

DEMOCRATIC AND POPULAR REPUBLIC OF ALGERIA
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

UNIVERSITY SAAD Dahleb- Blida 1



Faculty of Sciences
Computer Science Department
MASTER THESIS
Specialty « Information Systems Security »

Design and Implementation of a SIEM
(Security Information and Event Management) System in
BADR Bank

Author: BOUATTOU Wissam

Advisor: Dr. SAHNOUNE Zakaria

President: Prof. BOUSTIA Narhimene

Supervisor: Mr. LADJOUZI Mahfoud

Examiner: Dr. NASRI Ahlem

Date: 06th July 2022

Academic year:

2021/2022

Acknowledgment

Above all, I thank the merciful A.L.L.A.H. the Almighty for giving me the knowledge, strength, and courage to carry out this work.

My deep gratitude goes to my promoter, Doctor SAHNOUNE Zakaria, for his availability and especially for his precious advice.

My thanks are also addressed to the jury members.

My sincere gratitude to the director of D.P.C.A.S.S.I., Madam OUSSADIT Akila, for welcoming me in her direction and to my supervisor Mr. LADJOUZI Mahfoud for giving me the great honor of leading my project under the best possible conditions, as well as for his mentoring, advice and availability.

I express my acknowledgment to all the staff of B.A.D.R. Bank.

I express my gratitude and warm thanks to my source of life and hope and my symbol of sacrifice to my parents, who supported, encouraged, and motivated me.

I would also like to thank my sisters Asma, Selma, and Sabrina for their understanding and their warm encouragement, without forgetting my little nephews Mohamed Amine and Aya Sidra.

I would like to express my gratitude and esteem to my teachers who have contributed to my training throughout my studies.

To my SSI classmates who were by my side, with them, I trained how to work as a team Oussama MOUSSIOU & Aghiles ZIBANI.

Immense thanks go to my dearest friends OUIR. Fatima, AMROUCHE Wafa, BOUSSAHOUA Mouna & BENALLAL Fella for their encouragement and their unfailing support.

Abstract

With the development and steady growth of several technologies, I.T. security has become a crucial issue for any business. As the information system is an essential asset of the company, the latter's security is paramount. This project consists of designing and implementing S.I.E.M. security information and event management for B.A.D.R. Bank, a tool that provides a dashboard to monitor the collected events and real-time analysis and alerts from the organization logs network equipment in case of anomalies. Open source tools E.L.K. and Wazuh are used for this solution.

The system includes agents installed in multiple endpoints to collect logs from the different sources and send these data in a raw format to the processing server for standardization and parsing.

The logs are indexed and stored and used by visualization tools as graphs and tables gathered in dashboards; alerts are created when identifying an attempted attack or anomaly in the system.

Keywords: Alerts, attack, detection, E.L.K., logs, log management, monitoring, security, S.I.E.M., supervision, visualization, Wazuh.

Résumé

Avec le développement et la croissance constante de plusieurs technologies, la sécurité informatique est devenue un enjeu crucial pour tout type d'entreprise. Le système d'information étant un atout essentiel de l'entreprise, la sécurité de ce dernier est primordiale. Ce projet consiste en la conception et la mise en œuvre d'un SIEM d'information de sécurité et de gestion des événements pour BADR Banque, qui est un outil qui fournit un tableau de bord pour surveiller les événements recueillis et l'analyse en temps réel et les alertes de l'organisation enregistrer dans l'équipement du réseau en cas d'anomalies. Les outils open source ELK et Wazuh sont utilisés pour cette solution.

Le système comprend des agents qui sont installés dans plusieurs endpoints pour collecter les fichiers journaux des différentes sources et envoyer ces données dans un format brut au serveur de traitement pour la normalisation et l'analyse.

Les fichiers journaux sont indexés et stockés et utilisés par des outils de visualisation sous forme de graphiques et de tableaux rassemblés dans des tableaux de bord ; des alertes sont créées lors de l'identification d'une tentative d'attaque ou en cas d'anomalie dans le système.

Mots clés : Alertes, attaque, détection, ELK, fichiers journaux, sécurité, SIEM, supervision, surveillance, visualisation, Wazuh.

