Received: September 15, 2022. Revised: October 15, 2022.

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Secure Mechanism Applied to Big Data for IIoT by Using Security Event and Information Management System (SIEM)

Marwan Alaa Hussein[1]* **Ekhlas Kadhum Hamza[1]**

[1]*Department of Control and Systems Engineering, Technology University, Baghdad, Iraq*
* Corresponding author's Email: Ekhlas.K.Hamza@uotechnology.edu.iq

Abstract: It is estimated that the number of devices and sensors connected to the Internet of Things (Internet of Things) will grow to around 125 billion by the end of this decade, compared to 21 billion this year. The Internet of Things promises tremendous advantages in many applications such as industrial environment, smart homes, smart cities, smart environment, agriculture, control of critical infrastructure and smart health. However, as the number of IoT devices increases and more information is shared between IoT devices, massive amount of data is transmitted between these devices and providing security becomes a major concern for researchers, developers and users, since IoT devices have low power and limited computing and storage capabilities. Where the application of strong and complex encryption processes requires significant capabilities in terms of computing and storage, which makes these devices more vulnerable to attacks and security risks that threaten the integrity of corporate and institutional data and other information. This article proposes implementing a security solution based on the "all-in-one" architecture for Wazuh and Elastic Stack as a tester, in order to implement proof of concept to detect anomalies occurring in devices on a network, which constitute the Wazuh proxy. In this way, the security contribution proactively with the collection of logs in real time, allows this system in question to generate alerts in the event of attempted attacks and implement the active response, a measure that allows mitigation of the detected incident. This project promotes open-source software solutions, and proves to be a complete business security solution in the context of analysing log data to secure a host for the internal business network. He concluded that the solution is ideal for business environments of any type, and even more so for small environments such as our simulated environments. Considering that the method of automating responses to security incidents offers a great alternative in the field of information technology.

Keywords: Industrial IoT, Big data, Cybersecurity, SIEM, WAZUH.

1. Introduction

In the last decade, organizations and companies have been immersed in sophisticated and diverse attacks on computer systems, compromising their data and computer assets, forcing them to focus their attention on information management and security. The massive use of information and communication technologies (ICT) has caused the dependence of society on these, companies do not escape this reality. In particular, today the use of that technology, including the Internet, makes people's lives easier, but unfortunately, they are not very familiar with security; thereby increasing the possibility of attacks [1]. This scenario complicates the problems for security systems due to the absence of efficient and effective controls in the use of ICT, and the lack of legal regulations that regulate the use of cyberspace, which leads to the continuous increase of vulnerabilities in the information assets of organizations [2].

As security log anomaly detection increasingly receives increased attention, authentication events constitute an important component of these logs, and with them reliable and accurate predictions can be produced, minimizing the effort of cyber-experts to stop fake attacks [3].

The internet of things (IoT) is a network of interconnected devices that may communicate with one another and deliver information to consumers over the Internet. Recent IoT growth has been fueled

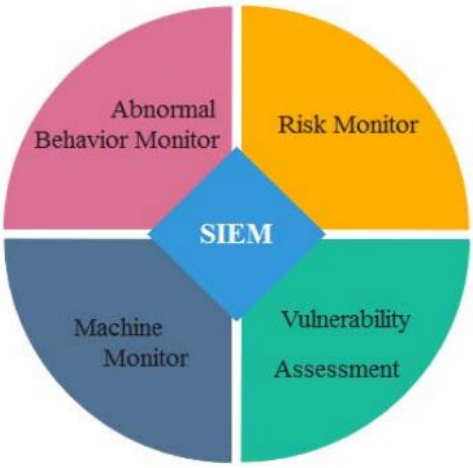

Figure. 1 Security information and event management

in part by its broad applicability, scalability, and help for intelligent applications. The majority of IoT apps perform functions automatically, with minimal or no human interaction[4].

Industrial IoT (IIoT) is a subcategory of the IoT in which IoT devices are primarily employed in closed industrial contexts. IIoT has been successful in generating substantial resources and boosting productivity [1]. The IIoT is a crucial enabler for Industry 4.0, often known as the Fourth Industrial Revolution [5]. The number of IoT-connected devices is predicted to reach 41 billion by 2027, there are already more than 8 billion. The global IoT market size was forecast to exceed $380 billion in 2021 and is anticipated to surpass $1.8 trillion by 2028, expanding at a CAGR of 25.4% from 2021 to 2028 [6]. According to Kaspersky researchers, the number of cyberattacks against IoT devices increased by more than 100 percent in a single year (2020-2021), from 639 million to 1.5 billion. Cybercriminals have turned their attention to this space in order to steal data, mine cryptocurrencies, and create botnets. The cybersecurity sector has made substantial progress in developing sophisticated security tools and approaches for protecting users and data in traditional IT systems. However, IoT/IIoT-based systems can not immediately take these actions. Numerous existing strategies are insufficient to combat novel threats that can compromise IoT networks, necessitating a deeper dive into advanced forensic techniques to detect and analyses malicious behavior [7].

As the number of IoT devices increases and more information is shared between them, their susceptibility to cyberattacks increases. Complex cyberattacks can be executed by exploiting internet of things infrastructure vulnerabilities. These cyberattacks include denial-of-service (DoS/DDoS), man-in-the-middle (MITM), brute force, replay, information gathering, injection, malware, and security and privacy threats. It is difficult to incorporate security protections into small IoT devices due to their limited resources, making them more susceptible to various attacks. Attackers can manipulate system data by exploiting vulnerabilities in IoT systems[8, 9].

From the above, this research arises that aims to propose an architecture of data analysis through Big Data tools through the use of events or security logs, which allow improving the identification, integration and correlation of security events [10].

Security information and event management (SIEM) combine security information management (SIM) and security event management (SEM). It can analyze, audit, discover and anticipate hardware, network and all application events in real time. Specifically, SIEM software products and services provide real time analysis of security alerts generated by network hardware and applications. SIEM aggregate data from software, equipment or hosting services, analyze and manipulate log data and security records, as well as report security compliance events. SIEM's goal is to support organizations more quickly respond to attacks and links events together into meaningful bundles. A key focus is to monitor and help manage user and service privileges, directory services and other system-configuration changes; as well as providing log auditing and review and incident response.[11]

The term security information event management (SIEM), coined by mark nicolett and amrit williams of gartner in 2005. [11] It is the first time to combine SIM and SEM together. SIM focuses on internal control, which monitors the authorized parties' behavior and the access to internal resource, and provides the compliance management. Meanwhile SEM deals with internal and external behavior monitoring, emergency response to security incident, which focuses more on security itself. A full SIEM solution should have following functions [12]:

a. the product capabilities of gathering, analysing and presenting information from network and security devices [12] identity and access-management applications [12].
b. vulnerability management and policy-compliance tools [12].
c. operating-system, database and application logs[12].
d. external threat data [12].

