Using Users Profiling to Identifying an Attacks

Aarthi M[a], Nivetha N[b], Sharvesha J[c], Udhaya Kumar M[d], and Yuvaraj T[e]
[a,b] Assistant Professor, Department of Computer Science and Engineering, K. Ramakrishnan College of Technology, Trichy
[c,d,e] UG Student, Department of Computer Science and Engineering, K. Ramakrishnan College of Technology, Trichy

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 16 April 2021

Abstract: Nowadays crime activities have increased in almost all areas, but this paper only focuses on who performs illegal activities within the organization. A user may perform insider and phishing attacks in the organization. A legitimate user of an organization may try to login in the administrator id, and then perform some illegal activities. Due to these activities, sensitive data can be modified or corrupted. Identification of illegal user's behavior is very difficult within the organization. The scope of this work is to analyze the log files, to filter out the user profiles of those who are involved in suspicious activity and to detect the suspicious activity of the user. In any organization, large number of log files is being generated, log manger system helps to take an optimal solutions. Although, a variety of log supervisor gadget exists, however, they are not providing that much efficiency. This paper analyses the ELK stack working principles and compare it with Splunk. ELK stack include many additional features such as indexing, preprocessing a large amount of logs and producing graphical representation output using kibana.

Keywords: Insider, Phishing, Elasticsearch, Logstash, Kibana.

1. Introduction

System administrators should have great knowledge about the current trend of attacks and its behavior trails of the attacks, then only they could maintain an organization more secure and prevent the problem. April 2017 cyber-attack statistics[7] says that 74.1% cybercrime activities happened on internet and 21.2% activities were of cyber espionage activities, it means using a computer network to access the confidential information and mainly targeted on government sector. 3.5% were hackactivism activities, it means unauthorized access to private file with the organization[7]. 1.2% were cyber warfare, it means using computer technology to disturb the state or organization. These are the reasons behind the motive of the attack. Attacker activities are now rapidly detected using ELK stack[9,10] to analysis and detect illegal activities in the organization.

Log control system [9] offers with a high quantity of log information generated with the aid of using computers. There is lots of the process includes inclusive of log collection, centralized aggregation, long-time period retention, log evaluation, logs looking and reporting. The evaluation of those logs now no longer only facilitates the agencies of their decision making, however additionally in improving their protection and services. An insider attack is a one of the malicious attack on a computer system by a user with, authorized system access and perform some illegal activities [14]. For example: legitimate user impersonation as an administrator. Phishing attack[13] is a one kind of attack and its attempt to obtain confidential information such as credit card details, account holder name, passwords often for malicious reasons, these are issues in trustworthy entity in an electronic communication [15]. Some phishing e-mails also contains malicious or unwanted software that can track your activities or slow your computer. For example: Creating a fake website and inserting it within the legitimate webpage.

2. ELK

ELK is a hard and fast of 3 additives, they are, Elasticsearch, Logstash and Kibana [10]. Although, Logstash and Elasticsearch paintings on separately, the 3 additives are valid for use as an incorporated solution, currently known as the Elastic Stack [9]. These components are available in open sources and it provided by Elastic company. Fig.1 gives descriptive overview of the working process of ELK Stack. Logstash collect data from filesbeats then which is transferred to the Elasticsearch. It creates indexes for data after visualizing the data.

