JEPIN
(Jurnal Edukasi dan Penelitian Informatika)
ISSN(e): 2548-9364 / ISSN(p) : 2460-0741

Vol. 9
No. 2
Agustus
2023

Simple, Fast, and Accurate Cybercrime Detection on E-Government with Elastic Stack SIEM

Ichsan Yudhianto[#1], Cutifa Safitri[*2]

[#]*Informatics Study Program, Faculty of Computing, President University*
Jl. Ki Hajar Dewantara, Kota Jababeka, Cikarang Baru, Bekasi 17550 - Indonesia
[1]ichsan.yudhianto@student.president.ac.id
[2]cutifa@president.ac.id

***Abstrak*— Increased public activity in cyberspace (Internet) during the Covid-19 pandemic has also increased cybercrime cases with various attack targets, including E-Government services. Cybercrime is hidden and occurs unnoticed in E-Government, so handling it is challenging for all government agencies. The characteristics of E-Government are unique and different from other service systems in general, requiring extra anticipation for the prevention and handling of cybercrime attack threats. This research proposes log and event data analysis to detect cybercrime in e-Government using System Information and Event Management (SIEM). The main contribution of this research is a simple, fast, and accurate cybercrime detection process in the e-Government environment by increasing the level of log and event data analysis with the SIEM approach. SIEM technology based on machine learning and big data is implemented with Elastic Stack. The implemented technique can be used as a mitigation program against cybercrime threats that often attack and target e-Government. With simple, accurate, and fast cybercrime detection, it is expected to improve e-Government security and increase public confidence in public services organized by government agencies.**

Kata kunci*— *Security, Cybercrime Detection, SIEM, Log Analysis, E-Government

I. INTRODUCTION

As people's internet activities increased during the Covid-19 pandemic, the number of cybercrime attacks that caused significant harm to everyone has also increased drastically [1]–[3]. Cybercrime attacks have been recorded to have targeted various facilities in the banking sector, government organizations, and business companies during this pandemic [3]. The threat of cybercrime on government facilities, especially e-Government services, deserves its attention because of its enormous impact on the interests of society at large [4].

Efficient, fast, accessible public services and accountable processes are the goals of e-Government [4]–[6]. However, the public's sustainable use of e-Government services is highly dependent on the level of security. Services provided through the Internet and mobile connections in e-Government are very vulnerable to cybercrime threats[4], [7]. And in fact, government institutions as service providers may not always be able to predict and defend against all types of cybercrime attacks [4], [7]–[9].

E-Government is a mechanism of governance and public services based on information and internet communication technology to improve services from public sector organizations or government [4]–[6], [10].

As a particular domain for information network applications with large amounts of data, e-Government has its characteristics and is different from service systems in general. The distinctive features of e-Government [10] are: the data and information are highly confidential and sensitive, as a means of monitoring the administration of government services online and as a unique means of public services through the internet network. With these distinctive characteristics of e-Government, handling risks and threats to the security of networks, data, and service systems is very important to be considered [10].

Many cybercrime incidents occur due to the victim's lack of awareness of security threats or their lack of ability to protect themselves from them [11]. However, organizations that can protect themselves from threats, like government institutions, are often victims of cybercrime attacks. The level of attractiveness, such as public reputation, is one of the factors that cause a government institution to become the primary target for cybercrime attacks [11]–[13].

According to sources [12], cybercrime refers to criminal activities through the Internet and other computer networks to obtain secure information or authorization rights. Unlike financial losses, which are easily noticeable, the loss or corruption of information resulting from cyberattacks may go undetected as they are often difficult to detect [14]. Many users or administrators are oblivious to the vulnerability of their computers and networks, and as a result, their systems may be attacked or hacked without their knowledge [11], [14], [15]. This is because physical access restrictions do not limit cybercrime and can occur without the victim's knowledge [16].

Literature studies in [13], [17], [18] have conducted a comprehensive review of several types of methods for

detecting cybercrime, including statistical techniques, machine learning, neural networks, deep learning, fuzzy logic neural, data mining, computer vision techniques, biometric techniques, cryptography, forensic tools, and penetration tests and DNS analysis.

Abdulghani Ali Ahmed [17] collects and analyses digital evidence to detect cybercrime. The approach focuses on monitoring network activity and analyzing cybercrime behaviour. The detection and identification of cybercrime involve six stages: data collection, detection, investigation, reporting, evidence collection, and maintenance. The surveillance network captures all network traffic (normal and malicious) using Raspberry Pi as the data collection tool. In the detection phase, predefined rules examine network traffic and filter out suspicious ones. Evidence from the negative traffic logs is used to verify whether the traffic is malicious. Cybercrime activity log files are stored in the database during the maintenance and collection phases. New rules are updated to enhance system performance by reducing false alarm rates. The logged data, including timestamps, source and destination IPs of alerts, and activity patterns, can be viewed in the reporting phase.

In their research, Khan et al. [13] utilized data mining techniques to identify instances of cybercrime. This approach involves extracting data from databases and identifying patterns by deriving association rules. The researchers then employed clustering, which groups data with similar characteristics, to "discover patterns" in the sequence of events in the system log file. Using cluster analysis with the clustering technique helps identify data patterns that exhibit high similarity and consistently occur within the log file. The cluster analysis process for the log file involves several steps, including evaluating the log file, performing the mining process based on time and date, scanning data, and adding data found in the main file. The procedure records data containing standard and abnormal (malicious) patterns. Using clustering techniques makes identifying repeatedly occurring data possible, allowing for identifying Denial of Service cyberattacks through patterns with similar features that persistently appear in log data.

The methods used to detect cybercrime in the previous research [13], [17] utilize log data as one of the research data sources because log data is the only source available from a system that contains detailed status, behaviour, and runtime information while operating in a production environment [19], [20], [21]. Log files are also generated by-products from significant hardware and software manufacturers that are widely used [22].

Correlated logs can be analyzed to assist in the process of identifying security incidents, policy violations, fraudulent activities, threat intelligence, and security troubleshooting on a particular network system [17], [23], [20], [21], [24], [25]. Instead of analyzing log data directly, a different approach is to use log management techniques that utilize SIEM [22], [26]. As a log management system, SIEM provides near-real-time analysis of log data[27]. Log data is collected, correlated, and generated from various resource pools within an organization's IT infrastructure [22], [24].

Based on this background, an idea was created to propose an alternative method of detecting cybercrime that is suitable to the characteristics of E-Government, namely by analyzing log data centrally based on SIEM (Security Information and Event Management). The SIEM system is implemented with Elastic Stack, advanced log management that combines three open-source projects: Elasticsearch, Logstash, and Kibana.

With Elastic Stack's easy and real-time extensive data management, analyzing log and event data to detect and mitigate cybercrime in e-Government services will become more efficient.

II. Literature Review

A. E-Government

E-Government, in theory, is classified into four levels [5]. The first level is the publication of government information through website-based media. At the second level, the interaction between government agencies and the public is carried out by utilizing email or other electronic correspondence platforms which are carried out online. At the third level, government agencies and the user community transactions are carried out reciprocally using web-based and mobile online application systems. While the fourth level of E-Government is the development of the third level, where the reciprocal interaction of government and society carried out through the internet network is aimed at making decisions from government officials that can be legally binding.

B. Cybercrime

Cybercrime emerged along with internet technology in various fields to support human life in the digital era. In general, cybercrime is a combination of crime and cyberspace. Where a crime occurs, it implies an act of the attacker or perpetrator that is considered dangerous and potentially harms people or society. Graeme R. Newman characterizes cybercrime as a type of conduct in which "criminal activity involves the use of a computer or network as either a tool, a target, or a place." [15]. Cybercrime often goes unnoticed and remains concealed due to the cyber world's virtual nature, making it challenging to detect.

C. Log Files

Computer systems generate log files daily, amounting to thousands or millions of recorded activities [25]. Managing and utilizing log data is critical for developers and operators in an operational environment. Logs are textual data generated by logging statements within a system's source code, and they typically do not follow a specific structure [19], [21]. Applications generate multiple informative and valuable logs for tracking and investigation purposes, with system logs being the most

important. These logs record the system's status and critical events at various key points to facilitate easy debugging [20].

D. Security Information and Event Management (SIEM)

SIEM is the combination or fusion of Security Information Management (SIM) and Security Event Management (SEM) [27], [28]. SIM and SEM work on collecting and analyzing relevant security-related data but with a different focus. SIM focuses on log management, such as aggregating and normalizing logs from various sources and historical analysis for long-term retention and improvement. At the same time, SEM focuses on identifying and analyzing specific security events based on data aggregation for a timely incident management process [27].

SIM is known for its capacity to automate and manage the gathering of logs and events from diverse sources, including intrusion detection systems, firewalls, servers, and antivirus systems. SIMs can efficiently store and standardize data, sift through events, and facilitate reporting and analysis. In contrast, SEM is known for its robust analytical capabilities, including real-time threat analysis and reporting, alert generation, and visualization tools such as charts and dashboards that assist security operations.

