Toward a Secure ELK Stack

Fatimetou Abdou VADHIL[1]
[1]Université de Nouakchott Al-assriya.
Nouakchott, Mauritanie
E-mail: fatiab38@gmail.com

Mohamed Lemine SALIHI[2]
[2]Université de Nouakchott Al-assriya.
Nouakchott, Mauritanie
E-mail: mlsalihi@gmail.com

Mohamedade Farouk NANNE[3]
[3]Université de Nouakchott Al-assriya.
Nouakchott, Mauritanie
E-mail: mohamedade@gmail.com

***Abstract*—Big Data technology is constantly evolving and becoming more and more complex. Companies using large amounts of data are beginning to invest in integrating solutions that can enhance the security of their products and services. In a platform of distributed search and analysis engines, it is likely that security has not been one of the main objectives. But the need to protect private data processed openly in this type of platform compelsthe concerned parties to pay more attention to this issue. In this context, we present a discussion on the work done to ensure the safety of the "Elastic" technology or the ELK Stack (Elasticsearch, Logstash, Kibana).**

***Keywords*—Big Data, Elasticsearch, Logstash, Kibana, Security, Privacy, SIEM, SPBD.**

I. INTRODUCTION

Data collection, processing and analysis must not be at the expense of users' privacy. As this data may include personal and sensitive information that can be used to target an individual, it is necessary to strengthen security and privacy by setting up mechanisms to protect private data.

Platforms for indexing and searching data (such as Elasticsearch, solr,...) are starting to be more used thanks to their powerful analytical features. They address problems that SQL search finds difficult or impossible, especially in an environment using large amounts of data. These platforms are used by many companies such as, Facebook, ebay, GitHub, StackOverFlow... etc.

A. Overview of ELK Stack

ELK is the acronym for three open-source projects, especially Elasticsearch, Logstash and Kibana.

Elasticsearch is among the most popular corporate search engines. There are other components that can be used with Elasticsearch to integrate and visualize the data more easily. Elasticsearch, Logstash, Kibana are the main components of the Elastic stack and are known as ELK.

- *Elasticsearch*: it is a distributed RESTful search and analysis system designed to respond to a multitude of use cases. It allows the analysis of different types of data (structured, semi-structured or unstructured), and to launch different types of searches [1].
- *Logstash*: it is an open source server-side pipeline for data processing, its main role is to integrate data from a multitude of sources, then transform and send them to Elasticsearch (or another storage system) [2].
- *Kibana*: is a web interface that allows to visualize the data stored in Elasticsearch. Kibana has the ability to take a data collection from Elasticsearch and then build different types of analytical graphs, diagrams and data tables, which can be shared [3].

This stack provides very interesting features in terms of search performance, indexing and scalability. However, during the development of the Elastic stack, safety was not considered as a main objective and constraint. The same is true for Big Data platforms, which have been designed for scalability requirements without taking security into account.

[4] presents a list of security threats and vulnerabilities detected in Elasticsearch in recent years. The majority of the vulnerabilities presented are due to the lacks in authentication and access control.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Therefore, the establishment of a methodology to avoid anything that could lead to illegitimate access of private data is essential. Perhaps the installation of security barriers at the database and visualization interface levels is already a very important step. But we must think about attacks of different types that are developing exponentially and about possible threats in the data life cycle.

The need for data security is combined with the need to protect personal data (e.g. GDPR[1]). Legislation in this field is constantly developing and becoming more and more restrictive, hence the need for better security and control of the data contained in these tools.

II. ELASTICSEARCH SECURITY LIMITATIONS

Among the most important security limits is the non-activation of all default security settings. When deploying the stack, it was important to deploy all security measures to protect the Cluster (a collection of nodes containing the data set) Elasticsearch from the Internet. More generally, security and privacy protection should be considered in a new approach to implement these tools by introducing the considerations allowing to get as close as possible to the principle of SPBD "Security and Privacy by design".