المُلخَص

مع التطور والنمو المطرد للعديد من التقنيات، و المشاكل الأمنية و الاختراقات التي تمس خصوصية المعلومات، أصبح أمن تكنولوجيا المعلومات قضية حاسمة لأي شركة. نظرًا لأن نظام المعلومات هو أحد الأصول الأساسية لها، فإن ضمان هذه الانظمة له أهمية قصوى.

من خلال هذا المشروع، نتطرق نحو تصميم وتنفيذ أداة معلومات أمنية وإدارة أحداث SIEM لبنك الفلاحة و التنمية الريفية، حيث توفر هذه الاداة لوحة متابعة لرصد الأحداث التي تم جمعها، والتحليلات والتنبيهات في الوقت الفعلي، لتسجيلات معدات شبكة المعلومات الخاصة بالبنك في حال حدوث حالات شاذة. لتنفيذ هذا الحل، تم إستعمال أدوات من المصادر المفتوحة ELK و Wazuh.

يتضمن النظام وكلاء مثبتين في نقاط نهاية متعددة لجمع السجلات من المصادر المختلفة وإرسال هذه البيانات في شكل خام إلى خادم المعالجة للتوحيد القياسي والتحليل. يتم فهرسة السجلات وتخزينها واستخدامها بواسطة أدوات التصور كرسوم بيانية وجداول مجمعة في لوحات القيادة؛ يتم إنشاء تنبيهات عند تحديد محاولة هجوم أو شنوذ في النظام.

الكلمات المفتاحية: تنبيهات، هجوم، كشف، ELK، سجلات، إدارة سجل، مراقبة، أمن، SIEM، إشراف، تصور، Wazuh.

Contents

Introduction.....	1
Chapter I: Host organization and study the existing.....	5
Introduction.....	6
I.1 Presentation of BADR Bank.....	6
I.1.1 History and advancement of BADR Bank	6
I.1.2 Missions of BADR Bank.....	6
I.1.3 Fundamental activities of BADR Bank.....	7
I.1.4 Strategic objectives of BADR Bank.....	7
I.2 Badr bank organization	8
I.2.1 Direction of the continuity plan of the information system	8
I.2.2 Audit of SSI Compartment general missions	11
I.3 Criticism of the existing.....	12
I.4 Problematic	13
Conclusion.....	14
Chapter II: Theoretical background.....	15
Introduction.....	16
II.1 IT supervision and security monitoring	16
II.1.1 IT supervision definition	16
II.1.2 Security monitoring mechanism.....	17
II.1.3 Security supervision interest.....	19
II.2 Machine learning in security.....	20
II.2.1 Machine learning definition.....	20
II.2.2 Machine learning for anomaly detection.....	20
II.2.3 Machine learning techniques	21
II.3 Log files	23
II.3.1 Logs collection protocol (Syslog)	24
II.4 SIEM.....	27

II.4.1	SIEM Definition	27
II.4.2	History and evolution of SIEM	28
II.4.3	Theory and process flow of SIEM.....	29
II.4.4	Benefits of SIEM	33
II.4.5	SIEM solutions type	34
II.5	Solution proposed	39
Conclusion.....		39
Chapter III: Design and implementation		40
Introduction.....		41
III.1	Conceptual study.....	41
III.1.1	Context definition	41
III.1.2	Actors	42
III.1.3	Context diagram	42
III.1.4	Modeling.....	43
III.1.5	Conceptual solution	54
III.2	ELK Stack.....	60
III.2.1	ELK definition.....	60
III.2.2	ELK Architecture	62
III.2.3	ELK components	62
III.2.4	Employing ELK Stack for SIEM solution.....	73
III.3	Wazuh	76
III.3.1	Wazuh definition	76
III.3.2	Wazuh features	76
III.3.3	Wazuh components.....	78
III.3.4	Wazuh agents and Wazuh server communication.....	82
III.4	Wazuh and ELK Stack communication.....	83
III.5	Implementation	84