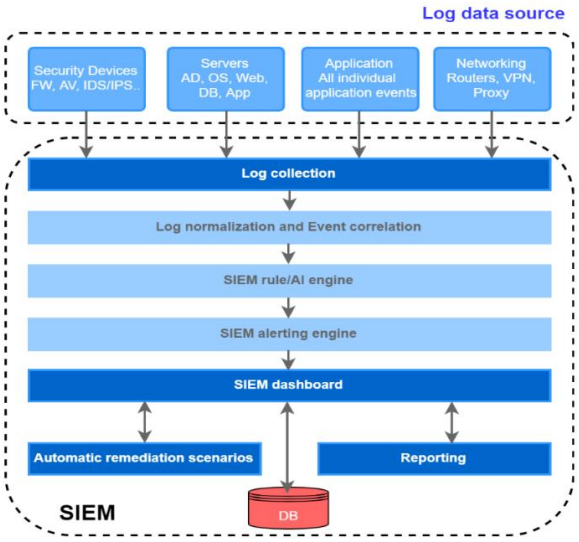


Figure. 1 Architecture of SIEM [21]

The core of SIEM is the collection of logs and analysis of events from various resources, and the results are visualized [16]. As a log analysis tool, the components of the SIEM architecture are illustrated in Figure 1.

E. Elastic Stack

The Elastic Stack comprises three open-source projects: Elasticsearch, Logstash, and Kibana, which can collaborate as a unified system [29], [30]. Logstash serves to process data on a server, gather log data from multiple sources through lightweight delivery agents, transform it, and then transmit it to Elasticsearch, which functions as a search engine [29]. Elasticsearch uses machine learning to perform indexing and real-time analytics, allowing for analyzing all logs and events related to service systems and resources [29]. Kibana, on the other hand, is a data visualization tool that utilizes Elasticsearch as an analytical framework, presenting informative charts and graphs [29]. In the context of a SIEM solution, the Elastic Stack provides high performance, flexibility, and extensibility in software systems [31].

With its construction, Elastic Stack as a modern SIEM technology solution can increase the effectiveness and ease of collecting, storing, and processing advanced data and security events on big data centralized in real-time more than the typical SIEM model [30], [31].

F. Endpoint Detection and Response (EDR)

Endpoints are an essential part often used as an entry point for cybercriminals to attack network operations [32]. The term endpoint refers to all devices that are the endpoints of network communications, such as PCs, laptops, tablets, surveillance cameras, bright lights, smart tv, sensors, and so on. As long as a device can receive internet signals, the device can be categorized as an endpoint. EDR, or Endpoint Detection and Response, is a security tool designed to protect endpoints from threats. This integrated endpoint security solution leverages continuous real-time monitoring and data collection, utilizing rule-based automatic responses and capability analysis, as noted in [32].

EDR tools continuously monitor end hosts' activity and raise threat alerts if any observed behaviour is potentially harmful [32]. Besides providing security alerts to endpoints, EDR responds automatically to threats/attacks. However, EDR security solutions have several drawbacks [32]. The effectiveness of EDR is impeded by various challenges such as a large number of false alarms generated, accumulation of investigative tasks leading to the backlog for analysts, the need for manual validation of threat alerts due to numerous low-level system logs, the difficulty of identifying important information in a large volume of data (also known as "needle in a haystack" problem), and a high resource load due to log retention.

III. METHODOLOGY

The stages of this research methodology are designed using a framework to be more structured and planned, as depicted in the illustration presented in Figure 2.

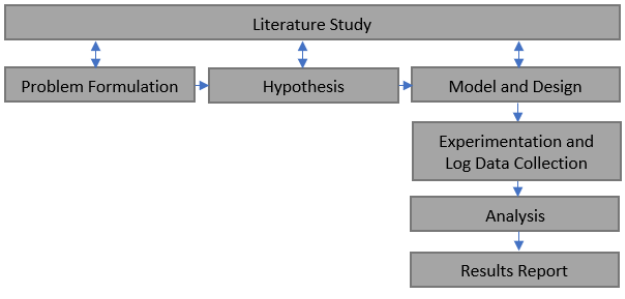


Figure. 2 The Research Methodology Framework

A. Proposed Model and Design

Based on the background problem formulation, a hypothesis emerges that the log data generated and recorded from the E-Government web service system network can detect cybercrime threats and attacks using SIEM-based log analysis. The authors propose a SIEM-based centralized log data analysis model on E-Government for cybercrime detection using an Elastic Stack, as shown in Figure 3.

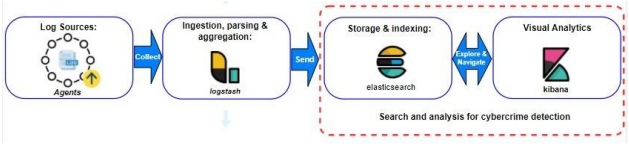


Figure. 3 The Proposed Model

The author developed the architecture design using the proposed model during the initial implementation stage, as shown in Figure 4. In addition to protecting device security, endpoint detection and Response (EDR) on various devices in an organizational environment is also used as an agent to record and collect log data. EDR will send log data to Elastic Stack via Logstash for parsing and then store it indexed in Elasticsearch. Security analyst performs threat hunting using Kibana tuning. Elasticsearch conducts the data analysis process through Kibana; all activities and events are thoroughly identified.

B. Experimentation and Log Data Collection

Research experiments were conducted using log events data from the E-Government web system network services at the Regional Personnel Agency of Jakarta Capital City Province Government (Badan Kepegawaian Daerah Provinsi DKI Jakarta). As the object of research, this local government agency provides several E-Government services for managing human resources of the state civil apparatus at the DKI Jakarta Provincial Government level.

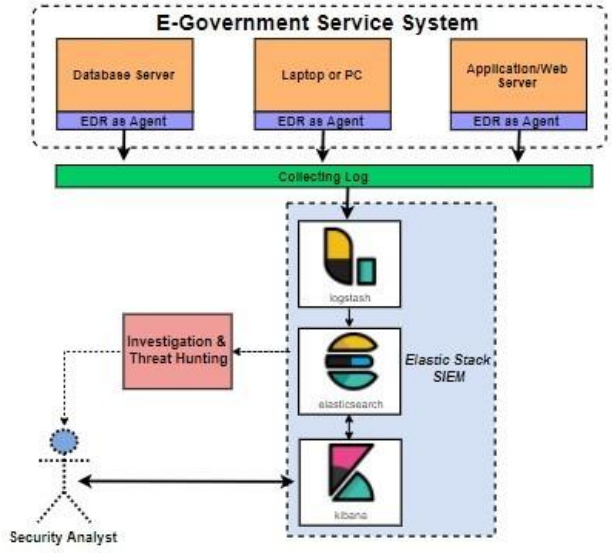


Figure. 4 The Architecture Design

The study was conducted using an experimental environment, as shown in Table I.

The infrastructure is established on a host that operates on a virtual machine (VM) from a Nutanix tool. The operating system is the Linux Ubuntu version of 20.04.1 LTS or Focal Fossa. Elastic Stack version 7.9.3 is implemented as the log management and analysis platform, with agent installation on all endpoints (servers and clients) tested using Elastic Agent version 7.16.2.

At the initial stage of the experiment, tests were carried out on the installed model system. The system configuration is set based on the proposed design. EDR as an agent is tested, whether it can record and send actual logs from each endpoint and send them to the system to be stored and analyzed in real-time with data visualization. The testing phase was carried out with two schemes: the case of an unauthorized user and the possibility of an authorized user trying to access the system server, as illustrated in Figure 5.

TABEL I
EXPERIMENTAL ENVIRONMENT INFRASTRUCTURE

Hardware:	
Host: *Nutanix AHV* *(Acropolis Hypervisor)*	One terabyte of RAM, a high processing speed of 134.35 GHz, and a storage capacity of 30 terabytes. It consists of four nodes, each with 256 gigabytes of RAM and *a 16-core Intel Xeon Silver 4208 CPU* operating at a speed of 2.10 GHz.
A virtual machine (VM) to execute the operating system (OS)	40 GB RAM, 5.1 TB data storage, and 16 vCPU
Software:	
Host:	*Nutanix AHV*
Host Server OS:	The version of *Linux Ubuntu* is 20.04.1 LTS, also known as *Focal Fossa*.
Analysis Tools for Log Files:	*Elastic Stack v.7.9.3 (Elasticsearch, Beats, Kibana, Logstash)*
EDR (Agent):	*Elastic Agent v.7.16.2*

Based on the results of the test scheme, the system can read the log data and visualize it according to the specified configuration. The authentication for the authorized users and alerts for the unauthorized users are visualized in Kibana.

Log data collection for research on the object environment was conducted from June 1, 2022, to July 31, 2022. All network data communication activities by endpoint devices were recorded on the EDR. The logs were collected into the system for further analysis, especially on channels used for E-Government services. In addition, log data is also taken directly from hosts that are indicated to have anomalies that lead to criminal acts based on the results of log analysis.

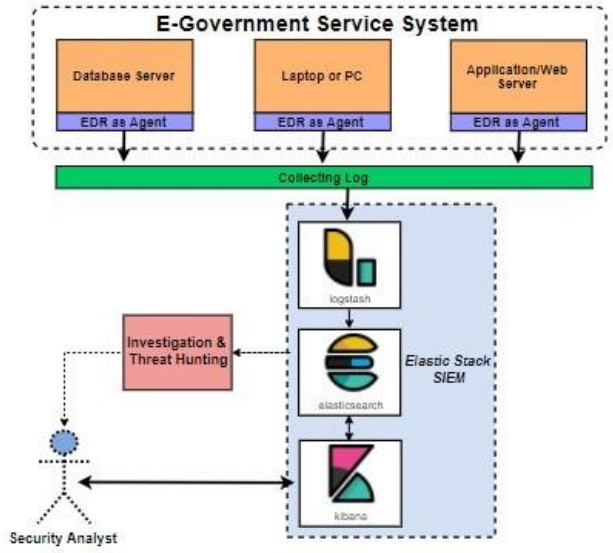


Figure. 5 Testing Scheme

C. Analysis

By filtering the event log data that had been collected, the author further analyzed the data to obtain information that was important for the identification of possible cases.