[5] presented a list of safety limits for the ELK stack at several levels. These are the possibilities of disclosure and usurpation in some cases. He highlights gaps in authentication and data access control.

He also mentioned that security plugins developed by third parties are not supported by Elastic.

III. RELATED WORKS

Big companies that use large amounts of data, buy very expensive tools and believe that they ensure maximum security with these tools, should know that commercial tools can face security limitations that are sometimes satisfied by Open Source tools.

[6] compared two log management and analysis platforms, namely the ELK stack and Splunk (Commercial Platform). This article suggests the use of the ELK stack for several reasons including the safety limits of the Splunk tool, it results in the ELK stack being more efficient than Splunk. It is very likely that this performance is the result of a security plugin integrated into the ELK stack, because the three main components of this stack do not provide any default security.

The need to protect personal data prompted the Elastic group to offer the X-Pack plugin. It is a commercial plugin, which supports the security of the ELK stack. Without using this plugin, Elasticsearch and Kibana do not offer security features. For example, without having a specialized authentication and authorization tool, any user with access to Kibana can search and visualize all the data stored by Elasticsearch [21].

[21] offers a flexible extension to Search Guard, capable of enforcing user and group access restrictions in the environment. This extension is capable of ensuring access restrictions for users and groups in a multi-user environment.

[7] has separated the Kibana dashboards to restrict users' access to dashboards from others. This feature is part of the [21] "Own Home" solution, except that Own Home has done so by User and Group.

Works in different fields relies on the ELK stack to monitor and systematically collect the recorded data appropriately indexed to retrieve the required parameters and display them in appropriate graphs [8].

Other work focusing on network security has taken advantage of ELK stack's efficiency to analyze log files.

[9] presented a survey that examined the effectiveness of IDS (Intrusion Detection Systems), and discussed external threats. He highlighted how internal threats affect the overall network security mechanism in another article [10], andused the ELK stack to receive the transferred logs.

[9], [11] and[12] are more interested in Intrusion Detection Systems and IDS/IPS Intrusion Prevention Systems. [9] presents a survey that examines the effectiveness of IDS (Intrusion Detection Systems), and [12] has introduced an intelligence model on threats, behaviors and models that are a relevant and important concern. He used ELK to identify various types of cyber incidents. On the other hand, [11] proposes a prototype called

[1]GDPR : General Data Protection Regulation (European union regulation)

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Smart SIEM[2], using ELK with intrusion detection systems to make detection and analysis faster and more efficient.

The use of ELK to help detect anomalies is beginning to take place in the research.[13] believes that the ELK stack is an advantage for organizations seeking to take advantage of threat research methodologies and use analytical data to detect anomalies in their environment.

[14] presents a security ecosystem of attacks and fatal threats. His work helps to develop a technique for aggregating and analyzing logs in real time via the dashboard.

It aims to increase knowledge about traffic patterns and trends, in addition to making an authentic decision over time on maleficent traffic using the ELK stack.

[15] processes IoT data from several subsystems using ELK to develop data analyses to extract and visualize meaningful information, he also discussed the lack of security that represents the major privacy challenge.

[16] made a comparison betweenthree SIEMs, including ELK (this one can be used as a SIEM, adding the Security aspect). This article highlighted the lack of alert and report management on the Kibana interface. He used an alternative to X-Pack's Alert functionality for managing alerts to be integrated with ELK. In fact, transforming ELK into a 100% SIEM solution will require the use of X-Pack, or some of the alternatives, and the integration of additional "third party" components to cover vulnerability assessment and intrusion detection.

[17] provides an encryption mechanism for information from some source on which the network circulates in a protected way and is transferred encrypted to its destination (Logstash then Elasticsearch). Although this encryption mechanism gives a value in terms of security, the results of this proposal show that this is at the expense of Good Search (one of the basic features of Elasticsearch) in terms of response time. In this case of this article, the difference between the duration with and without encryption is not very important, perhaps the amount of data used is not very large. On the other hand, if the amount of data is very large, it maybe proportional to the performance of the ELK stack.

