){0,1}){0,1})$)|\bdatacha0s\b|; widows|\\\r|a(?: href=|d(?:sarob
                         ot|vanced email extractor)|gdm79@mail\.ru|miga-aweb/3\.4|t(?:hens|tache|(?:omic_email_hunt|spid)er)|
                         utoemailspider)|b(?:ackdoor|lack hole|utch__2\.1\.1|wh3_user_agent)|c(?:h(?:e(?:esebot|rrypicker)|in
                         a(?: local browse 2\.|claw))|o(?:mpatible(?: ;(?: msie|\.)|-)|n(?:cealed defense|t(?:actbot/|entsmar
fields                   error
geoip.continent_code     EU
geoip.country_code2      RU
geoip.country_code3      RU
geoip.country_name       Russia
geoip.ip                 84.201.168.247
geoip.latitude           55.739
geoip.location.lat       55.739
geoip.location.lon       37.607
geoip.longitude          37.607
geoip.timezone           Europe/Moscow
input.type               log
log.file.path            C:\Program Files\filebeat\log_error\error_log.log
log.offset               117,278,008
loglevel                 error
message                  ›
```

Figure 3.6: Message detailed content containing script codes

Rely on information obtained, this can be a script inserted heading to crawl data on server. With script crawl data like that can lead to the issue that consuming resource of system or websites do not operate as beginning plan. However, script is prevented by the firewall and this does not affect on the system.

IV. CONCLUSION

Authorities have conducted log analysis system which have ability to update data as real time, helps the log server analysis become much easier and more intuitive t Kibana. System help the server privacy become easier when having errors and issue warning immediately about abnormal signals. This help the administrators can receive fastest error reports. Comparing to the manual or following the trouble, this can be considered as absolutely appropriate choice. Analyzing the log data have important meaning in server privacy, but only one particular log analysis system relying on ELK stack are not sufficient to ensure the server privacy. Combining different monitoring network tools, monitors the system to elevate navigating ability, prevents basic loopholes or announce errors to administrators, which is extremely vital. Firewall, Antivirus tools still are paramount software in term of protection of server system privacy. Those software help system prevent from beginning the detrimental causes help to reduce the workload which is needed for administrator's solving. This means that server system need to update firewall and antivirus software fastest and soonest.

REFERENCES

1. BulletProof (2019), *BulletProof Annual Cyber Security Report 2019*
2. https://www.bulletproof.co.uk/industry-reports/2019.pdf
3. Karen Kent, Murugiah Souppaya (2006), *Guide to Computer Security Log Management*, National Institute of Standards and Technology, Gaithersburg.
4. Kim Thanh (2019), *Đang có một chiến dịch tấn công có chủ đích nhằm vào các Server Public của Việt Nam*
5. https://www.sggp.org.vn/dang-co-mot-chien-dich-tan-cong-co-chu-dich-nham-vao-cac-server-public-cua-viet-nam-575735.html
6. Kiến Văn (2019), *Wikipedia xác nhận sự cố ngừng hoạt động do bị tấn công DdoS*
7. https://thanhnien.vn/cong-nghe/wikipedia-xac-nhan-su-co-ngung-hoat-dong-do-bi-tan-cong-ddos-1123957.html
8. https://www.elastic.co/what-is/elasticsearch (10/02/2020)
9. https://www.elastic.co/logstash (14/02/2020)
10. https://www.elastic.co/what-is/kibana (20/02/2020)
11. https://www.splunk.com/ (15/01/2020)
12. https://www.graylog.org/ (15/01/2020)
13. https://www.elastic.co/what-is/elk-stack (11.16/01/2020)
14. https://www.freeformatter.com/json-formatter.html

Dr. Phan Thi Ha is currently a lecturer atthe PTIT (Faculty of Information Technology at Posts and Telecommunications Institute of Technology in Vietnam) and FPT University(Computing Fundamental Department,FPT UniversityHanoi 10000), Vietnam as well. She received a B.Sc.in Math & Informatics, a M.Sc. in Mathematic Guarantee for Computer Systems and a PhD. in Information Systems in 1994, 2000 and 2013, respectively. Her research interests include machine learning, natural language processing and mathematics applications.

Email: hapt@ptit.edu.vn and ***hapt27@fe.edu.vn***

ThS. Trint Thi Van is currently a lecturer and a researcher at the Faculty of Information Technology at Posts and Telecommunications Institute of Technology in Vietnam and
Computing Fundamental Department, FPT University, Hanoi
10000, Vietnam as well. She received a B.Sc. in ElectronicsTelecommunications in 1993 (HUST), M.Sc. in Computer Science in 1998 (HUST). Her research interests include machine learning, NLP and opinion mining.

Email: vanh22@yahoo.comandanhttv20@fe.edu.vn