D. Log Analysis on Secure Shell (SSH) Protocol

Based on the allowed traffic statistics for June 17, 2022, Figure 6 displays that port 22 has the highest usage rate of 73.96% among the top 10 ports authorized for traffic. Port 22 is a standard port used in communication with the Secure Shell (SSH) protocol, so conducting detailed analyses for investigations related to activities on that port is necessary.

Figure 7 illustrates the outcomes of the data filter applied to port 22. The graph displays the extent of traffic activity and distinguishes between secure and potentially hazardous traffic, which may require additional examination. The chart shows a spike in traffic that needs to be studied and analyzed further.

Based on the tracing of the previous analysis results, login attempts were detected, and IP address sent packets on port 22 to many IP addresses (138 sources to 2,780 destinations). Figure 8 illustrates the results of the Kibana visualization, which has identified 788 instances of unsuccessful login attempts and 1689 successful login attempts.

Based on host authentication, three server hosts were detected that were affected by login attempts from 3 source IP Addresses, as shown in Figure 9.

In the following analysis stage, the Secure Shell Protocol (port 22) evaluation process is carried out on one of the hosts (**.**.**.102). The assessment was conducted on June 17, 2022, using log data directly extracted from EDR, depicted in Figure 10.

Figure 10 shows that an IP address sends packets with the SSH protocol (port 22) to many IP addresses at a persistent periodic time. EDR considers this abnormal traffic, so the traffic status is labeled as alert or anomaly and needs further investigation

E. Log Analysis on Service Message Block (SMB) Protocol

According to the traffic data from June 23, 2022, it is evident from Figure 11 that port 445 has the highest utilization rate of 41.61% among the top ten destination ports used in authorized traffic. Port 445 is commonly used for Service Message Block (SMB) protocol communication. Therefore, performing a detailed analysis to evaluate the activities occurring at that port is crucial.

The results of the data filter for port 445 show a graph as shown in Figure 12. From the presented graph, it is apparent that there is an unusual pattern of traffic spikes that warrants further investigation and analysis.

