Performance of ELK Stack and Commercial System in Security Log Analysis

Sung Jun Son
Igloo Security
Daejeon, South Korea
ssj3821@gmail.com

Youngmi Kwon
Dept. of Radio and Info. Communications Eng.
Chungnam National University
Daejeon, South Korea
ymkwon@cnu.ac.kr

***Abstract*—To utilize commercial security log analysis system, small or medium size companies meet big burden in pricing because the price is decided based on data volume to be processed in daily basis. And to build its own system with primitive Hadoop and MongoDB, etc. is too much complex and resource taking of time and man in the initial period. This paper suggests open source stack for the negotiation of these difficulties. We tested performance feasibility comparing with the 1st market share commercial software. It shows reasonable performance in experiments.**

Keywords—Big Data Analysis, Security Log, Splunk, ELK

I. INTRODUCTION

It is not possible to protect explosively increasing number of attacks in perfect ways. But it is important to minimize the aftermath by detecting the sources of attacks and applying appropriate reactions. Ponemon Institute says that it takes 201 days for detection of attacks and 70 days for suppression of accidents in average [1]. As it takes longer time for the reaction, it makes greater the size of damage and the money of solution. Big data solution can help to reduce the time of log analysis for the response of attacks. But commercial products of big data solution cost very expensive for small or medium size companies. This paper suggests the use of open source platforms and shows the feasibility of it by comparing the performance of log analysis between commercial product and open source platform. Chapter 2 shows the limitation of commercial product for companies. Chapter 3 shows how open source stacks can replace the role of commercial product and chapter 4 shows the comparison of performance in log analysis processing. Finally in chapter 5, we conclude this paper.

II. COMMERCIAL PRODUCT: SPLUNK

Splunk is a commercial software platform for big data analysis. It is used in wide field such as bank, hospital, communication, security, education, etc. The IDC report states that Splunk is the top market share with 28.5% and ranked number one in worldwide IT Operation Analytics (ITOA) software market share for 2015 [2]. Splunk provides powerful UI and users can apply their own UI configuration. More than 12,000 customers in enterprises, service providers and governments in over 110 countries use Splunk solutions in the cloud and on-premises. Gartner says that Splunk is selected as the leader of SIEM in 2016 Gartner Magic Quadrant for consecutive 4 years [3].

Although Splunk has many merits, its pricing plan is burdensome for enterprises. As shown in Table I, the pricing is calculated based on daily data volume [4]. Even though the contract is made based on 1GB data volume in cheap rate, as the data size is getting increased, new contract is necessary and the pricing is getting big burden for Splunk clients.

TABLE I. PRICING PLAN OF SPLUNK

Volume (per Day)	Perpetual License (per GB)	Annual Term License (per GB)	Volume Purchase Discount
1GB	$4,500	$1,800	0%
10GB	$2,500	$1,000	44%
50GB	$1,900	$760	58%
100GB	$1,500	$600	67%
>100GB	Contact sales		

Further, commercial end products has limitation in flexibility for each client's optimization. The flexibility here means easy adaptability to individual conditions and environment. The end product in a market does not support clients' individual special concern and purposes. Client ought to apply and trim their aims to the purchased products. With these pricing and flexibility limitation, companies might be willing to build their own big data analysis system with Hadoop and MongoDB, etc. But, it takes too much time and man costs with high complexity.

III. OPEN SOURCE PLATFORM: ELK STACK

ELK stack consists of several open source components (ElasticSearch, Logstash and Kibana) and can be executed in virtual hardware environment [5]. ElasticSearch is a search engine based on Apache Lucene. Logstash is a dynamic data collection pipeline component to collect necessary logs and sends to ElasticSearch component after transforming to JSON format. Kibana is a component for visualization in various types such as graph, table and map, etc. Setting up a centralized logging solution using these three components is not that quite

978-1-5386-3132-4/17/$31.00 ©2017 IEEE

difficult. But the problem is uncertainty of acceptable performance with naïve setting [6].

Fig. 1 shows the components of ELK stack and Fig. 2 shows the role of Logstash and ElasticSearch.