[2]SIEM : Security Information and Event Management

IV. DISCUSSION

Many Big Data platforms have limited security features enabled by default. For example, no authentication is required to access the default web service interfaces. This lack of security has led to several data breaches in recent years. But, some of them allow improvement through the integration of security plugins, because they are Open Source.

Elasticsearch is among those platforms that are free and open source. There is a security plugin for this platform called X-Pack [18] provided by Elastic, its components are as follows: Security, Monitoring, Alerting and Notification, Reporting, Graph, Machine Learning. This plugin offers a wide range of security features, such as encryption, authentication, access control, multi-layer security etc. Using this plugin, a user can manage access privileges to his private data. But these features are only available for a 30-day trial period or in the commercial version, which requires a very expensive license. Search Guard [19] is an alternative to X-Pack that provides similar features to these. It offers basic security features free of charge in the "Community Edition" version, but some are not free (those in the "Enterprise and Compliance Edition"). On the other hand, another alternative involving a Kibana plugin called Own Home [20] and some contributions to the Search Guard plugin was proposed by Takase et al. [21]. It is totally free and provides almost the same features as Search Guard. The two plugins, Search Guard and Own Home, separate the dashboards and visualizations according to the users preferences. This consists in placing them in separate Kibana indexes. The difference between these two alternatives is that Search Guard separates at the ElasticSearch level, while Own Home separates at the Kibana level. Both plugins have almost the same results. [22]

There is also Grafana. It is a visualization tool like Kibana. Both tools allow users to visualize and understand trends in large volumes of log data. Although, there are some security differences, for example Grafana, in addition to the Multi-tenancy feature, provides its users the authentication and access control mechanisms to protect their Dashborads, contrary to Kibana. But in terms of usage and contributors, Kibana is more popular than Grafana, according to Google Trends.

[23] presented Log analysis and visualization solutions (including Kibana and Grafana) in detail.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

X-Pack therefore brings some missing features to Kibana that were already present in Grafana, including authentication, user interface roles and permissions, and alerts. In addition to this, Grafana supports other Data Sources, such as Grafite, InfuxDB, etc. However, Kibana is highly recommended.

The solution implemented by [21] allows to protect data access at Kibana and Elasticsearch level. The problem is that instead of using Logstash, the solution uses Apache Flume, And although Flume is transactional (no data loss when duplicating streams), Logstash remains the most popular, appropriate and recommended for using the ELK stack. In addition, [21] needed to integrate some Patches to adapt Flume with the Elastic platform due to non-compatibility with the Serach Guard plugin.

A. *Research Questions*

- Works have been carried out to ensure the safety of the ELK stack by replacing one of its components with an alternative developed by third parties using missing safety features in this stack. However, it is recommended to use this stack as it is without replacing any of the components. On the other hand, it is necessary to ensure safety mechanisms with the respect of the order ELK.

- Is it possible to have anomalies during data collection before storage, given the storage and visualization's granular security? If so, what intelligent solution could be used to detect these anomalies?

- At which stage of the data lifecycle are they more vulnerable to attacks?

- How to feed a SIEM with the different security / tracking events?

It should also be noted that some of the features of X-Pack are strongly linked, such as 'Alerting' and' Machine Learning', especially with the intrusion detection and prevention (IDS/IPS), given that the detection of anomalies requires Machine Learning.

V. CONCLUSION

This document presents a discussion on the safety of the ELK stack, which is a very interesting solution for indexing, analyzing and visualizing data although there is very little mention of research in this field. He asked questions about new ways to secure this stack and its integration with a SIEM. Among the researchers who use this stack are those who target features in the Elastic stack that are not part of security. Knowing that X-Pack includes the functions: Security, Alerting, Monitoring, Reporting, Graph and Machine Learning, there are a number of alternatives for each component of this extension. This one is not limited to Security. Security therefore represents a small percentage of research in this field. Integration with an adapted SIEM could be used as an alternative security solution. It would allow us to get closer to the principle of SPBD which integration is rarely used in research on ELK stack. This document led us to think of the combination of two or three features of X-Pack: Security, Machine Learning and Alerting. These features can ensure the security of a SIEM. Again, the use of an ELK as a SIEM will require the integration of certain security features. On the other hand, it will allow you to benefit from a free and efficient SIEM.