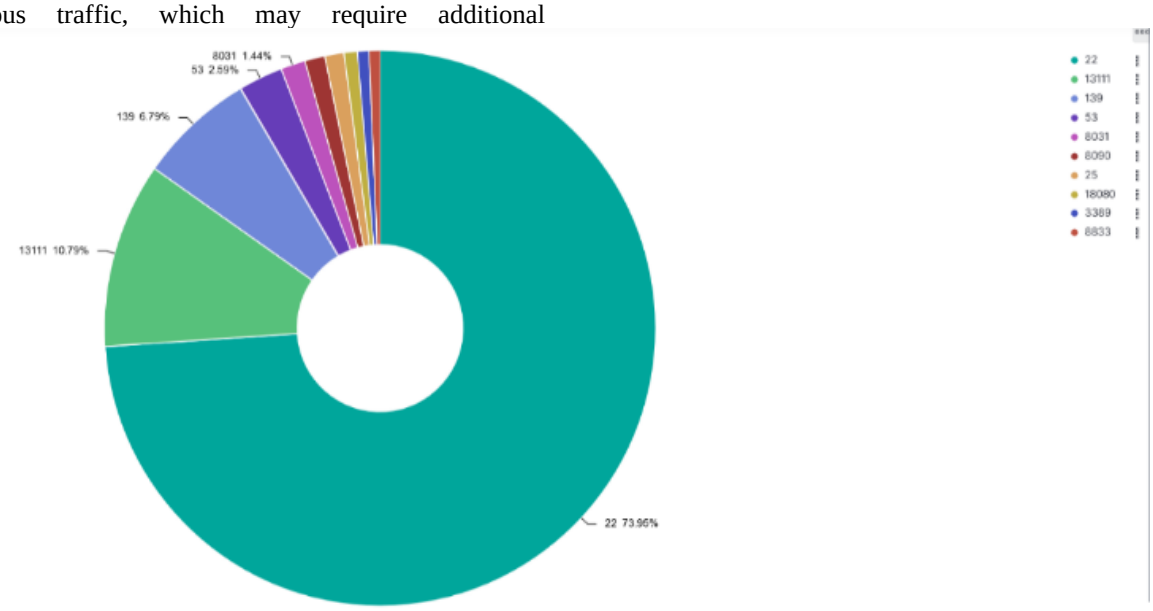


Figure. 6 Port 22 on the List of 10 Allowed Traffic Ports

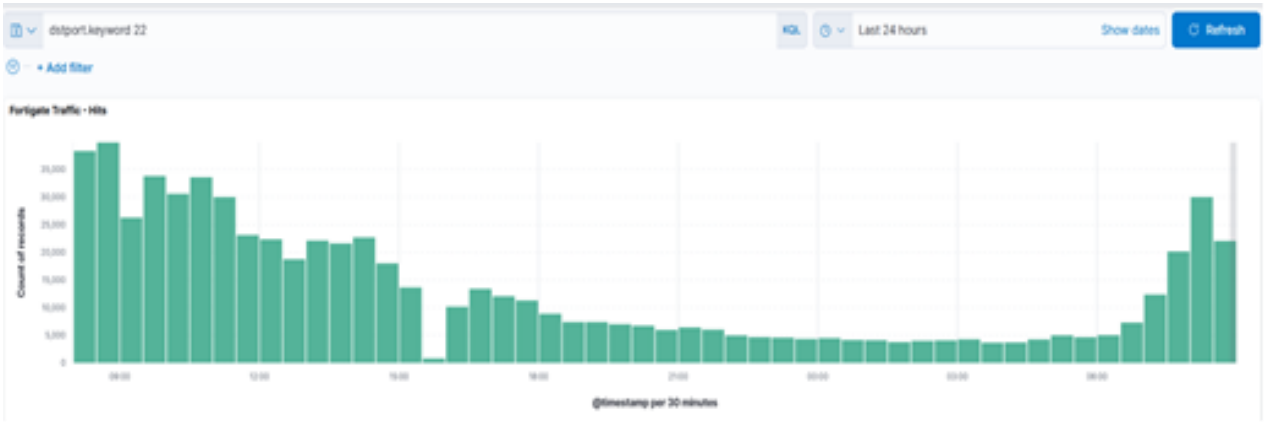


Figure. 7 Traffic Analysis Chart of Port 22

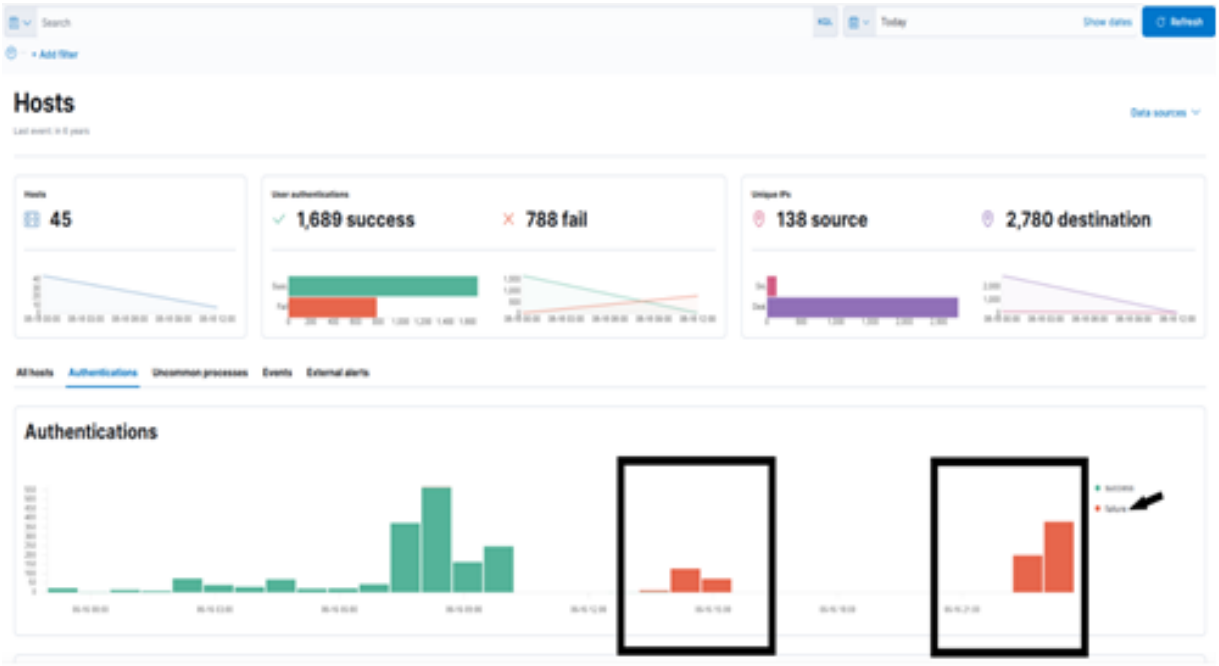


Figure. 8 Authentication Failures During the Login Process on Port 22

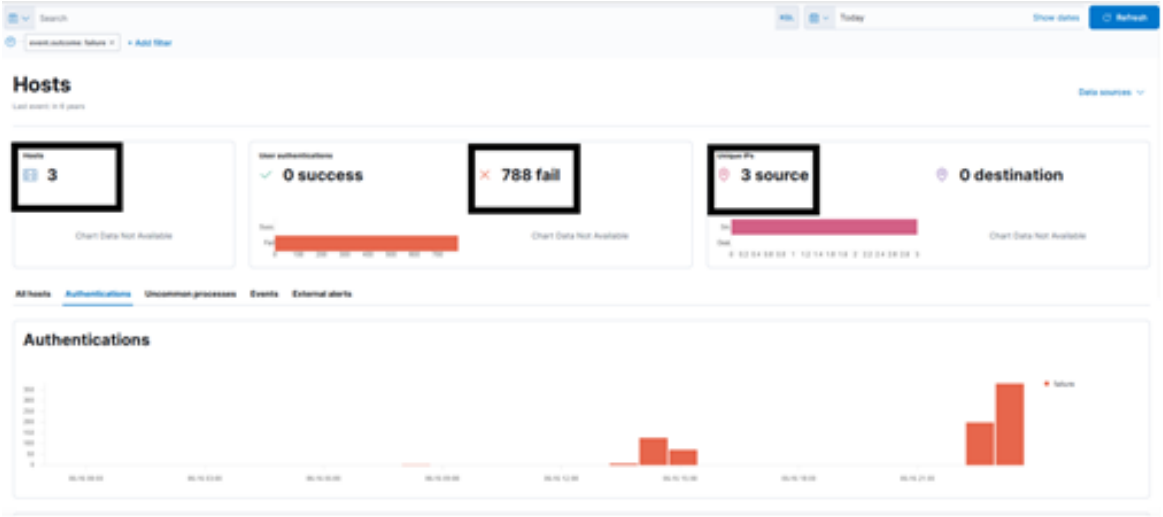
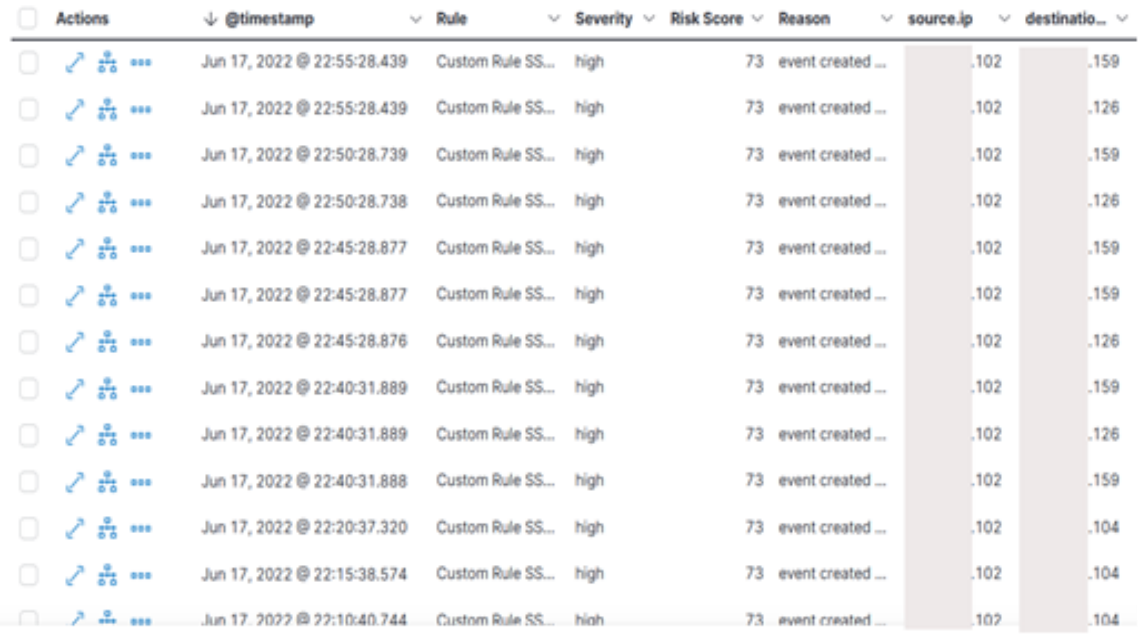


Figure. 9 Three Open Host Server



Actions	@timestamp	Rule	Severity	Risk Score	Reason	source.ip	destinatio...
	Jun 17, 2022 @ 22:55:28.439	Custom Rule SS...	high	73	event created ...	.102	.159
	Jun 17, 2022 @ 22:55:28.439	Custom Rule SS...	high	73	event created ...	.102	.126
	Jun 17, 2022 @ 22:50:28.739	Custom Rule SS...	high	73	event created ...	.102	.159
	Jun 17, 2022 @ 22:50:28.738	Custom Rule SS...	high	73	event created ...	.102	.126
	Jun 17, 2022 @ 22:45:28.877	Custom Rule SS...	high	73	event created ...	.102	.159
	Jun 17, 2022 @ 22:45:28.877	Custom Rule SS...	high	73	event created ...	.102	.159
	Jun 17, 2022 @ 22:45:28.876	Custom Rule SS...	high	73	event created ...	.102	.126
	Jun 17, 2022 @ 22:40:31.889	Custom Rule SS...	high	73	event created ...	.102	.159
	Jun 17, 2022 @ 22:40:31.889	Custom Rule SS...	high	73	event created ...	.102	.126
	Jun 17, 2022 @ 22:40:31.888	Custom Rule SS...	high	73	event created ...	.102	.159
	Jun 17, 2022 @ 22:20:37.320	Custom Rule SS...	high	73	event created ...	.102	.104
	Jun 17, 2022 @ 22:15:38.574	Custom Rule SS...	high	73	event created ...	.102	.104
	Jun 17, 2022 @ 22:10:40.744	Custom Rule SS...	high	73	event created ...	.102	.104