Fig. 1. Components of ELK stack

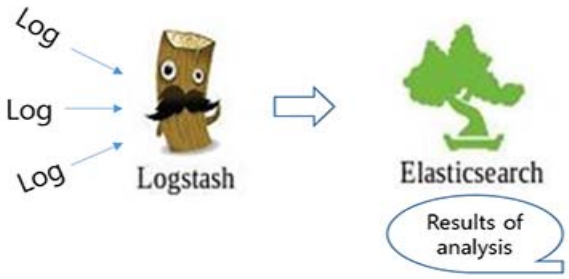

Fig. 2. Log Analysis Procedure in ELK Stack

Once source logs are gathered, Logstash filters and maps original logs to filtered logs according to particular conditions and stores filtered logs in ElasticSearch. To access these processed logs, users use Kibana tool. Fig. 3 is a screen to analyze logs after installation of ELK stack via Kibana. Number 1 area shows menus for search, visualization analysis, dashboard setup, etc. Number 2 area is to query logs. JSON type query is available. Number 3 area shows traffic in graph style. It shows on which time band it has more network traffic. Number 4 area is output of response for the query requests.

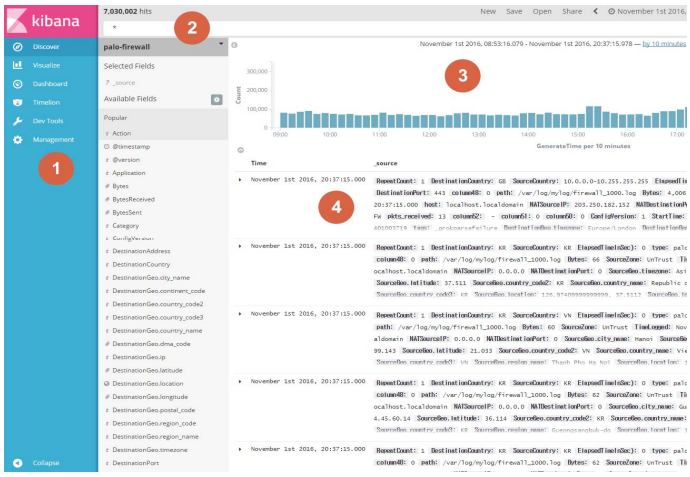

Fig. 3. Access to the Log Analaysis System via Kibana

With Kibana, ELK Stack can show various type of analysis results. Fig. 4 is an example of real-time analysis dashboard. It shows threat analysis screenshot that can be built for customized NOC purposes [7].

Fig. 4. Real-Time Anaysis Dashboard

IV. Performance Experiments

We built an analysis system for large volume security log using ELK stack and compared its performance with that of commercial product, Splunk. The hardware and software environment for ELK stack is shown in Table II. In hardware specification, CPU is a 2 core processor with CPU of 3.30 GHz. Memory is 8GB and CentOS 6.8 is installed. In software's view, Elasticsearch-5.3.0, Logstash-5.3.0, Kibana-5.3.0-linux-x86_64 and Java-1.8.0-openjdk-devel are used.

TABLE II. Specification of ELK stack Installation

Hardware specification	Software specification
CPU – 2 core processor i5-2500 CPU @ 3.30GHz Memory : 8GB OS : CentOS 6.8 (Vmware)	Elasticsearch-5.3.0.tar.gz Logstash-5.3.0.tar.gz Kibana-5.3.0-linux-x86_64.tar.gz Java-1.8.0-openjdk-devel.x86)64

To compare the performances of Splunk and ELK stack, 200 million logs (782 MB), 400 million logs (1.6 GB), 600 million logs (2.4 GB), 800 million logs (3.2 GB) and 1,000 million logs (3.9 GB) were used for analysis. After searching logs originated from a specific IP, result was stored in text file. It was done in CLI (Command Line Interface) mode to measure the exact execution time.

A. Log Data

The logs used in experiments are from Palo Alto Firewall [8] installed in a small company. The original log records are the streams of data separated by comma. Fig. 4 shows one example of log data.

```
<14> Dec 8 14:33:09 1,2015/12/08
14:33:09,001701002739,TRAFFIC,start,1,2015/12/08
14:33:08,203.230.46.147,104.20.5.36,0.0.0.0,0.0.0.0,I
PS,,,web-browsing,vsys1,Main-wire,Main-wire,ethernet1
/14,ethernet1/13,Log_Forwarding_Profile,2015/12/08
14:33:08,198088,1,7944,80,0,0,0x0,tcp,allow,1172,1106
,66,4,2015/12/08 14:33:09,0,any,0,5622694095,0x0,Korea
Republic Of,United States
```

Fig. 4. Log Data Sample

Splunk first stores full text data to hard disk and processes mapping later. But, ELK stack processes mapping before storing data to hard disk as shown in Fig 2. ELK stack needs pre-processing to store logs to ElasticSearch [7].