III.5.1	Simplified Network Architecture (DPCASSI - Badr bank)	85
III.5.2	Architecture deployment	86
III.5.3	Lab environment	89
III.5.4	Installation	90
III.6	Wazuh user interface	97
III.6.1	Home page	97
III.6.2	Agents page	98
III.6.3	Security information management	102
III.6.4	Auditing and Policy Monitoring	105
III.6.5	Threat detection and response	109
III.6.6	Regulatory compliance	113
III.6.7	Wazuh roles	114
III.7	Elastic user interface	115
III.7.1	Integrations	115
III.8	Alerts notifications	116
	Conclusion	118
	Chapter IV: Test and results	119
	Introduction	120
IV.1	Detecting a Brute-force attack:	120
IV.1.1	Overview	120
IV.1.2	Prerequisites	121
IV.1.3	Test Steps	121
IV.1.4	Test Results	121
IV.2	File integrity monitoring	124
IV.2.1	Overview	124
IV.2.2	Prerequisites	124
IV.2.3	Test Steps	124

IV.2.4	Test Results	124
IV.3	Detecting Unauthorised Processes.....	128
IV.3.1	Overview	128
IV.3.2	Prerequisites	128
IV.3.3	Test Steps.....	129
IV.3.4	Test Results	129
IV.4	Detecting an SQL Injection attack.....	130
IV.4.1	Overview	130
IV.4.2	Prerequisites	130
IV.4.3	Test Steps.....	131
IV.4.4	Test Results	131
IV.5	Detecting a Shellshock attack.....	132
IV.5.1	Overview	132
IV.5.2	Prerequisites	132
IV.5.3	Test Steps.....	132
IV.5.4	Test Results	132
IV.6	Detecting Suspicious Binaries	133
IV.6.1	Overview	133
IV.6.2	Prerequisites	133
IV.6.3	Test Steps.....	134
IV.6.4	Test Results	134
IV.7	Osquery integration.....	135
IV.7.1	Overview:	135
IV.7.2	Prerequisites	135
IV.7.3	Test steps	136
IV.7.4	Test Results	136
IV.8	Detecting malware - Virus Total integration	136

IV.8.1	Overview	136
IV.8.2	Prerequisites	137
IV.8.3	Test steps	137
IV.8.4	Test Results	138
	Conclusion.....	139
	Conclusion and future work	140
	Bibliography	142

List of figures

Figure 1: Badr Bank logo	6
Figure 2: BADR Bank general organizational chart	9
Figure 3: Part of the general chart	10
Figure 4: Organization Chart of the Direction of the Continuity and Security	11
Figure 5: Security monitoring	18
Figure 6: Log message example (He et al., 2020)	23
Figure 7: Log files sources	23
Figure 8: Deployment between devices and Syslog server	25
Figure 9: Syslog information format	25
Figure 10: SIEM steps	29
Figure 11: Event normalization example.....	31
Figure 12:Event aggregation example	32
Figure 13: Elasticsearch Logo	35
Figure 14: Wazuh Logo	38
Figure 15: Quadrant Sagan Logo.....	38
Figure 16: Context diagram	42
Figure 17: Package diagram	43
Figure 18: Log management use case diagram	44
Figure 19:Data visualization use case diagram	47
Figure 20: Information security use diagram	49
Figure 21: Detection and response to threats use case diagram	51
Figure 22: Threat detection and response use case diagram	52
Figure 23: Block diagram	54
Figure 24: Log collection module	55
Figure 25: Log normalization module	56
Figure 26: Log consolidation module.....	57
Figure 27: Event Analysis Module.....	58
Figure 28:Incident Management Module	58
Figure 29: Alert Management Module	59
Figure 30: Class diagram of the log normalization module	60
Figure 31 : ELK Stack	61
Figure 32: ELK Stack component	61