Figure. 10 Secure Shell Protocol Evaluation

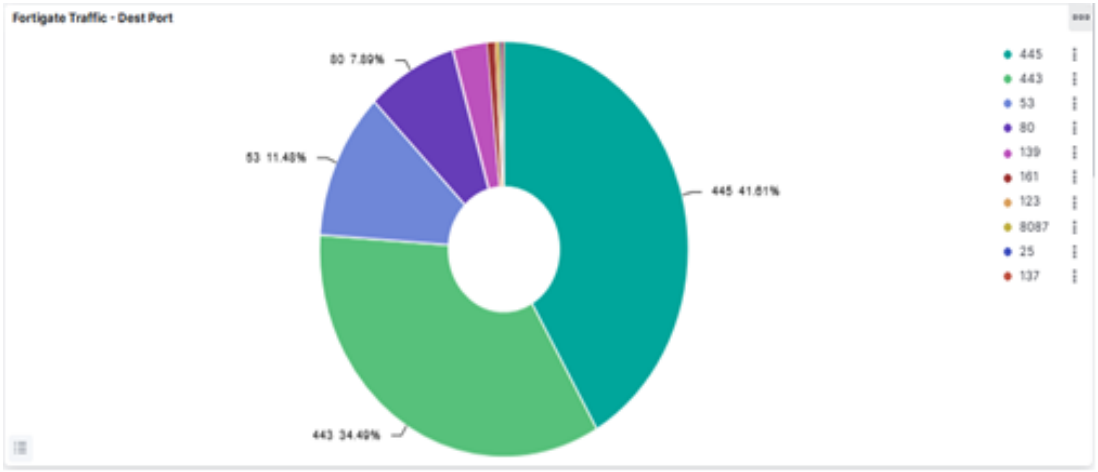


Figure. 11 Port 445 on the List of 10 Allowed Traffic Ports

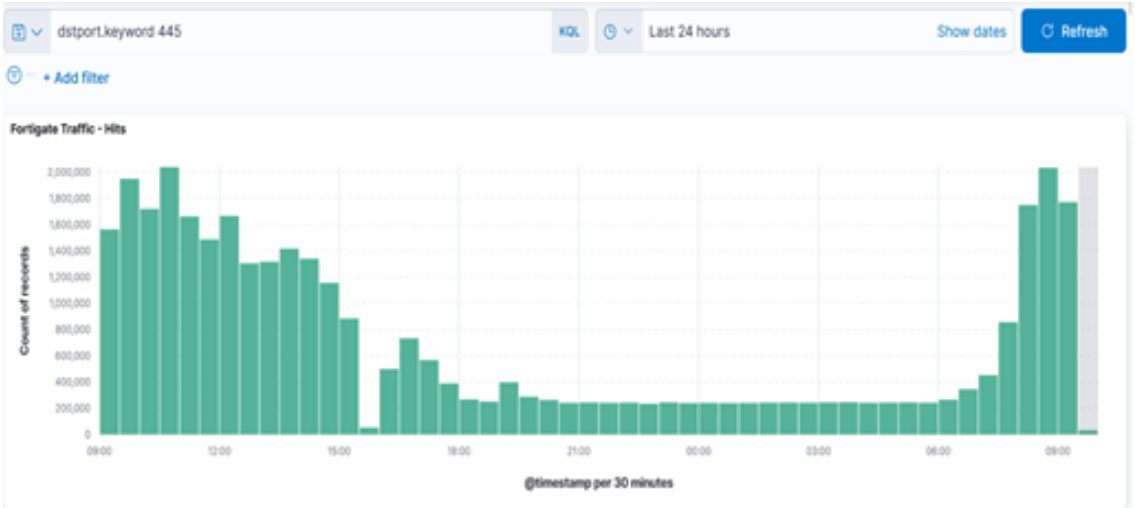


Figure. 12 Traffic Analysis Chart of Port 445

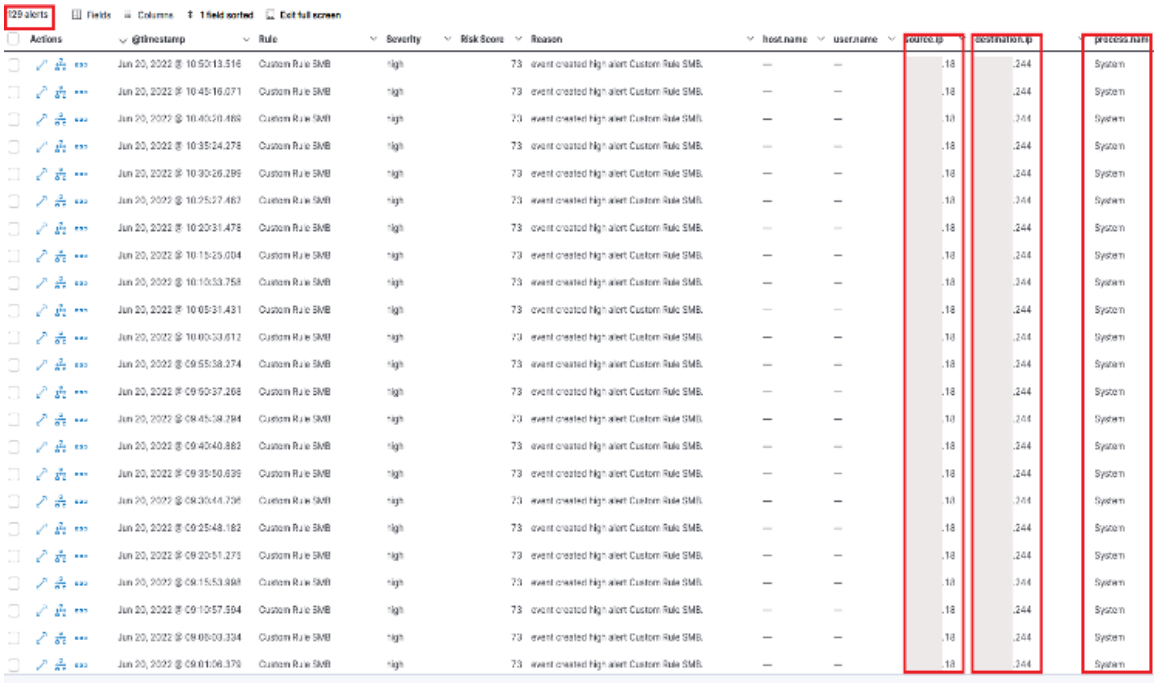


Figure. 13 Evaluation of Service Message Block (SMB) Protocol

During the subsequent analysis phase, the Service Message Block protocol is assessed using the log data gathered from EDR, as illustrated in Figure 13. The figure reveals that on June 20, 2022, an IP address transmitted packets on port 445 to one of the IP addresses at a remarkably consistent interval. Nonetheless, EDR classified this traffic as usual; therefore, the traffic status was not labeled as an alert or anomaly.

By looking at the anomaly traffic obtained through the visualization of the SMB Protocol evaluation, it is necessary to conduct a deeper investigation of potential threats and to find out more clearly about the activities of the service or service that runs the broadcast packet to the IP Address with the specific port 445 used by the SMB service or network sharing.

The investigation is also carried out with analysis on Elasticsearch to see the file service used on port 22, as shown in Figure 15. The PID 183087 label is given to Hydra during the port connection check evaluation process. It can also be seen that the parent process PID 183067 runs Hydra services with the argument: /usr/share/wordlists/metasploit/ipmi_passwords.txt.
Elasticsearch analysis also processed the MD5 hash: 6f93cbdfad177705fd55fa7d37f0b910.

IV. Result and Discussion

A. *SSH Investigation*

A deeper investigation was conducted on host 102 (**.**.**.102), which found anomalies in sending packets with high intensity. While investigating data packet connections on port 22, the author uses Wireshark as a network forensic tool that focuses on protocol analysis [33] to obtain accurate information to determine whether or not there is malicious activity on the victim host, as shown in Figure 14.

Based on the examination of data packet connections in Figure 14, it can be seen that there are IP addresses that SYN to several other IP addresses using port 22. This should be suspected of specific file processes that can threaten security.

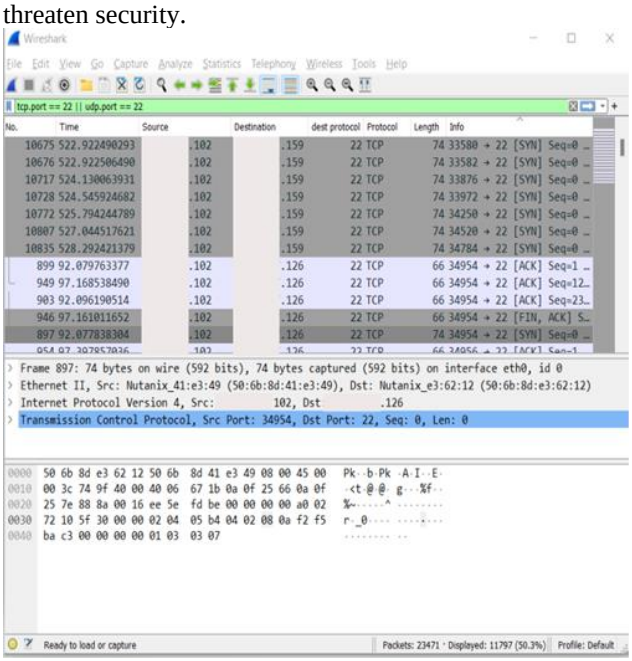


Figure. 14 Port 22 Data Packet Connection Investigation With Wireshark

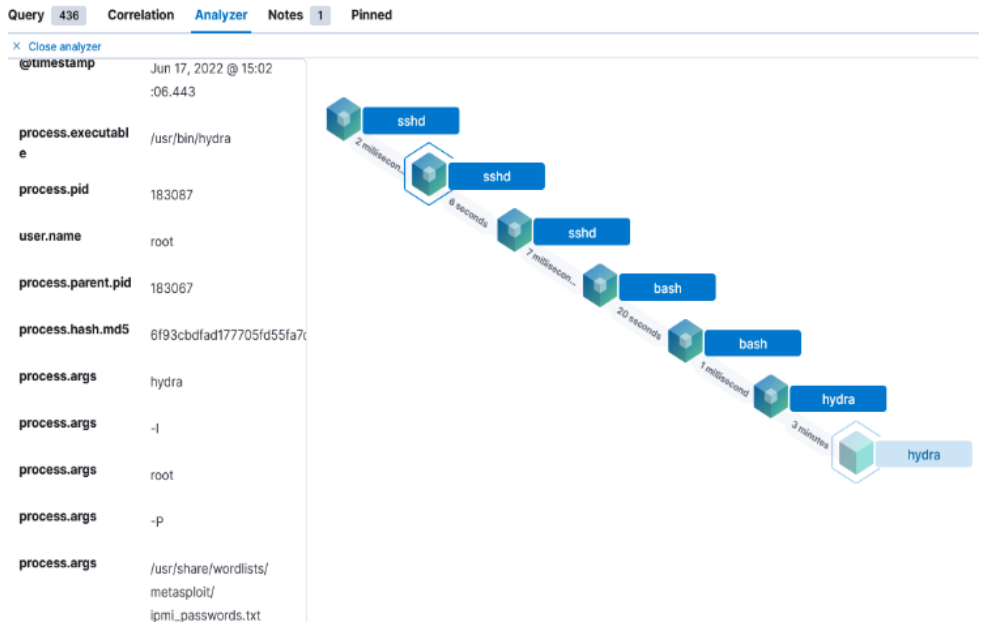


Figure. 15 Analysis of File Service in Elasticsearch

After obtaining the MD5 hash data from the discovered service file, the author cross-checked it for verification purposes using www.virustotal.com, an online security analysis, and a malware threat notification centre [34].

The aim was to determine the likelihood of a potential virus or malware attack, as illustrated in Figure 16. The results of hash matching on www.virustotal.com did not find any malware or specific threats. However, the hash matching results from www.virustotal.com detect that the service is running files with type ELF (Executable and Linkable Format), which is the format of executable files, removable object files, shared libraries, and core dumps.

Based on the hash data analysis search on www.virustotal.com, as shown in Figure 17, the results show that the application that runs the ELF format file is THC Hydra, a tool commonly used by hackers for password cracking [35].

This attack with THC Hydra utilizes a brute force method that uses a dictionary containing many passwords in trying passwords or usernames to enter a system illegally [35].