B. Measurement of Execution Time in CLI mode

Fig. 5 and Fig. 6 shows one sample of analysis time in ELK stack and Splunk system. In Fig. 5, ELK Stack took 18.19 sec to get the logs which matches specific conditions, whereas Splunk took 18.409 sec as shown in Fig. 6.

```
[elk@localhost root]$ time curl -XPOST 'localhost:9200/palo-200/_search?pretty' -d
> {
>   "query": { "match": { "SourceAddress": "10.110.10.144" } },
>   "size": 10000000
> }
> ' > /ELK_STACK/ELK_200.txt
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100 10.8M  100 10.8M    0     86   635k      4  0:00:17  0:00:17 --:--:-- 2542k

real    0m18.190s
user    0m0.001s
sys     0m0.039s
```

Fig. 5. Execution time of ELK when log size is 2 million

```
[root@localhost ~]# time /ELK_STACK/splunk/bin/splunk search 'sourcetype=firewall_200
src_ip=10.110.10.144 | fields *' -maxout '0' > /ELK_STACK/splunk_200.txt

real    0m18.409s
user    0m1.711s
sys     0m0.258s
```

Fig. 6. Execution time of Splunk when log size is 2 million

C. Results of Experiments

Table Ⅲ and Fig. 5 shows the results of whole experiments. ELK stack performs as well as Splunk. Rather, Splunk shows longer execution time than ELK stack. It might be caused by heavy functionalities embedded in Splunk to provide rich and general capabilities.

TABLE III. TIME FOR SEARCHING LOGS

System / Size of Logs	ELK Stack	Splunk
200 Million (782MB)	18.2 sec	18.4 sec
400 Million (1.6GB)	26.5 sec	28.8 sec
600 Million (2.4GB)	36.1 sec	40.1 sec
800 Million (3.2GB)	61.6 sec	65.9 sec
1,000 Million(3.9GB)	74.4 sec	82.2 sec

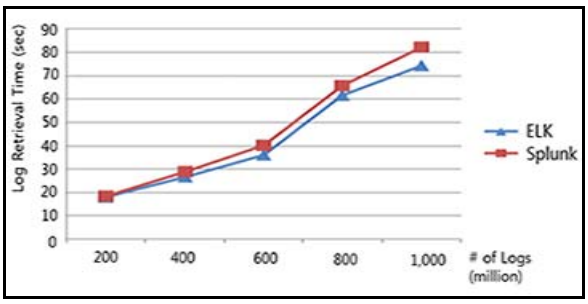

Fig. 5. Searching time for specific conditioned logs (in sec)

V. CONCLUSION

We suggested open source platform, ELK stack, to build a big security log analysis system for small or medium sized enterprises. It cuts the concerns about the installation cost of commercial product in the beginning stage and it makes start-ups free from the effort of building their own log analysis system with primitive Hadoop and MongoDB, etc. ELK stack shows similar or better performance in searching for particular security logs which matches specific conditions. For 1,000 million log file, ELK stacks took 1 min and 14.4 sec, whereas 1 min and 22.2 sec with Splunk. In addition, ELK solution provides various kinds of visualization tools which are useful for security administrators. So, ELK stack can be a powerful beginning security log analysis tool with acceptable performance compared to high cost commercial product.

REFERENCES

[1] I.S. Kim, "The biggest problem in cyber security is to reduce the response time in cyber accidents," article in CIO Biz, December 2016, http://ciobiz.etnews.com/20161221120007 (access in August 2017)

[2] T. Grieser and M.J. Turner, "Worldwide IT operations analytics software market shares, 2015: Special Report," (doc #US416638816), August 2016

[3] J.B. Park, "Splunk is selected as the leader of SIEM in 2016 Gartner Magic Quadrant for consecutive 4 years," article of Electronic Science, August 2016

[4] L. Dignan, “Splunk adds unlimited plan to enterprise pricing mix, “ article of ZDNet, February 2015, http://www.zdnet.com/article/splunk-adds-unlimited-plan-to-enterprise-pricing-mix/ (access in May 2017)

[5] Elastic homepage, “Elasticsearch, the company behind the popular search and analytics open source product, introduces the Elasticsearch ELK stack and Marvel, a real-time management and monitoring solution, “ articles in elastic press, January 2014, https://www.elastic.co/about/press/marvel-2-2 (access in May 2017)

[6] J. Reichardt, “Performance tuning ELK stack,” article in Practical System Administration, March 2015

[7] IPAM Software, “ELK for Network Operations,” http://operational.io/elk-for-network-operations/, June 2014 (access in Jan. 2017)

[8] Palo Alto Networks Firewall Essentials Installation and Configuration Guide, July 2016, https://www.netdevgroup.com/content/paloalto/documentation/netlab_pan7_pod_install_guide.pdf (access in Jan. 2017)