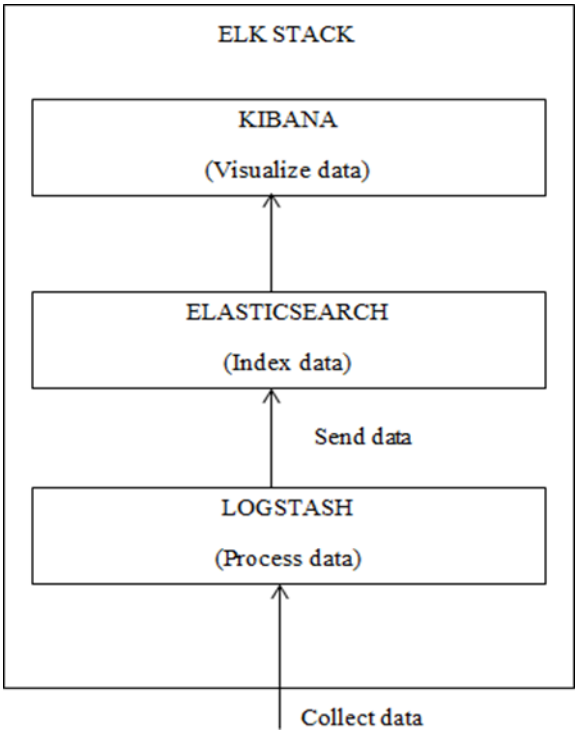

Fig. 1 Architecture of ELK Stack

A. Logstash

Logstash is a bottom of the ELK stacks and it is an open source component available in the cloud. Logstash[9] used to collect all log files with the help of Filebeat. It collect logs from several source to elasticsearch. In Fig. 2 describe the three stages of pipeline performed for each event.

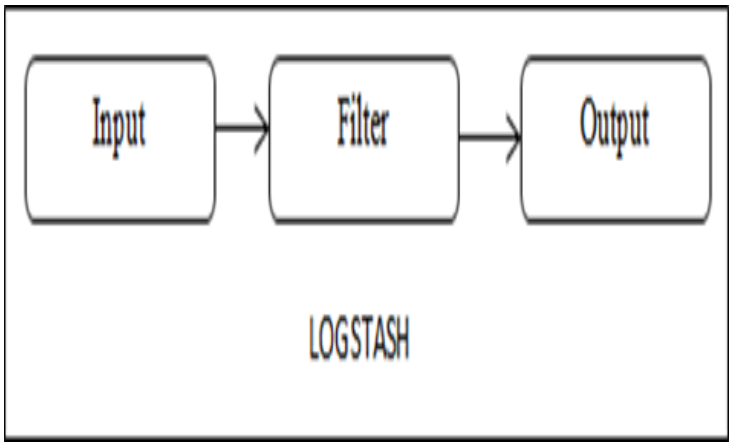

Fig. 2 Three stage pipeline of Logstash

```
input
{
file
{
path => "C:\Users\Aarthi\Downloads\filebeat-5.6.1-windows-x86\logs"
start_position => "beginning"
since_path => "/dev/null"
}
}
filter
{
mutate {convert => ["Ip_address", "integer"]
mutate {convert => ["Timestamp", "integer"]
mutate {convert => ["Method", "string"]
mutate {convert => ["Path", "string"]
}
output
{
elasticsearch
{
hosts => "localhost"
index => "Data"
document_type => "Data_log"
}
stdout {}
}
}
```

Listing 1. Logstash configuration

B. Elasticsearch

Elasticsearch is placed in the center of ELK stack and it is an open supply thing to be had withinside the cloud. Logstash ship records too Elasticsearch, it retrieves records, then create the index (Listing 2) for the records. Elasticsearch [9] allows governing and showing all factors of a grouping data.. Elasticsearch may be used like every other NoSQL statistics store. Elastic given wrappers for the Application Programming Interface in famous programming languages inclusive of DOT Net, Java, Groovy and .NET. Data maintain mechanism is important,to analyses the statistics of data from multiple sources. Delete operation is achieved on entire index.PUT 'http://localhost:9200'

```
{
  "name" : "cpZNhxT",
  "cluster_name" : "elasticsearch",
  "cluster_uuid" : "mEgbuczbR9a06-k-hgYqwQ",
  "version" : {
    "number" : "5.1.1",
    "build_hash" : "5395e21",
    "build_date" : "2016-12-06T12:36:15.409Z",
    "build_snapshot" : false,
    "lucene_version" : "6.3.0"
  },
  "tagline" : "You Know, for Search"
}
```

Listing 2. Index template

C. Kibana

The kibana is a pinnacle of ELK stack and it's far an open source additives to be had withinside the cloud. It has become in particular designed as a graphical platform for Elasticsearch. Kibana provide the graphical output and using interface to analyses the data. Kibana have four major parts: dashboard, discover the data, Graphical representation of output and managing the data. These are the parts are used to analyses the data from multiple sources.