The traditional log system has lots of problems:

a. Unfriendly user interface: lack of overall view about the security threats and risks which the organization is currently facing, the compliance status is not very clear;

b. Poor real-time ability: the security team cannot real-time detect security threats and risks happening in the organization. Often get incident information after user complaints;
c. Difficult to collect evidence–when the investigation after incident occurred, security team has to search within lots of security logs and production records.

Different format and different equipment require many members support across different departments;
d. Hard investigation - the root cause is often hard to be found out in the incident investigation;
e. Incomplete log chain – sometimes lack of some log data leads to be unable to provide compliance evidence when facing audit;
f. Guardian self-theft - IT staff, production management personnel can modify or delete log easily. SIEM (Security information and event management) is the best solution to free us from the passive "firefighting" mode and solve the traditional log system's problems. It aggregates data from different sources: events, users, systems, applications and risk (Fig. 1), so that organizations can link all security resources together, real-time identify security attacks or unauthorized behaviours, and have a full view of security status in whole network. At present SIEM products have two types: commercial products and open-source tools. There are some difficulties to implement open source SIEM in industry companies:
a. Hard to solve dependency problems between multi open-source software and integrate subsystems as a whole SIEM system;
b. Repeated authentication between different subsystems and each subsystem interface has different style;
c. Hard to share data between subsystems;
d. Hard to correlate data from different sources;
e. Reporting formats cannot unify;
f. Lack of a central dashboard to present critical monitoring information;
g. Unable to detect the network threat timely;
h. Heavy loading to maintain multi subsystems;
It may be feasible for those enterprises that have their own strong development teams, such as internet enterprise, E-commerce companies, to develop a SIEM system based on open-source tools. But for manufacturing companies, to choose one commercial product and do customization is a more efficient and more reliable way to build up a SIEM platform.

In this project, we select (elastic stack and its main components (Elasticsearch, Logstash, Kibana), and technologies such as Filebeat and Wazuh security detection, managed security in information assets such as communications equipment, data servers and applications, database engines, and end-user terminals) to build the SIEM system. The tool is able not only to read data from traditional IT terminals, but also to collect and index machine data from any source. It is very suitable for manufacturing companies that have lots different kinds of production machines. The tool delivers powerful search, analysis and visualization capabilities and decision maker can get valuable data from it, as well as IT operations become more convenient and clearer: a. Simple: it provides a Google-style search engine, a same entrance, for all department users. Learning is easy, maintenance cost is low. Just as Pareto principle, use 20% time to learning and obtain 80% information through the SIEM platform. So it's very suitable for a large number of different levels of users to use; b. Easy to realize functions SIEM required: correlation analysis, alert, event records, blacklist, and etc.; c. Cross-platform and reporting functions.

The above aspects and challenges are discussed and effective solutions provided. Several related studies in the same field as this work were conducted, and they are as follows in Table. 1

2. Theoretical framework Generalities

The discovery of useful information is an important topic among researchers because it represents the cornerstone in a progressive society. We live today in the age of computers and the Internet, a time when data accumulates through exponential growth every second. Therefore, it is the responsibility of the computer science researcher to invent a better way to obtain useful insights from this large-scale data or big data [21].

Currently, the use of the term big data tends to refer to the analysis of user behaviour, giving value to the stored data and formulating predictions through the observed behaviour patterns. This discipline, dedicated to the study of big data, is part of the ICT sector [22].
Big data deals with all activities related to systems that manipulate huge data sets, the most common difficulties being those linked to the management of these amounts of data that focus on collection and storage, search, sharing, analysis, and visualization [23]. The "management of big data", also called "data intelligence", "large-scale data" or "Big Data", are the terms that refer to the data sets so large and complex, that non-traditional computer applications of data processing are needed to treat them properly [10].

Table 1. Summary comparison for the pertinent studies

Ref.	Authors	year	Methodology	Security	Cloud Computing	Technology
[14]	D. D. Khudhur and M. S. Croock,	2017	Implement SSL/TLS encryption with MQTT-based Internet of Things (IoT) protocol to secure communication and encrypt data exchange between the system. The second aspect is an algorithm to detect and prevent a Denial of Service (DoS) attack.	Y	N	secure sockets layer (SSL)/transport layer security (TLS) encryption
[15]	N. A-hussein and A. D. Salman	2018	The protocol MQTT to exchange sensor information between two different devices. Node-Red and Thing Speak were designed as a website to share the data.	N	N	Uses the Esp8266 , Raspberry pi 3 and Node-Red and Thing Speak were designed as a website to share the data
[16]	H. Sun, H. Yu, et al.	2020	They have devised a fog-cloud enabled IoT architecture that has used the best features of fog and cloud	Y	Y	fog-cloud and ETCORA algorithm
[17]	Alzubi et al	2020	They have devised a scheme to provide security for IIoT data transfer using cloud services with a Hashed Needham Schroeder (HNS) cost-optimized deep machine Learning (CODML) technique that shows the need to deliver IIoT security securer channel.	Y	Y	a Hashed Needham Schroeder (HNS) cost-optimized deep machine Learning (CODML) technique
[18]	V. Teeraratchakarn and Y. Limpiyakorn	2020	The author explained how Elastic Stack applies security monitoring and analytics. Elastic stack, formerly known as ELK	Y	N	Used ELK
[19]	F. Mulyadi, et al.	2020	ELK + Wazuh efficacy demonstrated. It shows that despite using the docked ELK, the performance shows the good condition and the latency is considered small. Moreover, the CPU usage and the used memory usage remain stable for the host servers	Y	N	Security Information and Event Management (SIEM) , ELK + wazuh and docked ELK
[20]	F. Balseca-Chávez, et al	2021	The designed technology architecture was based on the integration of Elastic Stack and its key components (Elasticsearch, Logstash and Kibana), technologies such as Filebeat and Wazuh Security Detection (NIPS/HIDS).	Y	N	Elasticsearch, Logstash and Kibana), Filebeat and Wazuh
proposed system		2022	Technology architecture was based on the integration of Elastic Stack and its key components (Elasticsearch, Logstash and Kibana), technologies such as Filebeat and Wazuh Security Detection.	Y	Y	SIEM, wazuh and Azure cloud

Table 2. Big data features

Feature	Description
Volume	Data is produced in larger quantities than traditional data.
Variety	Different typologies and structures of the data coming from very different sources. Classified in: structured, unstructured and semi-structured.
Velocity	The processing of the data must be done in the shortest possible time and even in real time to access the data, to facilitate analysis and processing.
Value	The value of any data varies according to its content, it is necessary to identify the valuable information, transform it and extract the data for later analysis.
Veracity	The information collected must enjoy a high level of reliability, eliminate possible events of inaccuracy or uncertainty.
Viability	It relates to the ability of organizations to effectively use the large volume of data they handle.