REFERENCES

[1] https://www.elastic.co/fr/products/elasticsearch
[2] https://www.elastic.co/fr/products/logstash
[3] https://www.elastic.co/fr/products/kibana
[4] https://www.cvedetails.com/vulnerability-list/vendor_id-13554/Elasticsearch.html
[5] https://www.elastic.co/guide/en/elastic-stack-overview/current/security-limitations.html
[6] S. J. Son and Y. Kwon, 'Performance of ELK stack and commercial system in security log analysis', in 2017 IEEE 13th Malaysia International Conference on Communications (MICC), Johor Bahru, pp. 187–190, 2017.
[7] A. D. D. Fuente, O. O. Andreassen, and C. Charrondière, "Monitoring Mixed-Language Applications with Elastic Search, Logstash and Kibana (ELK)," in Proceedings of ICALEPCS, pp. 9–12, 2015.
[8] S. Dharur and K. Swaminathan, "Efficient surveillance and monitoring using the ELK stack for IoT powered Smart Buildings," in 2018 2nd International Conference on Inventive Systems and Control (ICISC), pp. 700–705, 2018.
[9] I. Y. M. AL-Mahbashi, P. Chauhan, S. Shukla, and M. B. Potdar, "Review on Efficient Log Analysis to Evaluate Multiple Honeypots using ELK", Vol.2 Issue.6, 2016.
[10] I. Y. M. Al-Mahbashi, M. B. Potdar,andP. Chauhan, "Network security enhancement through effective log analysis using ELK". In 2017 International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2017.
[11] M. EL ARASS and N. SOUISSI, "Smart SIEM: From Big Data logs and events to Smart Data alerts", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Vol.8, Issue. 8,June 2019.

[12] H. Almohannadi, I. Awan, J. Al Hamar, A. Cullen, JP. DissoandL. Armitage,"Cyber threat intelligence from honeypot data using elasticsearch". In: 2018 IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA). IEEE, pp 900–90, 2018.

[13] P. Delgado, "Developing an Adaptive Threat Hunting Solution: The Elasticsearch Stack",

[14] A. Kumar, A. Bandyopadhyay, H Bhoomika, I. Singhania and K. Shah, "Analysis of Network Traffic and Security through Log Aggregation", International Journal of Computer Science and Information Security (IJCSIS), Vol. 16, Issue. 6, June 2018.

[15] M. Bajer, "Building an IoT Data Hub with Elasticsearch, Logstash and Kibana." In Future Internet of Things and Cloud Workshops (FiCloudW), 2017 5th International Conference on, pp. 63-68. IEEE, 2017.

[16] N. Moukafih, S. Sabir, A. Lakbabi, and G. Orhano, "SIEM selection criteria for an efficient contextual security". In: 2017 International Symposium on Networks, Computers and Communications (ISNCC). IEEE, 2017.

[17] M. Sanchez and L. Urquiza, "Security Enhancement through Effective Encrypted Communication using ELK", ICBDE'19 Proceedings of the 2019 International Conference on Big Data and Education, 2019.

[18] https://www.elastic.co/fr/products/stack

[19] https://search-guard.com/

[20] https://github.com/wtakase/kibana-own-home

[21] W.Takase, T. Nakamura, Y.Watase, and T. Sasaki, "A solution for secure use of kibana and Elasticsearch in multi-user environment", International Symposium on Grids and Clouds (ISGC), 2017.

[22] https://stackoverflow.com/questions/44542708/whats-the-best-kibana-multi-tenancy-free-open-source-project

[23] A. Vainio, "Implementation of Centralized Logging and Log Analysis in Cloud Transition", 2018.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500