Figure 33: ELK Stack architecture	62
Figure 34: Elasticsearch and Lucene integration.....	63
Figure 35: Elasticsearch cluster	64
Figure 36: Elasticsearch index.....	64
Figure 37: Elasticsearch partition and shard.....	65
Figure 38: Elasticsearch replica.....	65
Figure 39: Logstash architecture	67
Figure 40: Logstash pipeline	68
Figure 41: Logstash internal architecture.....	69
Figure 42: Beats architecture	70
Figure 43: Kibana architecture	71
Figure 44: X-Pack components	72
Figure 45: Wazuh components architecture	78
Figure 46: Wazuh Agent architecture.....	79
Figure 47: Wazuh server architecture.....	81
Figure 48: Wazuh and Elasticsearch communication.....	84
Figure 49: DPCASSI Badr bank network architecture.....	85
Figure 50: SIEM ALL-in-one deployment architecture	87
Figure 51: SIEM Distributed deployment architecture	88
Figure 52: Lab environment architecture	90
Figure 53: Wazuh home page	97
Figure 54: Wazuh home page 2	97
Figure 55: Wazuh agent page	98
Figure 56: Deploying agent 1	99
Figure 57: Deploying agent 2	99
Figure 58: Agent dashboard 1	101
Figure 59: Agent dashboard 2	101
Figure 60: Security events dashboard.....	102
Figure 61: Security events, events.....	103
Figure 62: Integrity monitoring dashboard.....	104
Figure 63: Integrity monitoring events.....	104
Figure 64: Policy monitoring dashboard	105
Figure 65: Policy monitoring events	105
Figure 66: System auditing dashboard	106

Figure 67: System auditing events.....	106
Figure 68: SCA inventory	107
Figure 69: SCA events.....	108
Figure 70: VirusTotal dashboard.....	109
Figure 71: VirusTotal events	109
Figure 72: MITRE ATT&CK techniques.....	111
Figure 73: MITRE ATT&CK valid accounts technique	111
Figure 74:MITRE ATT&CK dashboard	112
Figure 75: MITRE ATT&CK events.....	113
Figure 76: Wazuh regulatory compliance	114
Figure 77: Wazuh roles.....	115
Figure 78: Elastic integrations	116
Figure 79: Stack platform icon	116
Figure 80: Integrating Slack with Wazuh.....	117
Figure 81: Alerts notification in Slack	117
Figure 82: Percentage of ICS computers on which malicious objects from various categories were blocked.....	120
Figure 83: Graphic of the alerts generated by Wazuh in response to the ssh attempt.....	122
Figure 84: Graphic of the alerts generated by Wazuh in response to the brute force attack	123
Figure 85:Graphic of the alerts generated by Wazuh in response to adding a file.....	125
Figure 86: Graphic of the alerts generated by Wazuh in response to deleting a file.....	126
Figure 87: Graphic of the alerts generated by Wazuh in response to adding a file 2.....	127
Figure 88:Graphic of the alerts generated by Wazuh in response to modifying a file	127
Figure 89:Graphic of the alerts generated by Wazuh in response to the execution of the black-listed Netcat command	130
Figure 90: Graphic of the alerts generated by Wazuh in response to SQL injection	131
Figure 91: Graphic of the alerts generated by Wazuh in response to the Shellshock attack	133
Figure 92: Graphic of the alerts generated by Wazuh in response to the suspicious binary	135
Figure 93: Graphic of the alerts generated by Wazuh in response to Osquery	136
Figure 94: Graphic of the alerts generated by Wazuh in response to malware detection .	138

List of tables

Table 1: Facilities numbers Syslog.....	26
Table 2: Severity levels Syslog	27
Table 3: Most popular SIEM tools 2022	35
Table 4: Pros and cons of Elasticsearch	36
Table 5: Pros and cons of OSSIM	37
Table 6: Pros and cons of Wazuh	38
Table 7: Pros and cons of Quadrant Segan.....	39
Table 8: Beats types.....	70
Table 9: Elastic Stack features subscriptions	75
Table 10: ELK Features.....	75
Table 11: Comparison table for ELK and Wazuh features	83
Table 12: Wazuh and ELK communication protocols	84
Table 13: All-in-one deployment hardware requirements.....	87
Table 14: Distributed deployment hardware deployment	88
Table 15: ELK Stack versions compatibility.....	89
Table 16: Rules type.....	108
Table 17: MITRE ATT&CK Enterprise Matrix.....	110
Table 18: Wazuh alerts information developed in response to the ssh attempt	122
Table 19: Wazuh alerts information generated in response to the brute-force attack	123
Table 20: Wazuh alerts information generated in response to the creation of a file	125
Table 21: Wazuh alerts information generated in response to deleting a file	126
Table 22: Wazuh alerts information generated in response to changing the content of a file	128
Table 23: Wazuh information generated in response to the black-listed Netcat Command	130
Table 24: Wazuh information generated in response to the SQL Injection attack.....	131
Table 25: Wazuh information generated in response to the Shellshock attack.....	133
Table 26: Wazuh information generated in response to the suspicious binary	135
Table 27: Wazuh information generated in response to Osquery	136
Table 28: Wazuh information generated in response to malware detection.....	138