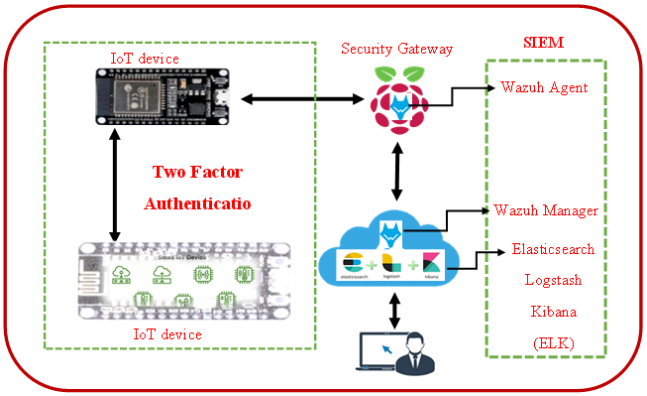

Figure. 2 SIEM reference architecture

The tendency to manipulate such amounts of data is due to the need to include such information for the creation of statistical reports and predictive models used by the managers of organizations for the analysis of business performance, marketing and customer loyalty, data on infectious-contagious diseases, industrial espionage and monitoring of citizens, or in the fight against organized crime, among other activities [23].

The concept of Big Data applies to all information that cannot be processed or analyzed, using traditional processes or tools; six key characteristics are considered, which are described in Table 2, which sets out the information related to big data in reference to information security [24].

2.1 Information security

The term information security has been coined by different authors, converging on the protection of confidentiality, integrity and access to information; encompasses technology, processes and people with the purpose of mitigating threats to information; using different technical measures, among which specialized antivirus and antispyware software stand out, biometric devices until reaching firewalls [25].

The modern security landscape is influencing sites on the network, and the cloud is constantly evolving; bringing with them threats are visualized from a variety of avenues [26]. It is vital to develop operational safety capabilities in all global contexts [27]. Information security includes those processes, good practices and methodologies that aspire to the protection of information and access to information systems, their use, disclosure, interruption, modification or unauthorized destruction. It also relates to the processes of registration and cancellation of users and personal aspects of users. All these aspects must be taken into account when deploying information security[28].

2.2 Referential architecture of the SIEM

The proposed referential architecture of a SIEM is made up of a complex technological structure, described in Fig. 2; whose components are physically and logically interconnected from their key functions.

2.3 Elastic stack components

Elastic stack es una Plataforma de código abierto para la ingesta confinable de dates de different fonts, in a variety of different formats; that allows to search, analyse and visualize data in real time [10]. It is used to create big data solutions, it is composed of: Elasticsearch (ES), Logstash and Kibana [29]. It is known as an open source search and analytics engine, based on Apache Lucene, considered as a full-text, distributed and multi-tenancy search engine with a RESTful web interface (Based on REST architecture, which is an interface to connect several HTTP-based systems, and serves to obtain and generate data and operations) and with JSON documents, designed to enable scale-out, reliability and easy management [30]. Elastic stack as a technology solution can be used for visualization and analysis not only of system failure logs, but also of common task logs in commercial systems [29].

Figure 3. Garner magic quadrant for insight engines

Table 3. Elastic stack solution specifications

Solution Technological	Description
Elastic search (ES)	Allows full-text search in unstructured data does not depend on the type of data source: structured, semi-structured or structured or unstructured.
Logstash	It is used to collect and analyze the data in a central service and send the result to ES for indexing. It can be customized and adapted to process any data source such as records, netflow, databases, databases, files, etc.
Kibana	It is the graphical interface of the solution that displays and searches es data. It contains a search syntax for querying the ES. It facilitates the creation of interactive dashboards, as it provides various types of predefined graphs such as cakes, histograms or trends.

The magic quadrant of Garner includes in the month of February 2021 Elastic, described in Fig. 3, among the top fifteen providers of information engine that combines search capabilities with artificial intelligence, to offer actionable information derived from the full spectrum of content and data, obtained inside and outside a company, thus helping application leaders make the best decision.

The three solutions (Elastic search, Logstash, Kibana) are combined to build a Big Data solution [29]. According to the specifications in Table 3.

At present, cyber threats are becoming increasingly sophisticated, security analysis and real-time monitoring are needed, for rapid detection and repair of threats. The Wazuh intrusion detection system is part of the SIEM, therefore it integrates with Elastic Stack. It is known as an open source platform that is used to collect, aggregate, index, and analyze security data, helping organizations detect intrusions, threats, and behavioral anomalies [31].

Meanwhile, the Filebeat component becomes a lightweight forwarder that is used to transmit logs over a network, usually to ES. It is used on the Wazuh server to send events and alerts to ES. It reads the output of the Wazuh analysis engine and sends events in real time through an encrypted channel [32].

3. Methodology

For the development of this work, an exploratory, descriptive research method was followed, generating value together with the experience and knowledge of experts specialized in information security and data analysis, residing its applicability in the environment of development of technological architectures, using tools from the big data ecosystem to improve the identification of computer threats.

Within the processes carried out in the research described in Fig. 4, a controlled scenario was proposed that simulates an information security operations and monitoring center of a company that provides technological services in Ecuador, with a practical-investigative character and. in order to analyze and evaluate possible information security breaches, based on the design of a technological, process, software, business architecture, thus supporting the phases of the big data process proposed by [33] for the identification of computer threats.

3.1 Acquisition and registration

This stage comprises the first step of processing, including the provisioning of various data sources: data servers, applications, user terminals, security equipment, sensor networks, among others which can produce staggering amounts of raw data.

The data was collected with potentially useful information for the analysis of information security events, from information assets such as data servers or applications, user terminals, security equipment, communications equipment, among others, in the form of events (logs).