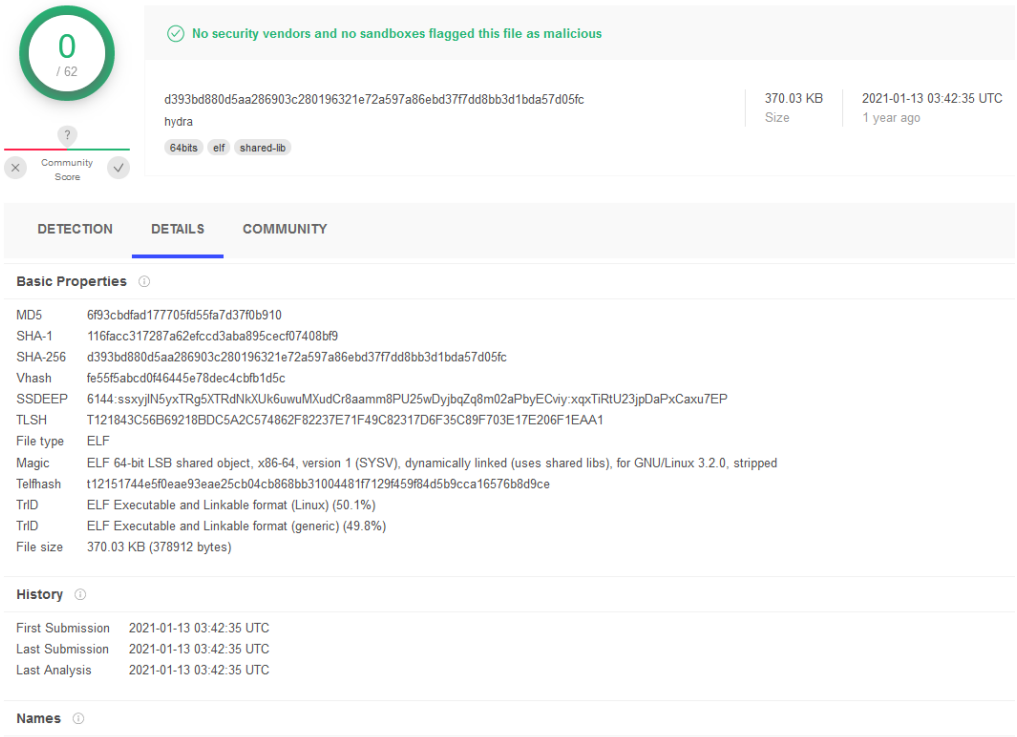


Figure. 16 ELF File Hash Identification on www.virustotal.com

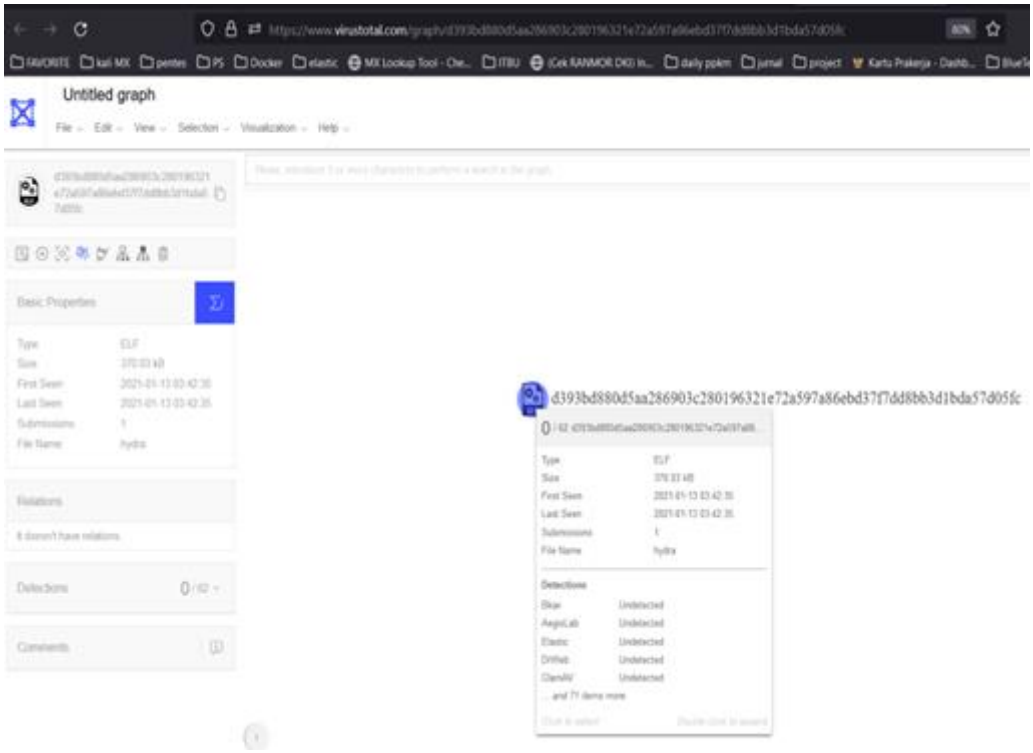


Figure. 17 Harmful ELF File Hash Data Matching on www.virustotal.com

B. SMB Investigation

To investigate the SMB protocol, the author collected raw logs from EDR on hosts with suspected anomalies, as indicated in Figure 18.

The investigation on the host was conducted using memory analysis with Mandiant Redline, a Windows-based memory forensics analysis and acquisition tool that facilitates the host investigation process [36], as illustrated in Figure 21.


```
{
  "_index": ".siem-signals-default-000006",
  "_type": "_doc",
  "_id": "da744569ed6b5fc9c64a808e8f172e3014099aaa030a69302f33d6f1aa074d38",
  "_score": 1,
  "_source": {
    "@timestamp": "2022-06-20T03:50:13.516Z",
    "destination.ip": "       .244",
    "process.name": "System",
    "source.ip": "       18",
    "event": {
      "kind": "signal"
    },
    "signal": {
      "_meta": {
        "version": 57
      },
      "parents": [
        {
          "id": "853a0119-e7f6-5d5d-90be-aaaaa957a750",
          "type": "event",
          "index": "auditbeat-*,logs-*,logs-endpoint.events",
          "depth": 0
        }
      ],
      "ancestors": [
```


Figure. 18 SMB Raw Logs from EDR

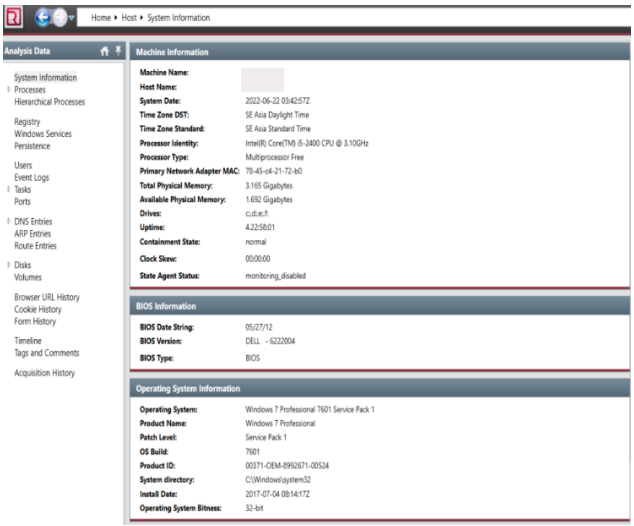


Figure. 19 Analysis of the host using Mandiant Redline

Based on Mandiant Redline analysis, it was found that there was suspicious activity running an outside service with the location of Port 445 by the System.exe application, as shown in Figure 20.

Based on the process hierarchy trace, it is found that there is a sequential malicious activity by System.exe with PID 4, which runs another application service named csrss.exe with PID 480 using commands in the "C:\Windows\system32" directory.

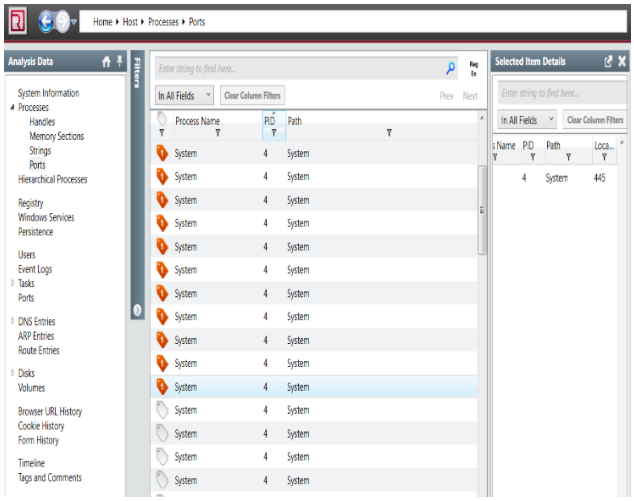


Figure. 20 System.exe Anomaly Activity Analysis

The results of further hierarchical process analysis show that the csrss.exe process runs an abnormal service (anomaly) with the argument: *"%SystemRoot% \system32\csrss.exe ObjectDirectory=\Windows SharedSection=1024,12288,512 Windows=On SubSystemType=Windows ServerDll=basesrv,1 ServerDll=winsrv: UserServerDllInitialization,3 ServerDll=winsrv:ConServerDllInitialization, 2ServerDll=sxssrv,4 ProfileControl=Off MaxRequestThreads=16"* as shown in Figure 21 With the Mandiant Redline forensic tool, the hash of the csrss.exe file can be determined for further investigation.

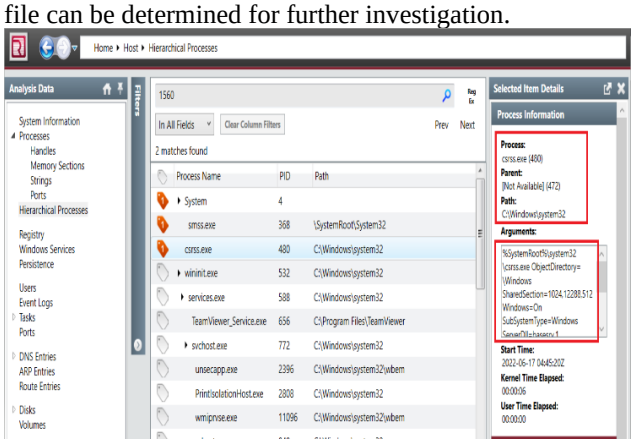


Figure. 19 System.exe Anomaly Running Service File csrss.exe

By matching the hash data owned by the csrss.exe file on www.virustotal.com, no security analysis detects that the file is malicious or a virus, as illustrated in Figure 22.

However, after further analysis in the "*exe*" category at www.virustotal.com for the csrss.exe file, it shows that the file has a relationship with another similar file with the name csrss.exe.exe on specific hosts that are connected and detected as malicious by several security analyzers based on searches in the distribution map in Figure 23 and Figure 24.