3. SPLUNK

Splunk is an industrial software program for facts,analysis platform. It is used in lots of fields along with communication, security, hospital, education. Splunk affords an effective User Interface (UI) and customers can use their personal UI configuration [16]. More than 12,000 clients in enterprises, provider vendors and governments in over a hundred and ten nations use Splunk answers in the cloud and onpremises.

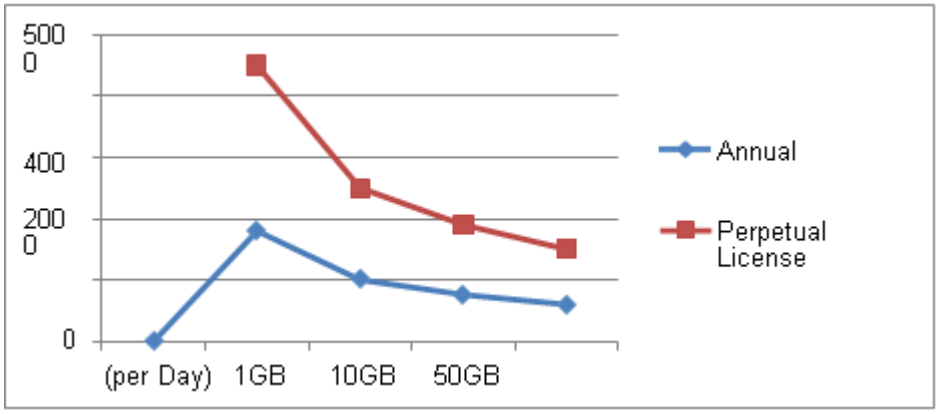

Fig. 3: Splunk price plan

4. Related Work

Tarun Prakash, et al., [1] The internet penetration charge goes greater complex, large variety of log documents is being generated, that have hidden statistics having enormous value. To release the hidden returns, log control gadget enables in making decisions [18]. Nowadays masses of log control exists, however,they both are high priced or fail to scale. These works [1] demonstrates the running of ELK ecosystem, i.e. Elasticsearch, Logstash and Kibana integrate collectively too Correctly offer an interactive, examine the log documents and easily comprehensible insights.In this paper done the Geo identity of customers primarily based totally at the gate admission to logs the usage of ELK stacks, They have used a touch amount of log statistics set simply to illustrate the usability of ELK stack. It is cost,powerful and had a large,energetic contributor base which makes it greater aggressive compared to different.

Andrei Talaş, et al., [2] Elasticsearch provide advantages to user, they can change definitions at any time, while in other traditional approaches should be created before transferring the information and if a definition must be changed, they require resending of data. ELK Stack provide high performance in retrieving, processing, searching or analyzing large volumes of data[11].

Sung Jun Son, et al., [3] This paper compares to performance analysis and execution time for ELK stack and Splunk. ELK stack can be a powerful tool for security log analysis and also it is an open source product compared to the other log analysis tool which are high cost. It is freely available and easily understandable to work and analysis large amount of data within a particular period[19]. ELK stack can provide support many languages (Java, Python, Curl, etc.) to work with any platform, but Splunk has lots of restrictions and it is a commercial product. Splunk has a price plan list based on that pay to use product.

Fadi Thabtah, et al., [4] Phishing websites are used to retrieve sensitive data for users and information that are used to modify valuable data. Traditional approaches use C4.5 Algorithm to create rules. Each URL has five parts (Protocol, Sub-domain, Resource name, Top-level domain, File path) that are used to create rules. These rules are used to identify the phishing websites and avoid attack activities. Rule 1, length of the URL is less than 54 character then that is a legitimate website otherwise called phishing website. Rule 2, domain name part contains '@' symbol that is a phishing website. Rule 3, domain name part contains '_' symbol that is a phishing website. These rules are generated with C4.5 algorithm.

Sven Kohler, et al., [5] within the organization it very difficult to identify attack activities and without legitimate knowledge attacker can perform some illegal activism[18]. Sometimes legitimate user also performs illegal operation to hide them from organization. Logic-rule-based static analysis approach used to automatically identify the insider attack and also identify how many ways to perform attack activities[20]. This algorithm more effectively identifies the insider attacker of the organization.