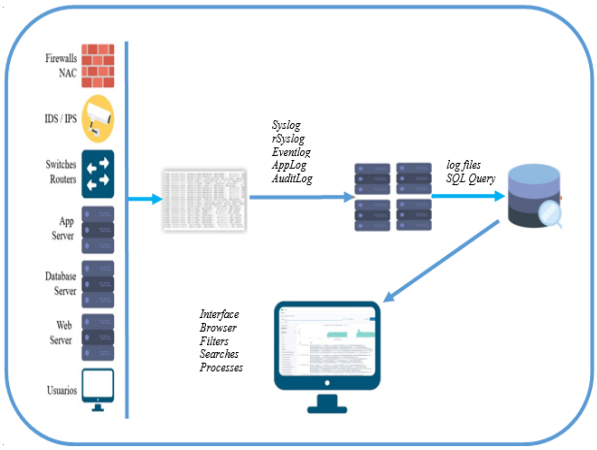

Figure. 4 Technological architecture design process

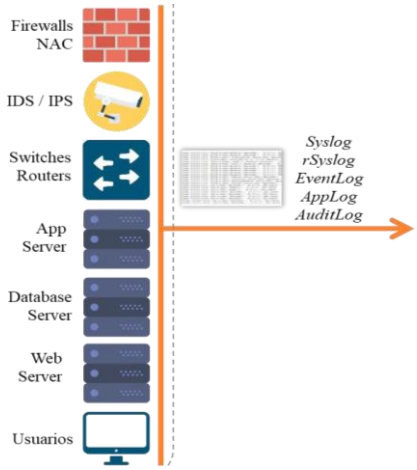

Figure. 5 Acquisition and registration phase

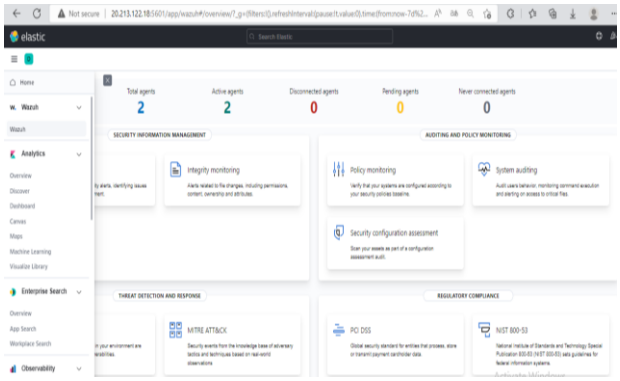

Figure. 6 Graphical interface, management of deployed agents

In this phase, Filebeat agents are used, as a platform for data managers and subsequent sending from the information assets to subsequent processes (Logstash and ES), see Fig. 5. This lightweight application allows you to forward and centralize logs; it is installed as an agent on your servers. Filebeat monitors log files or required locations, collects log events, and forwards them to ES or Logstash for indexing [34].

Figure. 7 Graphical interface, threat identification overview

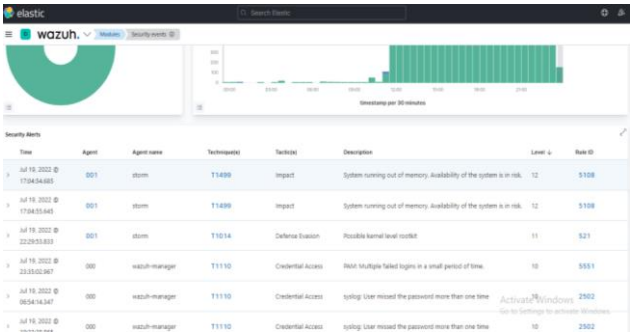

Figure. 8 Elastic stack graphical interface

3.2 Extraction, cleaning, annotation

In this phase necessary information is extracted from the underlying sources and expressed in a structured form, suitable for analysis. Wazuh performs intrusion detection based on an open source host (HIDS), providing log analysis, monitoring the integrity of operating system files, detection of rootkits and vulnerabilities, evaluation of configurations and responsiveness to information security incidents [35].

This end-to-end solution integrates with additional tools such as OpenSCAP and ES, as described in Figs. 6 and 7. Logstash acts as an open-source data processing channel on the server side of big data, which ingests data from a multitude of sources simultaneously, transforms it and sends it for storage.

3.3 Integration, aggregation and representation

This phase (Fig. 8) has as its main function the processing and transformation of data, which can be stored and displayed in a more understandable way. The aggregation of data was carried out through a specialized software, by the ES, as a search engine and analysis, capable of adapting to a growing number of use cases.

The software in question is responsible for storing the data in a centralized and distributed way, allowing advanced searches to be executed. Using the Elastic stack graphical interface, described in Fig. 8, logs are

Figure. 9 Graphical interface, security dashboard

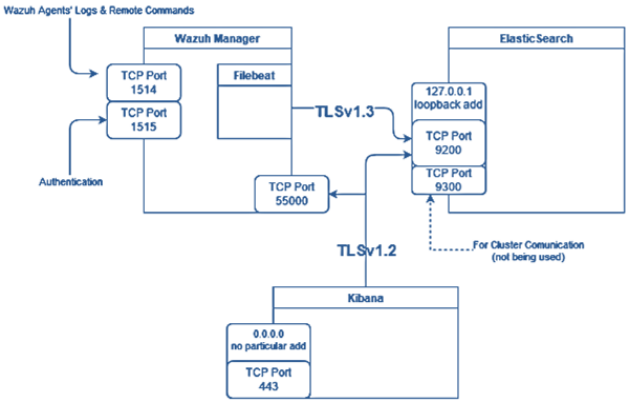

Figure. 10 Diagram of installation wazuh manager and ELK

tracked in real time, through a unified, customizable display. Log data correlates with infrastructure UI metrics, making it easier to diagnose potential incidents.

4. Analysis and modelling

Within the analysis and modelling of big data the set of applications, components and standards that process large amounts of data, being their results, key elements for decision making are considered as tools.

The technology used in the analysis was Kibana, user interface in the Elastic stack. It facilitated the analysis of time-based events through visualization. With the creation of interactive dashboards, it allowed for greater understanding and visibility.

The index created in Elastic search is used in Kibana to analyse and create visualizations. It also contemplates anomaly detection, alerts and monitoring based on machine learning [36]. Fig. 9 describes the use of Kibana with the visualization of the data of the searches executed in ES, and navigate through the Elastic stack at the level of generated Dashboards, showing all the information that is stored and indexed as unstructured data [37].