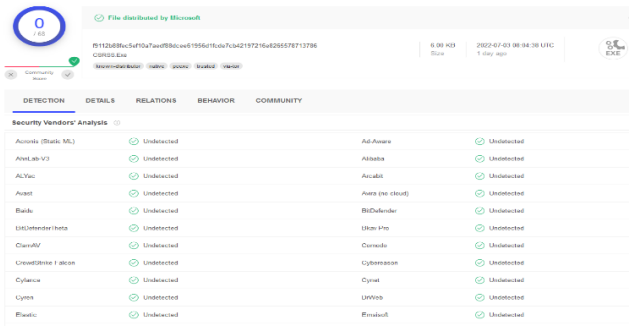


Figure. 22 An Examination of the csrss.exe File Using Hash Analysis was Performed on www.virustotal.com

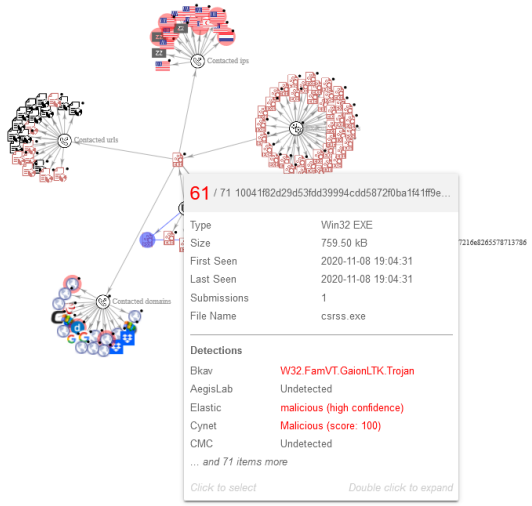


Figure. 23 Analyzed The csrss.exe file using Distribution Map on www.virustotal.com

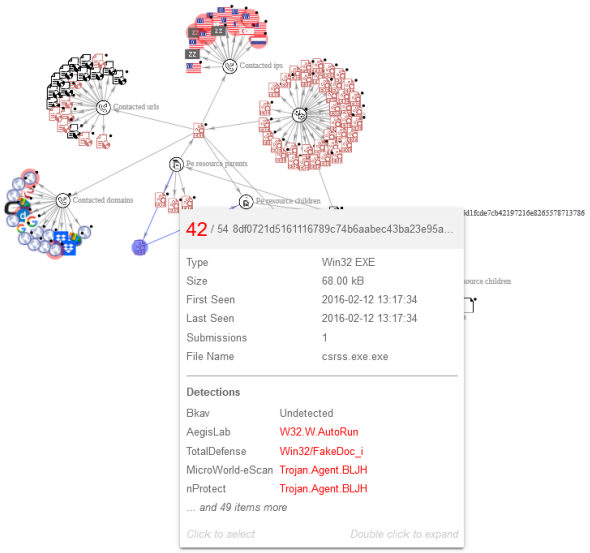


Figure. 20 File Distribution Map of csrss.exe.exe Related to csrss.exe

The hash data on the csrss.exe.exe file associated with csrss.exe is taken to be matched again based on the security analysis results at www.virustotal.com. According

to the security analysis conducted on the hash data of the csrss.exe file on www.virustotal.com, the findings indicate that the file is identified as either a Trojan-type virus or malicious software (malware), as depicted in Figure 25.

The author also revalidates the hash of the csrss.exe file installer found from the victim host at www.virustotal.com. The results are identical to the analysis of the csrss.exe.exe file identified as a Trojan virus.

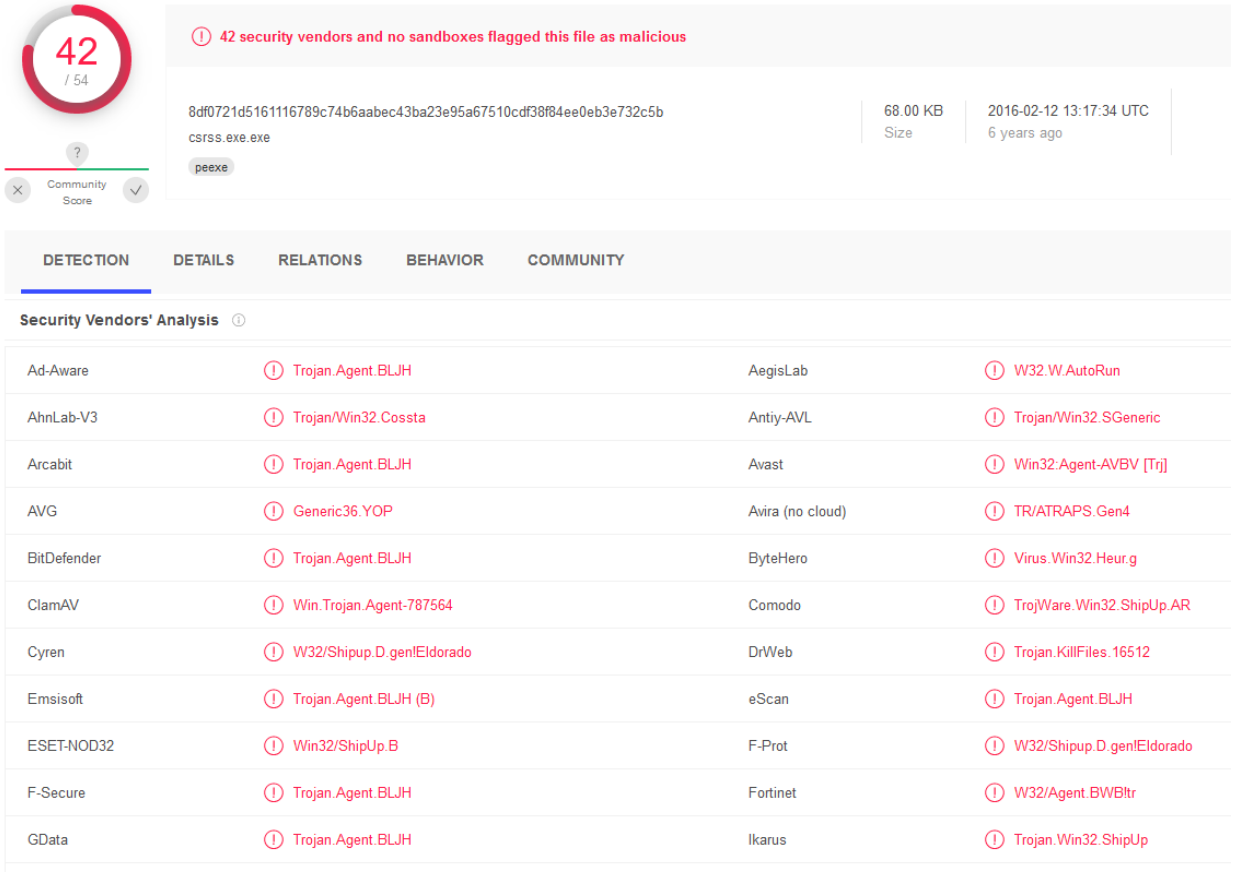


Figure. 21 Hash Data Identification of csrss.exe.exe File Detected as Malware

C. Results

With the techniques used in this experiment, cybercrime detection in E-Government is more straightforward for security analysts. The system will display the results of monitoring several events and occurrences in real-time by setting filters for several specified analysis logs. The results displayed graphically are pretty attractive and accurate, making it easier for security analysts to investigate each event and incident, and can make it a tool for proving security checks.

Upon analyzing the logs and investigating the findings through a digital forensic approach, this study has identified cybercrime attempts and activities targeting the web-based E-Government service network system, namely:

- Brute force attack via SSH protocol: the evidence from the detection results depicted in Figure 17 revealed that on June 17, 2022, at 15:02, a brute force attack was conducted via SSH protocol using the THC Hydra tool with the argument process: /usr/share/wordlists/ metasploit/ipmi_ passwords.txt.
- Trojan virus attack via SMB protocol: proof of detection results found a malware attack on one of the hosts that utilize the SMB protocol with the file name csrss.exe, with the argument: "%SystemRoot%\system32\csrss.exe ObjectDirectory=\WindowsShared Section=1024,12288,512 Windows=On SubSystemType=WindowsServer Dll=basesrv,1

  ServerDll=winsrv:UserServerDllInitialization,3 ServerDll=winsrv:ConServerDllInitialization,2 ServerDll=sxssrv,4 ProfileControl=Off MaxRequestThreads=16". By investigating using a digital forensics approach, the file was a Trojan virus, as illustrated in Figure 25.

V. CONCLUSION

Based on the test results, the SIEM-based centralized log data analysis model using Elastic Stack can effectively detect cybercrime threats on E-Government service network systems with relatively fast and accurate results. However, the investigation stage with a digital forensic approach using several analysis techniques is still needed to validate and identify the detection results of the proposed model, so the existing process must get knowledge assistance from security analysts in its implementation.

REFERENCES

[1] J. Ahmed and Q. Tushar, "Covid-19 Pandemic: A New Era Of Cyber Security Threat And Holistic Approach To Overcome," 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), 2020, pp. 1-5, doi: 10.1109/CSDE50874.2020.9411533.

[2] A. A. Najar and M. Naik S, "Covid-19 Impact on Cyber Crimes in India: A Systematic Study," 2022 IEEE India Council International Subsections Conference (INDISCON), 2022, pp. 1-8, doi: 10.1109/INDISCON54605.2022.9862935.

[3] S. Hakak, W. Z. Khan, M. Imran, K. -K. R. Choo and M. Shoaib, "Have You Been a Victim of COVID-19-Related Cyber Incidents? Survey, Taxonomy, and Mitigation Strategies," in IEEE Access, vol. 8, pp. 124134-124144, 2020, doi: 10.1109/ACCESS.2020.3006172.