5. Methodology

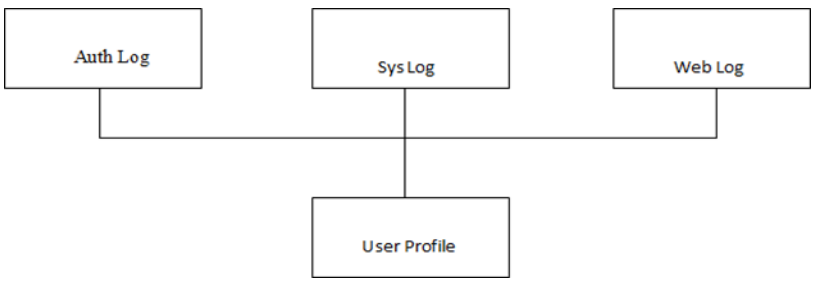

Fig. 4: User Profiling

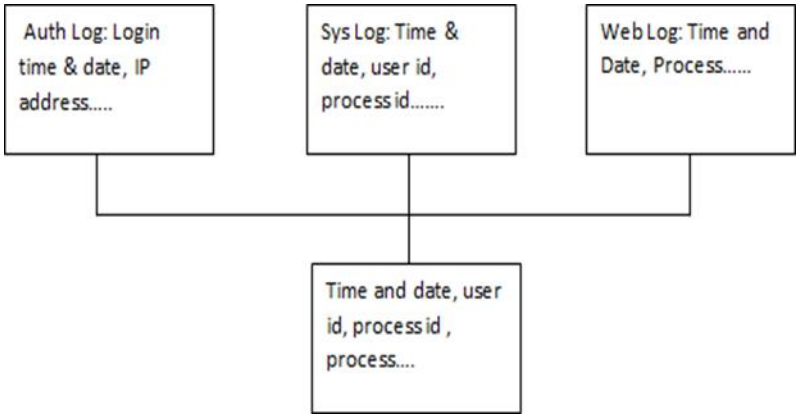

Fig. 5: Correlation of Log files

In Fig. 4 and Fig. 5 describe user profile by correlation of log file[8]. Sys log, Auth log, web log merges to make the user profile. Based on login, timestamp, process id the proposed architecture perform the correlation operation[6]. Fig. 6 gives descriptive overview of working process. Filebeats is used to collect all logs, then shipping to the Logstash Management System (LMS) [21]. It collects all log files with the help of Filebeats, then apply the filter. Elasticsearch perform a filter operation and pattern matching algorithm are used to find the suspicious Traffic IP. Kibana is used to visualize data in graph and snapshot, etc.