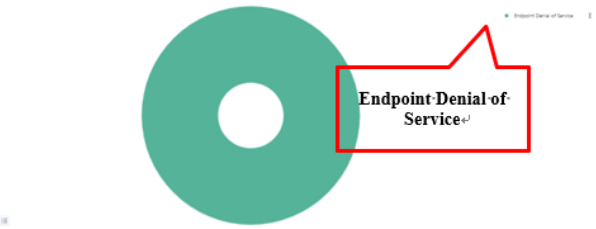

Figure. 11 Represented denial of service (DoS)

5. Results

Once the data inventory was carried out and in accordance with the purpose of the research, we proceeded to the evaluation of information security gaps, applying a technological-business architecture that allowed the implementation of tools of the Big Data ecosystem. For the purpose of executing the validation tests of the proposed infrastructure, five use cases of simulated computer attacks, under a controlled environment, were applied to the technological infrastructure deployed as part of this research, which have been developed for the exploitation of the most common computer vulnerabilities.

This agent monitors the security gate while collecting logs and sending them to the Wazuh manager. Wazuh manager is installed on the Azure cloud. Elasticsearch and Kibana are also installed. Elasticsearch is a free and open-source search and analytics engine for all the types of data that Wazuh uses to process its logs. And Kibana is a data visualization and management tool from Elasticsearch, which also has a Wazuh plugin that allows it to visualize Wazuh logs and alerts. In addition, the open source Filebeat software sends Wazuh logs to Elasticsearch. Additional details of this installation can be shown in Fig. 10.

5.1 Detecting a denial of service (DoS) attack

This denial-of-service test is carried out on the wazuh agent device or on the wazuh manager, as this attack is considered one of the most dangerous attacks as it floods the victim with a flood of requests after which the victim falls and cannot continue and provide his services. It is expected that Wazuh will be able to correctly detect and identify the patterns of this attack. To highlight the capabilities of the monitoring and recording tool to detect and correctly identify an electronic attack, when wazuh alerts are observed, it identifies this attack with its criticality and the Fig. 11. Shows how to fill the memory of the victim.

Fig. 12 shows some of the details that appear in the wazuh dashboard alerts, it also shows the type and

date of the attack, and the Table 4 describes the rest of the important details.

Figure. 12 Details of wazuh alerts the rule 5108

Table 4. The attack information

AField Name	Alert Details
rule.id	5108
Rule.group	syslog, linuxkernel, service_availability
rule. description	System running out of memory. Availability of the system is in risk.
rule.mitre.id	5108
rule. level	12
rule. mitre.tactic	Impact
rule. mitre.technique	Endpoint Denial of Service

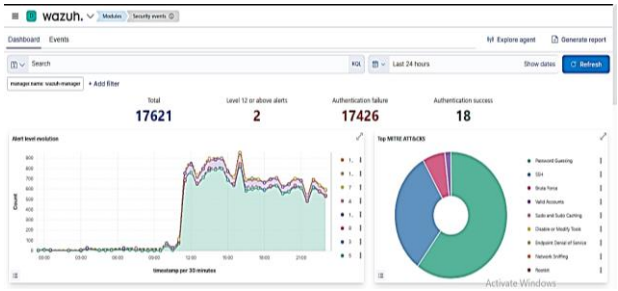

Figure. 13 Graphic of the alerts generated by Wazuh

5.2 Detecting a brute-force attack

This test performs a brute force attack on a device with the Wazuh agent installed. To identify and demonstrate the capabilities of Wazuh to detect attacks. After executing the previous command, it goes to Wazuh Alerts in the Security Events module of the Wazuh Kibana plug-in. There it was discovered that it was a response and as shown in Fig. 13 where the failed authentication is displayed, which represents the number of failed attempts. Whereas, Wazuh created multiple versions of two alerts with base ID numbers 5710, 5712 and 5551. They can be analysed in Fig. 14 and brief details of each can be found in Table 5.

Table 5. The truncated details of both Wazuh alerts generated in response to the brute-force attack

Field Name	Alert Id:5710	Alert Id:5712	Alert Id: 5551
rule.description	sshd: Attempt to login using a non-existent user	sshd: brute force trying to get access to the system.	PAM: Multiple failed logins in a small period of time.
rule.firedtime	20	21	24
rule.id	5710	5712	5551
rule.level	5	10	10
rule.mitre.tactic	Credential Access	Credential Access	Credential Access
rule.mitre.technique	Brute Force	Brute Force	Brute Force

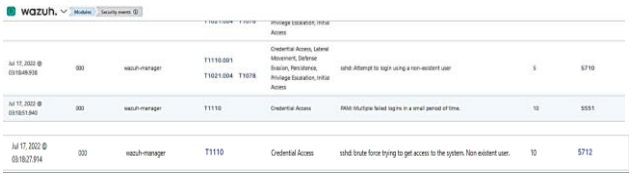

Figure. 14 Details of wazuh alerts the rule 5710-5712-5551

Additionally, the fields above, alerts with rule ID 5712 and 5551 have an additional entry called previous_output. This field contains the multiple unsuccessful attempts that have been aggregated to trigger a warning with rule ID 5710.

5.3 Detecting file changes

This test aims to check if Wazuh can detect changes to files within the system. Additionally, it seeks to see what information is logged when these changes are detected and if it is possible to track the origin or author of the change.

An integral part of recognizing and tracing the route of cyber-attack is the logging of system file changes. These logs should include information on when these changes happen and who performed these

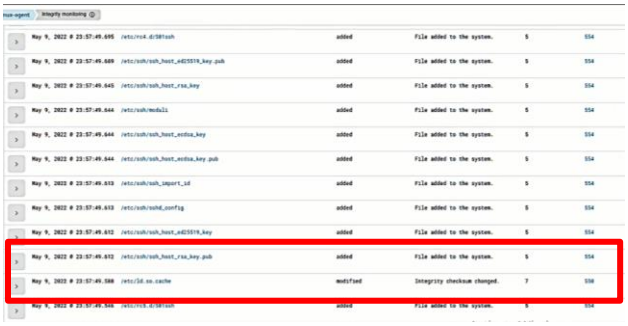

Figure. 15 Details of wazuh alerts the rule 554 - 550

Table 6. Alert generated by Wazuh in response to the creation of the file

Field Name	1st Alert Details
rule.id	554
rule.description	File added to the system
rule.level	5
syscheck.audit.login_user.name	ptin_admin
syscheck.audit.process.name	/usr/bin/touch
syscheck.event	added
syscheck.path	/etc/systemd/system/multi-user.target.wants/snap-core20-1581.mount