Acronyms and Abbreviations

API	Application Programming Interface
CLF	Common Log Format
DNS	Domain Name Server
ELK	Elasticsearch, Logstash et Kibana
HTTP	Hypertext Transfer Protocol.
IDS	Intrusion Detection System
IP	Internet Protocol
IPS	Intrusion Prevention System
IS	Information System
IT	Information Technology
OS	Operating System.
PCA	Principal Component Analysis
TCP	Transmission Control Protocol
SEM	Security Event Management
SIEM	Security Information and Event Management
SIM	Security Information Management
SOAR	Security Orchestration, Automation, and Response
SOC	Security Operation Center
SQL	Structured Query Language.
SSH	Secure Shell
SVM	Support Vector Machine
UEBA	User and Entity Behavior Analytics

Introduction



Introduction

With the development and steady growth of several technologies in all areas of our lives, it has become indispensable for us to identify with our personal information. However, securing our information is a significant challenge; it doesn't include only the protection of personal data but also the data of organizations that can be very sensitive: trade secrets, product designs, customer information, etc. In addition, this information must be protected against unauthorized access by internal or external persons of the organization or theft by these competitors.

More technological developments mean more cyber threats, which are becoming a massive nightmare in the computer science field. With the evolution of cybersecurity, cyber threats have become more sophisticated. It is no longer enough to count on firewalls, intrusion protection systems (IPS), intrusion detection systems (IDS), antimalware, and antivirus to protect the environment against these threats.

There are many reasons for the security measures in computer networks, such as vulnerabilities, security policy errors, incorrect configurations, etc.; it is becoming easier for cyber criminals to exploit network configuration and security policy vulnerabilities to execute different penetration strategies. These strategies are directed to various network resources.

The information system is an essential asset of any organization; thus, it has become necessary to help improve an organization's threat detection and response capabilities by monitoring real-time events and maintaining long-term data to detect abnormal use patterns and alert organizations whenever necessary.

The complexity of managing the security of computer networks leads to the need to develop powerful automated security analysis components. For example, a Security Event and Information Management System (SIEM) can receive and capture events from network equipment, such as switches or routers, and security equipment, such as proxies or gateways, or directly from certain services installed on machines such as Web server logs. This system must make it possible to find and correct errors in the network configuration, detect different security threats, identify critical network resources and choose an effective security policy and security mechanisms that are appropriate to current threats.

As part of our graduation project, we implement a “SIEM” Security Event and Information Management System. This monitoring tool allows the security analyst to verify that the network configuration settings and security procedures provide the necessary level of security. Furthermore, perform analyses on the log files collected from the different sources related to the network organization system. The analysis makes it possible to detect anomalies, identify bugs, detect threats and vulnerabilities, and also available to evaluate the security level. Furthermore, this tool offers dashboard functionality that allows real-time monitoring and alerting the administrator in case of suspicious behavior—in the end, proposing a platform for the organization’s employees to facilitate the communication between them and get notified of all the alerts the system receives.

Internship objective

As part of this project, we will perform the theoretical study (operation and design of a hypothetical SIEM) and implementing a "SIEM" security management tool. The features expected by the BADR for the security system are the functionalities of a modern SIEM (known as analytical SIEM), so we went through: collect, normalize, aggregate, correlate and analyze the data of events from the machines, systems, and applications (firewalls, IDS, IPS, Network machines such as routers and switches, security machines, applications, various databases, servers, etc.) and finally create active response mechanisms in the event of a security incident.

This project focuses on the following points:

- Collecting data from various sources;
- Normalization and aggregating of the data collected;
- Analyzing the data to discover anomalies;
- Identifying security vulnerabilities and allowing organizations to investigate alerts;

The details are mentioned in the chapter “Design and implementation”.

The structure of our document is as follows

Chapter 1: “Host organization and study the existing”: We introduce the host organization BADR Bank and its strategies and mission, including the purpose of our internship, and identify the organization's problems and the problematic.