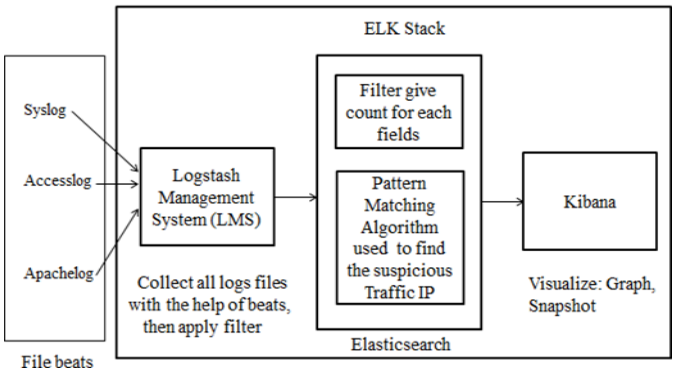

Fig. 6 :Overall Architecture

6. Experimental Results

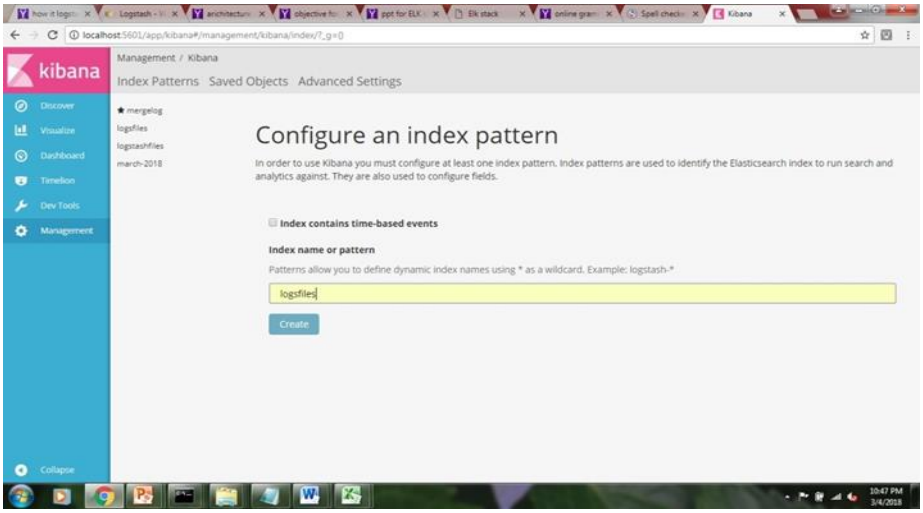

Fig. 7 : Create an index pattern with help of Elasticsearch

Fig. 7 describes the index pattern creation with the help of Elasticsearch. In browser, localhost:5601 show visualization of data. First, we have created the index pattern for data based on that we have visualized the data.

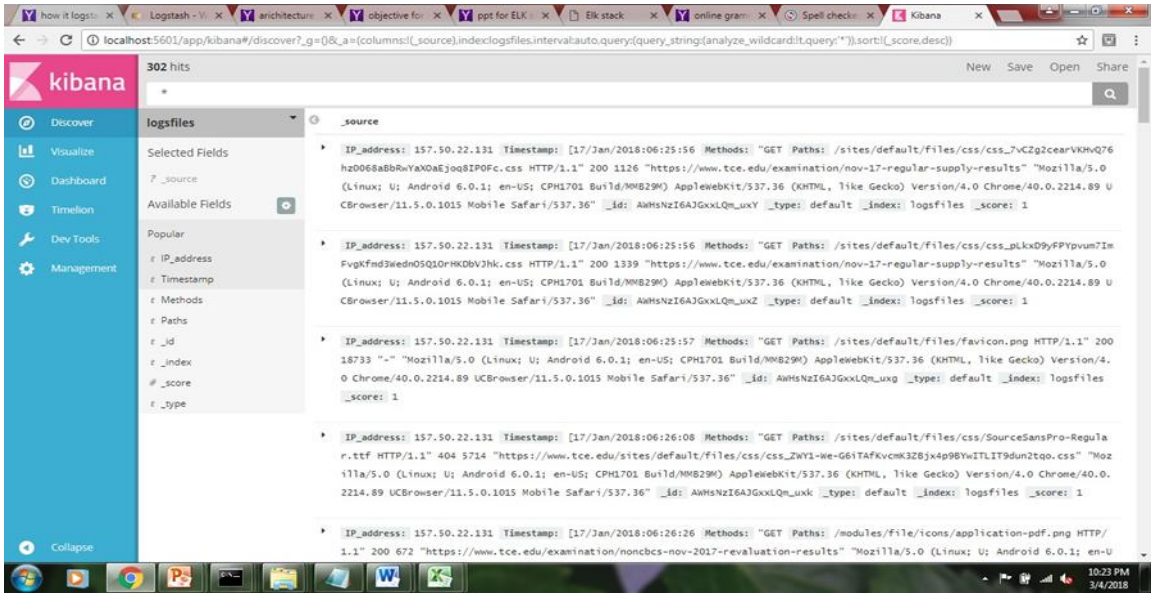

Fig. 8 Retrieve logs from Elasticsearch

Fig. 8 describes the way to retrieve the log files from Elasticsearch. Once created an index pattern for data that is used to retrieve the entire log from elasticsearch is created.