Table 7. Alert generated by Wazuh in response to the second modification of the file

Field Name	2nd Alert Details
rule.id	550
rule.description	Integrity checksum changed.
rule.level	7
syscheck.audit.login_user.name	admin
syscheck.uname_after	root
syscheck.audit.process.name	/usr/bin/nano
syscheck.event	modified
syscheck.path	/etc/cups/subscriptions.conf

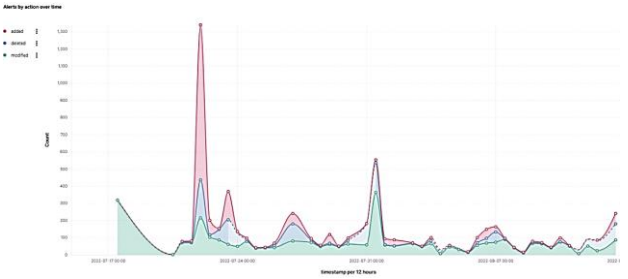

Figure. 16 Alerts the files change

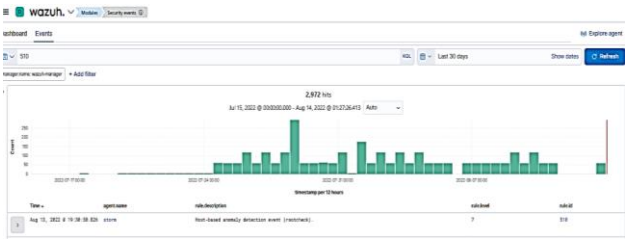

Figure. 17 The suspicious binaries

changes. This capability of these tools is crucial for performing damage recovery and preventing repeated attacks from happening. As such, this test aims to implementing a monitoring and logging system in the Azure cloud. Demonstrate this ability and observe if Wazuh provides satisfying amounts of details when file systems are changed. As shows in Fig. 15.

Table 6 describes creating a file and shows the rest of the important details such as the rule ID number and the level of severity.

Takes a closer look at the alerts generated, it is possible to see that they contain fields that allow the identification of the author of the changes and what process was used to make the changes. In the case of this test, it's possible to see that the file was created by the user "admin" with the command "touch" and then altered with the process "nano". After that, the user "admin", as effective user "root", changed the file again with "nano" and then deleted the file with the command "rm". Table 7. A comparative look at the alerts generated in the detecting file changes test.

Additionally, both modification-related alerts have fields containing the details of the change made to the file. Wazuh can catalogue and show both the author of the change and the changes made to the file. Therefore, Wazuh can provide crucial information on the aftermath of an attack for its identification and tracking. As shown in Fig. 16.

5.4 Detecting suspicious binaries

This test aims to determine if a Wazuh installation can detect malware attacks by identifying suspicious binaries such as trojans or viruses. An original binary system is replaced by a "harmless" Trojan version of this test. Malware-related attacks are one of the most common forms of cyber-attack. By using a combination of file monitoring and pattern recognition, monitoring and logging tools can detect such malicious files within the system before they can cause damage to the infected system. As shown in Fig. 17. This test aims to demonstrate this ability to monitoring and recording tools and to see if Wazuh has such capabilities. Wazuh already comes with a list of common Trojan locations and signatures in the /var/ossec/etc/shared/rootkit_trojans.txt file.

By combining the contents of this file with the capabilities described in the file change detection test, Wazuh should be able to detect system changes that are caused by malware and Additionally, Wazuh must be configured to check for malware and Trojans every 12 hours by default upon installation, this can be checked in the monitored device configuration file /var/ossec/etc/ossec.conf. The Table 8 shows alert details.

Table 8. Detecting suspicious binaries

Field Name	Alert Details
data.file	/usr/bin/w
data.title	Trojaned version of file detected.
full_log	Trojaned version of file '/usr/bin/w' detected. Signature used: 'uname -a|proc\.h|bash' (Generic).
rule.description	Host-based anomaly detection event (rootcheck).
rule.groups	ossec, rootcheck
rule.id	510
rule.level	7

Table 9. The truncated details of the Wazuh alert generated in response to the black-listed

Field Name	Alert Details
location	process list
rule.description	Netcat listening for incoming connections.
rule.groups	ossec, process_monitor
rule.id	100051
rule.level	7

Table 10. The truncated details of the Wazuh alert generated in response to the usage of the

Field Name	Alert Details
data.title	Process '1045' hidden from /proc.
full_log	Process '1045' hidden from /proc. Possible kernel level rootkit.
rule.description	Possible kernel level rootkit
rule.id	521
rule.level	11
rule.mitre.tactic	Defense Evasion
rule.mitre.technique	Rootkit

Figure. 18 details of wazuh alerts the rule 521

5.5 Detecting unauthorised processes

This test aims to see if Wazuh can detect the execution of black-listed processes. Monitoring running processes on a system is a valuable capability of monitoring and logging tools. This capability alerts when crucial processes or functions stop unexpectedly and identifies unknown or unwanted ones suddenly starting within the system. As shown in Table 9. This test aims to demonstrate this ability of monitoring and logging tools and see if Wazuh has such capabilities.

After executing the previous command, access the Wazuh Alerts in the security events module of the Wazuh Kibana plugin. can see the alert created in response to the execution of the black-listed Netcat command.

5.6 Exposing rootkits and hidden processes

This test aims to see if Wazuh can detect rootkits that can hide from the kernel module list and hide their processes from the "ps" command. This test can be seen as a continuation of the previous one, as it demonstrated that Wazuh was capable of monitoring running processes. Still, it did so by analyzing the list provided by the system. This test aims to see if Wazuh can detect processes if those have been hidden by a rootkit, a standard cover tactic used by malicious attackers. The rootkit Diamorphine will be installed on the machine to perform this test. As shown that in Fig. 18.

The Wazuh agent is configured by default to check for rootkits and hidden processes every 12 hours. It does this by utilising system calls such as setsid(), getpid() and kill() that indirectly expose any processes running in the system. This function can be sped up for the sake of this test in the "rootcheck" and "syscheck" sections of the var/ossec/etc/ossec.conf file.