Chapter 2: “Theoretical background”: We define the concept of supervision and security monitoring and machine learning integration in security supervision. Moreover, we describe the utility of log files in our work and the interest in this technique within an organization. In addition, we mention in detail the monitoring tool that we use, the SIEM, and explain its functionality. Next, we will compare open-source tools to choose a solution to implement.

Chapter 3: “Design and implementation ”: We present the architecture and modeling of the system using the UML language. Next, we offer the detailed architecture of each component of our solution and explain the main functions of visualization, including the installation and configuration steps; also, we illustrate the different functionalities that our solution offers.

Chapter 4: “Tests and results”: We perform tests on our environment to observe the solution's functionality and register the results for validating the solution.

Chapter I



Host organization
and study the existing

Introduction

Financial institutions are vulnerable to cyberattacks, and security measures and controls must be included to identify the blind spots in their information system. Security personnel is qualified for this mission.

On these terms, we were headed to Badr bank, one of the most significant financial institutions in the country.

This chapter is dedicated to the host organization for our practical internship; it contains a brief passage of their history and advancement, including their mission, fundamental activities, and a list of their few strategic objectives, as well as a presentation of their general organizational chart and, in particular, the service Audit of SSI service, where our internship has been. Finally, we clarify the purpose of the problem and the objectives we must reach at the end of the internship.

I.1 Presentation of BADR Bank



Figure 1: Badr Bank logo

I.1.1 History and advancement of BADR Bank

The Bank of Agriculture and Rural Development is a national financial institution abbreviated as a suite B.A.D.R, and it is a company per share with a current share capital of 33.000.000.000DA. (BADR Banque - 318 Mots | Etudier).

I.1.2 Missions of BADR Bank

The Bank of Agriculture and Rural Development (BADR) was made on 13 March 1982 by declaration N°82-106 laying out the resolutions of the BADR. By ethicalness of Law 90/10 of 14 April 1990 on cash and credit, as changed and enhanced by Order 03-11 of 26

August 2003, BADR turned into a legitimate individual doing the activities of receipt of public assets, the tasks of giving credits, as well as making the method for installment and the board accessible to clients.

BADR is an institution per share, which is liable for guiding public monetary endeavors and helping them in the utilization and the executives of the method for installment made accessible to them, inconsistent with banking secrecy. (BADR Banque - 318 Mots | Etudier).

I.1.3 Fundamental activities of BADR Bank

The formation of BADR expects to guarantee the country's monetary freedom, increase the expectation of living of rural populations, and restructure the agricultural system. Among the fundamental undertakings of the BADR (Présentation de La Banque BADR - CAW JIJEL) :

- The handling of all credit, foreign exchange, and treasury transactions;
- Opening accounts for anybody applying.
- Receiving demand and term deposits.
- Support in the assortment of investment funds.
- Commitment to the improvement of the agricultural sector.
- Guaranteeing the advancement of agricultural, agro-food, agro-industrial, and art exercises.

As well, the Bank for Agriculture and Rural Development has exercises abroad, including:

- Establishing effective foreign relations for the execution of unfamiliar exchange tasks.
- Exchange, including account keeping.
- Opening accounts and directing banking.

I.1.4 Strategic objectives of BADR Bank

The improvement of the nature of administration and client relations makes the BADR the most significant bank with more noteworthy productivity. These goals will be accomplished through (Présentation de La Banque BADR - CAW JIJEL):

- Expanding assets at the best costs and making them beneficial through practical and differentiated advances in consistence with the principles.
- The bank's rigorous cash management in both dinars and currencies.
- Guaranteeing the smooth advancement of the bank in the space of its business.
- Growing and redeploying its organization.
- The fulfillment of its clients by offering them items and administrations liable to address their issues.
- Adjusting dynamic collections management.
- Business advancement through the presentation of new administrative procedures like advertising and the inclusion of another product offering.

I.2 Badr bank organization

BADR is a public bank whose mission is the improvement of the agricultural sector and to advance the rural world. Its network currently consists of more than 303 agencies and 41 regional directorates. It has more than 7,000 directors and employees working in central, regional, and local structures.