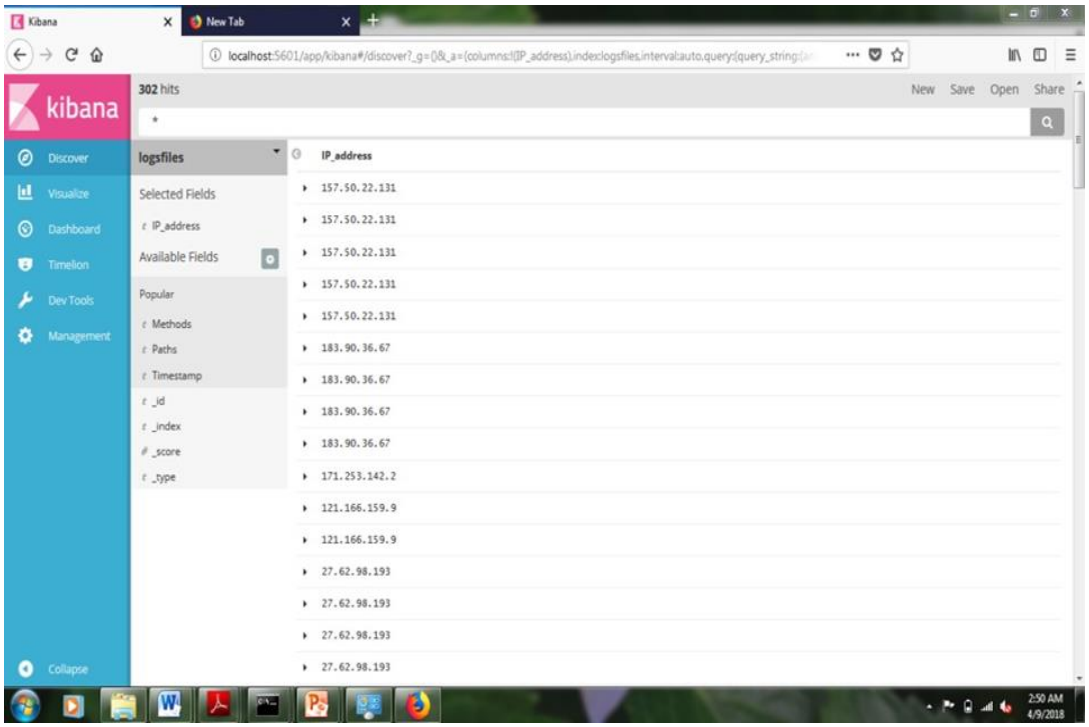

Fig. 9 Frequently login IP address.

Fig. 9 shows frequent login, IP address. A legitimate user only login per day once or more than two, based on that analysis the most frequently login, IP address that user may perform the insider attack.

Fig. 10 Frequently Login User Proflie

Fig. 10 depicts frequently login profile. Based on the most frequent login, IP address will collect relevant timestamp and URLs.

7. Conclusion

We have counseled open supply platform of ELK , to construct a large safety log evaluation device for small or medium sized enterprises. It offers information about the setup value of business merchandise withinside the beginning level and it makes startups unfastened from the attempt of constructing their personal log evaluation device with primitive Hadoop and MongoDB, etc. ELK stack display comparable or excessive overall performance in locating for unique safety logs which fits particular conditions. In addition, ELK answer offers diverse sorts of visualization gear which can be beneficial for safety administrators. So, ELK stack may be a effective safety log evaluation device with proper overall performance in comparison to excessive value business product as proved from our investigation results.

References

1. Prakash, Tarun, Misha Kakkar, and Kritika Patel, "Geo-identification of web users through logs using ELK stack." In Cloud System and Big Data Engineering, 2016 6th International Conference, pp. 606-610. IEEE, 2016.
2. Tales, Andrei, Florin Pop, and Gabriel Neagu, "Elastic stack in action for smart cities: Making sense of big data." In Intelligent Computer Communication and Processing (ICCP), 2017 13th IEEE International Conference on, pp. 469-476. IEEE, 2017.
3. Son SJ, Kwon Y. Performance of ELK stack and commercial system in security log analysis. In2017 IEEE 13th Malaysia International Conference on Communications (MICC) 2017 Nov 28 (pp. 187-190). IEEE.
4. Mohammad RM, Thabtah F, McCluskey L. Intelligent rule-based phishing websites classification. IET Information Security. 2014 Mar 4;8(3):153- 60.
5. Sarkar A, Köhler S, Ludäscher B, Bishop M. Insider attack identification and prevention in collection-oriented dataflow-based processes. IEEE Systems Journal. 2015 Oct 8;11(2):522-33.
6. Colombini, Clara Maria, Antonio Colella, and Italian Army. "Digital scene of crime: technique of profiling users." Journal of Wireless Mobile Networks, pp. 50-73, 2012.
7. Z. Kerravala. Configuration management delivers business resiliency. The Yankee Group, Nov. 2002.
8. Peng, Jian, Kim-Kwang Raymond Choo, and Helen Ashman. "User profiling in intrusion detection: A review." Journal of Network and Computer Applications, pp. 14-27, 2016.
9. Elasticsearch 5.1.1 online docementation – Elasticsearch product description, http://www.elastic.co/products/elasticsearch