As the Wazuh is already configured, it is just a matter of downloading and installing the Diamorphine rootkit. After executing the previous steps, access the Wazuh Alerts in the security events module of the Wazuh Kibana plugin. can see the alert created in response to the Diamorphine rootkit hiding

Table 11. Comparison between the performance of the proposed system and the performance of the in previous works

Author (year)	Algorithm and protocol	IoT authentication	Experiment simulation	SIEM type	Cloud	Protection type
D. D. Khudhur and M. S. Croock, (2017) [10]	MQTT	No	(NodeMCU 12), Raspberry Pi3 as the sensors (DS18B20, actuators.	No	No	secure sockets layer (SSL)/transport layer security
N. A-hussein and A. D. Salman (2018) [13]	MQTT	No	Uses the Esp8266, Raspberry pi 3 and Node-Red and ThingSpeak were designed as a website to share the data	No	IoT platform	No
H. Sun, H. Yu, et al. (2020) [15]	ETCORA algorithm	No	fog-cloud and ETCORA algorithm		Yes	No
V. Teeraratchakarn and Y. Limpiyakorn (2020) [17]	No	No	Elastic Stack	ELK	No	IDS
F. Mulyadi, et al. (2020) [18]	No	No	Y	ELK + wazuh and docked	N	IDS
F. Balseca-Chávez, et al. (2021) [20]	JSON	No	(NIPS/HIDS)	Elasticsearch, Logstash and Kibana), Filebeat and Wazuh	N	IDS
Proposed system (2022)	HTTP and JSON.	TFA by Token	Using ESP8266, (Raspberry pi),	Elasticsearch, Logstash and Kibana), Filebeat and Wazuh	Azure	Authentication, IPS and IDS

a process from the "ps" command, meaning that Wazuh has successfully detected the hidden processes running in the machine. As shown in Table 10.

6. Comparison and discussion

From the results of the previous work, it was found that the proposed system achieved a desirable

performance in preventing and detecting various attacks such as (DOS, DDOS, MITM), brute force and dictionary attacks. The main contribution of this work is the implementation of the authentication model between IoT devices. And the application of the security portal compared to previous research. Table 11 shows a comparison between the proposed method and the latest relevant research.

7. Conclusion

Conclusions In this work, a theoretical basis is laid that lays out a proposal for a security solution based on anomaly detection for network hosts such as servers. The benefits that Wazuh provides as an IDS when detecting large-scale anomalies is not only to deal with a set of rules that detect and identify attacks, but also the way it can respond to security incidents that may occur. For this reason, active responses can be seen as an essential tool when proposing a security solution. The combination of Elastic stack and Wazuh with the goal of providing security for a business environment has been successfully implemented due to the multiple functions and benefits of these tools, thanks to the implementation of a lab environment whose operation has been validated, implementation of Wazuh as a plug-in is a successful way to exploit all the functions it provides as IDS / HIDS. The installation that is implemented in this solution, that is why security solutions are increasing every day and in the same way critical systems breaching and attacking techniques like as BDDs, it is important to keep Wazuh up to date without first thinking that this may affect how it works in tandem with Elastic stack. The DoS attack when making partial HTTP requests generates multiple errors that are detected by Wazuh because they come from the same source, so by detecting the active response configured, it implements and adds a web server firewall rule to deny the source IP for the detected attack. The SQL injection attacks simulated in this work are based on the most commonly used and simplest in practice, but it is their simplicity that makes them the most dangerous. This is why Wazuh solutions like HIDS and Elastic Stack appear to be suitable for any type of malicious attack. The active responses that the Wazuh plus Elastic stack security solution has are very efficient in terms of the response time at which an attack is detected, i.e. the network administrator performs traditional monitoring to identify the attacks and block the IP addresses of the attack origin. It takes several minutes, considering that among the hundreds of possible alerts that can be generated, the responsible person detects them and implements them manually. So with this solution it is instant.

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

The paper conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review, and editing, visualization, have been done by 1st author. The supervision and project administration has been done by 2nd author.

References

[1] M. E. Arass and N. Souissi, “Smart SIEM: From big data logs and events to smart data alerts”, *Int. J. Innov. Technol. Explor. Eng*, Vol. 8, No. 8, pp. 3186–3191, 2019.

[2] A. K. Obaid, “an Improved Data Confidentiality Protocol Based on Timestamp”, *Iraqi J. Comput. Commun. Control Syst. Eng.*, Vol. 12, No. 1, 2012.

[3] G. Kaiafas et al., “Detecting malicious authentication events trustfully”, In: *Proc. of NOMS 2018-2018 IEEE/IFIP Network Operations and Management Symposium*, pp. 1–6, 2018.

[4] M. Zolanvari, M. A. Teixeira, L. Gupta, K. M. Khan, and R. Jain, “Machine learning-based network vulnerability analysis of industrial Internet of Things”, *IEEE Internet Things J.*, Vol. 6, No. 4, pp. 6822–6834, 2019.

[5] N. Moustafa, M. Keshky, E. Debiez, and H. Janicke, “Federated TON_IoT Windows datasets for evaluating AI-based security applications”, In: *Proc. of 2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, pp. 848–855, 2020.

[6] J. Sengupta, S. Ruj, and S. D. Bit, “A comprehensive survey on attacks, security issues and blockchain solutions for IoT and IIoT”, *J. Netw. Comput. Appl.*, Vol. 149, p. 102481, 2020.

[7] N. Koroniotis, N. Moustafa, E. Sitnikova, and B. Turnbull, “Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-iot dataset”, *Futur. Gener. Comput. Syst.*, Vol. 100, pp. 779–796, 2019.

[8] B. K. Mohanta, D. Jena, U. Satapathy, and S. Patnaik, “Survey on IoT security: Challenges

and solution using machine learning, artificial intelligence and blockchain technology", *Internet of Things*, Vol. 11, p. 100227, 2020.

[9] H. Wu, H. Han, X. Wang, and S. Sun, "Research on artificial intelligence enhancing internet of things security: A survey", *IEEE Access*, Vol. 8, pp. 153826–153848, 2020.

[10] W. Sholihah, S. Pripambudi, and A. Mardiyono, "Log event management server menggunakan elastic search logstash kibana (elk stack)", *JTIM J. Teknol. Inf. dan Multimed.*, Vol. 2, No. 1, pp. 12–20, 2020.

[11] A. Ključnikov, L. Mura, and D. Sklenár, "Information security management in SMEs: factors of success", *Entrep. Sustain. Issues*, Vol. 6, No. 4, p. 2081, 2019.