[4] J. P. Kesan and L. Zhang, "An Empirical Investigation of the Relationship between Local Government Budgets, IT Expenditures, and Cyber Losses," in IEEE Transactions on Emerging Topics in Computing, vol. 9, no. 2, pp. 582-596, 1 April-June 2021, doi: 10.1109/TETC.2019.2915098.

[5] Huda, Mirojul & Yunas, Novy, "The Development of e-Government System in Indonesia", Jurnal Bina Praja. 08., 2016, pp. 97-108, doi: 10.21787/JBP.08.2016.97-108.

[6] O. S. Al-Mushayt, "Automating E-Government Services With Artificial Intelligence," in IEEE Access, vol. 7, pp. 146821-146829, 2019, doi: 10.1109/ACCESS.2019.2946204.

[7] A. J. Horta Neto and A. Fernandes Pereira dos Santos, "Cyber Threat Hunting Through Automated Hypothesis and Multi-Criteria Decision Making," 2020 IEEE International Conference on Big Data (Big Data), 2020, pp. 1823-1830, doi: 10.1109/BigData50022.2020.9378213.

[8] S. Byeon and W. Suh, "A Study on the Government's Countermeasures Against Cyber Attacks," 2020 IEEE International Conference on Big Data and Smart Computing (BigComp), 2020, pp. 495-499, doi: 10.1109/BigComp48618.2020.00-17.

[9] A. A. Ali and M. Zamri Murah, "Security Assessment of Libyan Government Websites," 2018 Cyber Resilience Conference (CRC), 2018, pp. 1-4, doi: 10.1109/CR.2018.8626862.

[10] J. P. Kesan and L. Zhang, "An Empirical Investigation of the Relationship between Local Government Budgets, IT Expenditures, and Cyber Losses," in IEEE Transactions on Emerging Topics in Computing, vol. 9, no. 2, pp. 582-596, 1 April-June 2021, doi: 10.1109/TETC.2019.2915098.

[11] Zhang, Y., Xiao, Y., Ghaboosi, K., Zhang, J. and Deng, H. (2012), A survey of cyber crimes. Security Comm. Networks, 5: 422-437. https://doi.org/10.1002/sec.331

[12] G. Tsakalidis and K. Vergidis, "A Systematic Approach Toward Description and Classification of Cybercrime Incidents," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 49, no. 4, pp. 710-729, April 2019, doi: 10.1109/TSMC.2017.2700495.

[13] W. A. Al-Khater, S. Al-Maadeed, A. A. Ahmed, A. S. Sadiq, and M. K. Khan, "Comprehensive Review of Cybercrime Detection Techniques," in IEEE Access, vol. 8, pp. 137293-137311, 2020, doi: 10.1109/ACCESS.2020.3011259.

[14] M. Xu, K. M. Schweitzer, R. M. Bateman, and S. Xu, "Modeling and Predicting Cyber Hacking Breaches," in IEEE Transactions on Information Forensics and Security, vol. 13, no. 11, pp. 2856-2871, Nov. 2018, doi: 10.1109/TIFS.2018.2834227.

[15] B. Arief and M. A. Bin Adzmi, "Understanding Cybercrime from Its Stakeholders' Perspectives: Part 2--Defenders and Victims," in IEEE Security & Privacy, vol. 13, no. 2, pp. 84-88, Mar.-Apr. 2015, doi: 10.1109/MSP.2015.44.

[16] B. Arief, M. A. B. Adzmi and T. Gross, "Understanding Cybercrime from Its Stakeholders' Perspectives: Part 1--Attackers," in IEEE Security & Privacy, vol. 13, no. 1, pp. 71-76, Jan.-Feb. 2015, doi: 10.1109/MSP.2015.19.

[17] A. A. Ahmed and Y. W. Kit, "Collecting and Analyzing Digital Proof Material to Detect Cybercrimes," 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech), 2018, pp. 742-747, doi: 10.1109/DASC/PiCom/DataCom/CyberSciTec.2018. 00129.

[18] A. Sørensen, M. J. Remy, N. Kjettrup, R. V. Mahmoud and J. M. Pedersen, "An Approach to Detect and Prevent Cybercrime in Large Complex Networks," 2018 International Conference on Cyber Security and Protection of Digital Services (Cyber Security), 2018, pp. 1-8, doi: 10.1109/CyberSecPODS.2018.8560687.

[19] P. He, J. Zhu, S. He, J. Li, and M. R. Lyu, "Towards Automated Log Parsing for Large-Scale Log Data Analysis," in IEEE Transactions on Dependable and Secure Computing, vol. 15, no. 6, pp. 931-944, 1 Nov.-Dec. 2018, doi: 10.1109/TDSC.2017.2762673.

[20] Y. Xie, K. Yang, and P. Luo, "LogM: Log Analysis for Multiple Components of Hadoop Platform," in IEEE Access, vol. 9, pp. 73522-73532, 2021, doi: 10.1109/ACCESS.2021.3076897.

[21] S. Locke, H. Li, T. -H. P. Chen, W. Shang, and W. Liu, "LogAssist: Assisting Log Analysis Through Log Summarization," in IEEE Transactions on Software Engineering, vol. 48, no. 9, pp. 3227-3241, September 1 2022, doi: 10.1109/TSE.2021.3083715.

[22] O. Podzins and A. Romanovs, "Why SIEM is Irreplaceable in a Secure IT Environment?," 2019 Open Conference of Electrical, Electronic and Information Sciences (eStream), 2019, pp. 1-5, doi: 10.1109/eStream.2019.8732173.

[23] S. Kobayashi, K. Otomo, K. Fukuda, and H. Esaki, "Mining Causality of Network Events in Log Data," in IEEE Transactions on Network and Service Management, vol. 15, no. 1, pp. 53-67, March 2018, doi: 10.1109/TNSM.2017.2778096.

[24] J. Lee, J. Kim, I. Kim and K. Han, "Cyber Threat Detection Based on Artificial Neural Networks Using Event Profiles," in IEEE Access, vol. 7, pp. 165607-165626, 2019, doi: 10.1109/ACCESS. 2019.2953095.

[25] N. Afzaliseresht, Y. Miao, S. Michalska, Q. Liu, and H. Wang, "From logs to Stories: Human-Centred Data Mining for Cyber Threat Intelligence," in IEEE Access, vol. 8, pp. 19089-19099, 2020, doi: 10.1109/ACCESS.2020.2966760.

[26] F. Özdemir Sönmez and B. Günel, "Evaluation of Security Information and Event Management Systems for Custom Security Visualization Generation," 2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT), 2018, pp. 38-44, doi: 10.1109/IBIGDELFT.2018.8625291.

[27] H. Mokalled, R. Catelli, V. Casola, D. Debertol, E. Meda and R. Zunino, "The Applicability of a SIEM Solution: Requirements and Evaluation," 2019 IEEE 28th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), 2019, pp. 132-137, doi: 10.1109/WETICE.2019.00036.

[28] Florian Menges, Tobias Latzo, Manfred Vielberth, Sabine Sobola, Henrich C. Pöhls, Benjamin Taubmann, Johannes Köstler, Alexander Puchta, Felix Freiling, Hans P. Reiser, Günther Pernul," Towards GDPR-compliant data processing in modern SIEM systems", Computers & Security, Volume 103,

2021,102165, ISSN 0167-4048, https://doi.org/10.1016/j.cose.2020.102165.

[29] Elastic Corporation, "ES," Elastic Corporation, (24/10/2021). What is the ELK Stack? Why, it's the Elastic Stack. Available: https://www.elastic.co/what-is/elk-stack [Accessed: 30-March-2022]

[30] I. Kotenko, A. Kuleshov and I. Ushakov, "Aggregation of elastic stack instruments for collecting, storing and processing of security information and events," 2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), 2017, pp. 1-8, doi: 10.1109/UIC-ATC.2017.8397627.

[31] F. Mulyadi, L. A. Annam, R. Promya and C. Charnsripinyo, "Implementing Dockerized Elastic Stack for Security Information and Event Management," 2020 - 5th International Conference on Information Technology (InCIT), 2020, pp. 243-248, doi: 10.1109/InCIT50588.2020.9310950.

[32] W. U. Hassan, A. Bates, and D. Marino, "Tactical Provenance Analysis for Endpoint Detection and Response Systems," 2020 IEEE Symposium on Security and Privacy (SP), 2020, pp. 1172-1189, doi: 10.1109/SP40000.2020.00096.

[33] S. Sandhya, S. Purkayastha, E. Joshua and A. Deep, "Assessment of website security by penetration testing using Wireshark," 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2017, pp. 1-4, doi: 10.1109/ICACCS.2017.8014711.

[34] C. Rathnayaka and A. Jamdagni, "An Efficient Approach for Advanced Malware Analysis Using Memory Forensic Technique," 2017 IEEE Trustcom/BigDataSE/ICESS, Sydney, NSW, Australia, 2017, pp. 1145-1150, doi: 10.1109/Trustcom/BigDataSE/ICESS. 2017.365.

[35] T. Kakarla, A. Mairaj and A. Y. Javaid, "A Real-World Password Cracking Demonstration Using Open Source Tools for Instructional Use," 2018 IEEE International Conference on Electro/Information Technology (EIT), Rochester, MI, USA, 2018, pp. 0387-0391, doi: 10.1109/EIT.2018.8500257.

[36] W. Ahmed and B. Aslam, "A comparison of Windows physical memory acquisition tools," MILCOM 2015 - 2015 IEEE Military Communications Conference, Tampa, FL, USA, 2015, pp. 1292-1297, doi: 10.1109/MILCOM.2015.7357623.