10. Kibana User Guide [5.2], Plugin development https://www.elastic.co/guide/en/kibana/current/plugin-development.html (accessed 10.01.2017)
11. Kılıc, Ugur, and Isıl Karabey. "Comparison of Solr and Elasticsearch Among Popular Full Text Search Engines and Their Security Analysis.", In Future Internet of Things and Cloud Workshops, 2015 6th International Conference on, pp. 163-168. IEEE, 2015.
12. Churilin, Artyom. "Choosing an open-source log management system for small business." PhD diss., Master's Thesis, Faculty of Information Technology, Tallin University of Technology, Tallinn, Estonia.
13. YZhang, J. Hong, and L. Cranor. "CANTINA: A content based approach to detect phishing websites." In Proceedings of the 16th International Conference on World Wide Web, pp. 139-164, 2014.
14. Sarkar, Anandarup, Sven Köhler, Sean Riddle, Bertram Ludaescher, and Matt Bishop. "Insider attack identification and prevention using a declarative approach." In Security and Privacy Workshops (SPW), pp. 265-276. IEEE, 2014.
15. Alhazmi, Omar H., Yashwant K. Malaiya, and Indrajit Ray. "Measuring, analyzing and predicting security vulnerabilities in software systems." Journal of Computers & Security, pp. 219-228, 2013.
16. Bajer, Marcin. "Building an IoT Data Hub with Elasticsearch, Logstash and Kibana." In Future Internet of Things and Cloud Workshops, 2017 5th International Conference on, pp. 63-68. IEEE, 2017.
17. Yasu, Yoshiji, and Andrei Kazarov. "Performance of Splunk for the TDAQ Information Service at the ATLAS experiment." In Real Time Conference (RT), pp. 1-6. IEEE, 2014.
18. Aarthi, M. and Bhuvaneshwaran, A., 2021. Iot Based Drainage and Waste Management Monitoring and Alert System for Smart City. Annals of the Romanian Society for Cell Biology, pp.6641-6651.
19. Avudaiappan, T., Balasubramanian, R., Pandiyan, S. S., Saravanan, M., Lakshmanaprabu, S. K., & Shankar, K. (2018). Medical image security using dual encryption with oppositional based optimization algorithm. Journal of medical systems, 42(11), 208.
20. Sivakumar M, Reddy US. Aspect based sentiment analysis of students opinion using machine learning techniques. In2017 International Conference on Inventive Computing and Informatics (ICICI) 2017 Nov 23 (pp. 726-731). IEEE.
21. Pavithra, M., Sindhana, A.M., Subajanaki, T. and Mahalakshmi, S., 2021. Effective Heart Disease Prediction Systems Using Data Mining Techniques. Annals of the Romanian Society for Cell Biology, pp.6566-6571.
22. Tresa, M., Francina, S., Jerlin Oviya, V. and Lavanya, K., 2021. A Study on Internet of Things: Overview, Automation, Wireless Technology, Robotics. Annals of the Romanian Society for Cell Biology, pp.654
23. Mallikarjunan KN, Shalinie SM, Sundarakantham K, Aarthi M. Evaluation of security metrics for system security analysis. InComputational Intelligence: Theories, Applications and Future Directions-Volume I 2019 (pp. 187-197). Springer, Singapore.