[12] G. G. Granadillo, S. G. Zarzosa, and R. Diaz, "Security information and event management (SIEM): analysis, trends, and usage in critical infrastructures", *Sensors*, Vol. 21, No. 14, p. 4759, 2021.

[13] A. F. Y. Althabhawee and B. K. O. C. Alwawi, "Fingerprint recognition based on collected images using deep learning technology", *IAES Int. J. Artif. Intell*., Vol. 11, No. 1, pp. 81–88, 2022, doi: 10.11591/ijai.v11.i1.pp81-88.

[14] D. D. Khudhur and M. S. Croock, "Developed security and privacy algorithms for cyber physical system.", *Int. J. Electr. Comput. Eng.*, Vol. 11, No. 6, 2021.

[15] N. A. hussein and A. D. Salman, "IoT Monitoring System Based on MQTT Publisher/Subscriber Protocol", *Iraqi J. Comput. Commun. Control Syst. Eng*., Vol. 20, No. 3, pp. 75–83, 2020, doi: 10.33103/uot.ijccce.20.3.7.

[16] H. Sun, H. Yu, G. Fan, and L. Chen, "Energy and time efficient task offloading and resource allocation on the generic IoT-fog-cloud architec?ture", *Peer Peer Netw. Appl.*, Vol. 13, No. 2, pp. 548–563, Mar. 2020.

[17] J. A. Alzubi, R. Manikandan, O. A. Alzubi, I. Qiqieh, R. Rahim, D. Gupta, and A. Khanna, "Hashed needham schroeder industrial IoT based cost optimized deep secured data transmission in cloud", *Measurement*, Vol. 150, Art. No. 107077, Jan. 2020.

[18] S. A. Rubaye, E. Kadhum, Q. Ni, and A. Anpalagan, "Industrial Internet of Things Driven by SDN Platform for Smart Grid Resiliency", *IEEE Internet Things J.*, Vol. 6, No. 1, pp. 267–277, 2019, doi: 10.1109/JIOT.2017.2734903.

[19] F. Mulyadi, L. A. Annam, R. Promya, and C. Charnsripinyo, "Implementing Dockerized Elastic Stack for Security Information and Event Management", in *2020-5th International Conference on Information Technology (InCIT)*, pp. 243–248, 2020.

[20] F. B. Chávez, A. M. C. Vargas, and M. A. E. Mina, "Identificación de amenazas informáticas aplicando arquitecturas de Big Data", *INNOVA Res. J.*, Vol. 6, No. 3.2, pp. 141–167, 2021.

[21] B. Gupta, R. Kumar, and A. Kumar, "Towards Information Discovery On Large Scale Data: state-of-the-art", In: *Proc. of 2018 International Conference on Soft-Computing and Network Security (ICSNS)*, pp. 1–9, 2018.

[22] S. A. Rubaye, E. Kadhum, Q. Ni, and A. Anpalagan, "Industrial internet of things driven by SDN platform for smart grid resiliency", *IEEE Internet Things J.*, Vol. 6, No. 1, pp. 267–277, 2017.

[23] H. M. Hasan and S. A. Jawad, "IoT protocols for health care systems: A comparative study", *Int. J. Comput. Sci. Mob. Comput.*, Vol. 7, No. 11, pp. 38–45, 2018.

[24] A. I. Alaoui, K. Baïna, K. Benali, and J. Baïna, "Towards smart incident management under human resource constraints for an iot-bpm hybrid architecture", In: *Proc. of International Conference on Web Services*, pp. 457–471, 2018.

[25] M. Lněnička, R. Máchová, J. Komárková, and I. Čermáková, "Components of big data analytics for strategic management of enterprise architecture", In: *Proc. of 12th International Conference on Strategic Management and its Support by Information Systems*, 2017.

[26] H. H. Hassan and M. A. A. Khodher, "Data Hiding by Unsupervised Machine Learning Using Clustering K-mean Technique", *Iraqi J. Comput. Commun. Control Syst. Eng.*, Vol. 21, No. 4, pp. 37–49, 2021.

[27] Q. F. A. A. Doori, "Design of a smart power manager for digital communication systems", *University of Salford (United Kingdom)*, 2017.

[28] A. R. Nasser et al., "Iot and cloud computing in health-care: A new wearable device and cloud-based deep learning algorithm for monitoring of diabetes", *Electron.*, Vol. 10, No. 21, p. 2719, 2021, doi: 10.3390/electronics10212719.

[29] S. S. Sekharan and K. Kandasamy, "Profiling SIEM tools and correlation engines for security analytics", In: *Proc. of 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, pp. 717–721, 2017.

[30] A. Talaş, F. Pop, and G. Neagu, "Elastic stack in action for smart cities: Making sense of big data", In: *Proc. of 2017 13th IEEE International*

Conference on Intelligent Computer Communication and Processing (ICCP), pp. 469–476, 2017.

[31] K. D. Salman and E. K. Hamza, "Visible Light Fidelity Technology: Survey", *Iraqi J. Comput. Commun. Control Syst. Eng.*, Vol. 21, No. 2, pp. 1–15, 2021, doi: 10.33103/uot.ijccce.21.2.1.

[32] F. Mulyadi, L. A. Annam, R. Promya, and C. Charnsripinyo, "Implementing Dockerized Elastic Stack for Security Information and Event Management", In: *Proc. of 2020 5th International Conference on Information Technology (InCIT)*, pp. 243–248, 2020.

[33] M. D. Pratama, F. Nova, and D. Prayama, "Wazuh sebagai Log Event Management dan Deteksi Celah Keamanan pada Server dari Serangan Dos", Vol. 3, No. 1. pp. 1–7, 2022.

[34] J. Chen et al., "Big data challenge: a data management perspective", *Front. Comput. Sci*., Vol. 7, No. 2, pp. 157–164, 2013.

[35] S. H. Hashem1, "Proposed Integrated Wire/Wireless Network Intrusion Detection System", *Iraqi J. Comput. Commun. Control Syst. Eng.*, Vol. 14, No. 2, 2014.

[36] A. M. Kadhum and E. K. Hamza, "Implementation of Spectrum Sensing based OFDM Transceiver using Xilinx System Generator", *Iraqi J. Comput. Commun. Control Syst. Eng.*, Vol. 21, No. 2, pp. 59–69, 2021, doi: 10.33103/uot.ijccce.21.2.5..

[37] R. Máchová and M. Lněnička, "Evaluating the quality of open data portals on the national level", *J. Theor. Appl. Electron. Commer. Res.*, Vol. 12, No. 1, pp. 21–41, 2017.