I.2.1 Direction of the continuity plan of the information system

The General Directorate of BADR is divided into Adjuncts, General Directorates, and Divisions. The following divisions demonstrate the general organizational chart of BADR Bank:

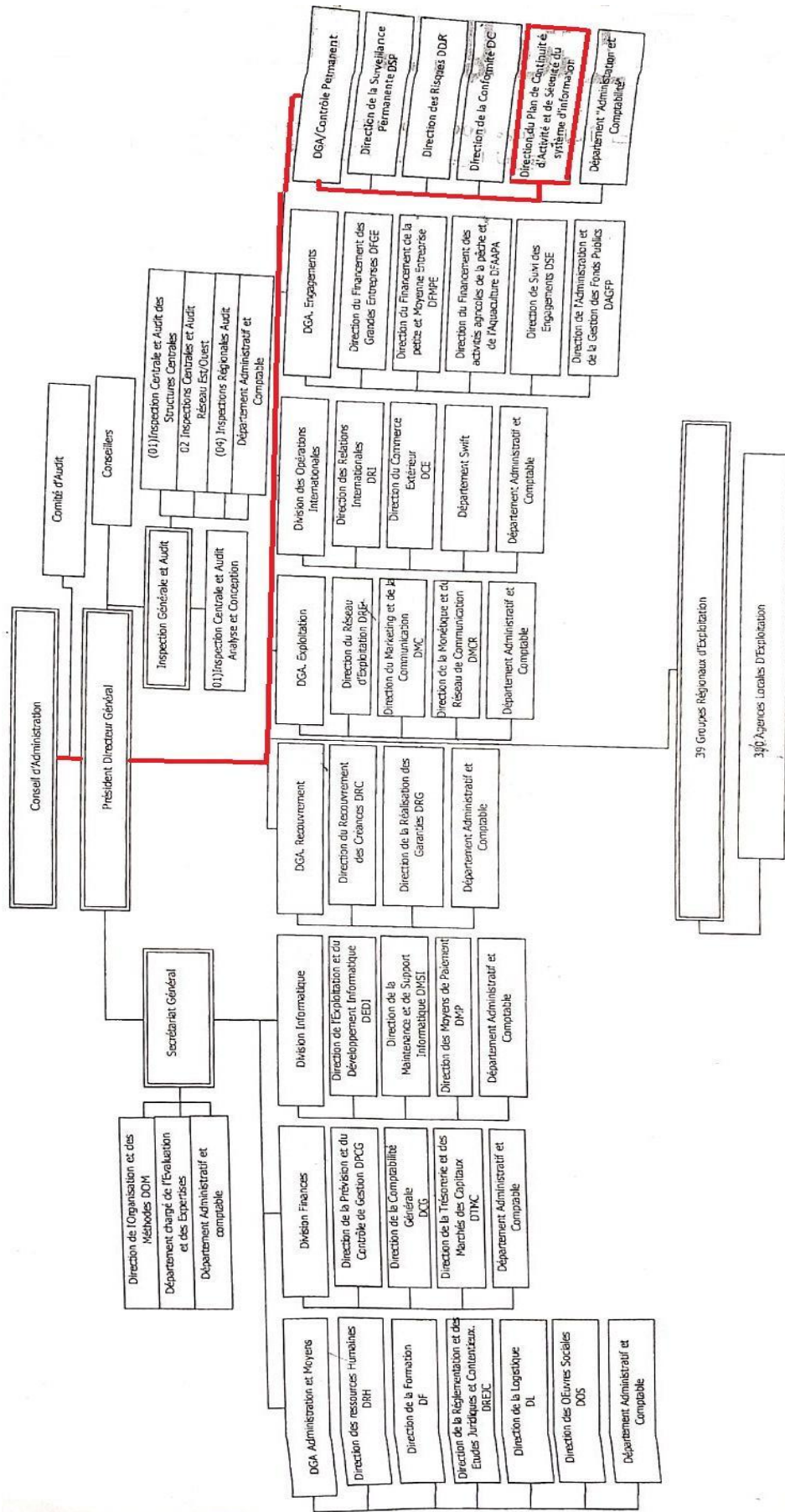


Figure 2: BADR Bank general organizational chart

We focus on the Permanent Control Directorate, which is divided into directions.

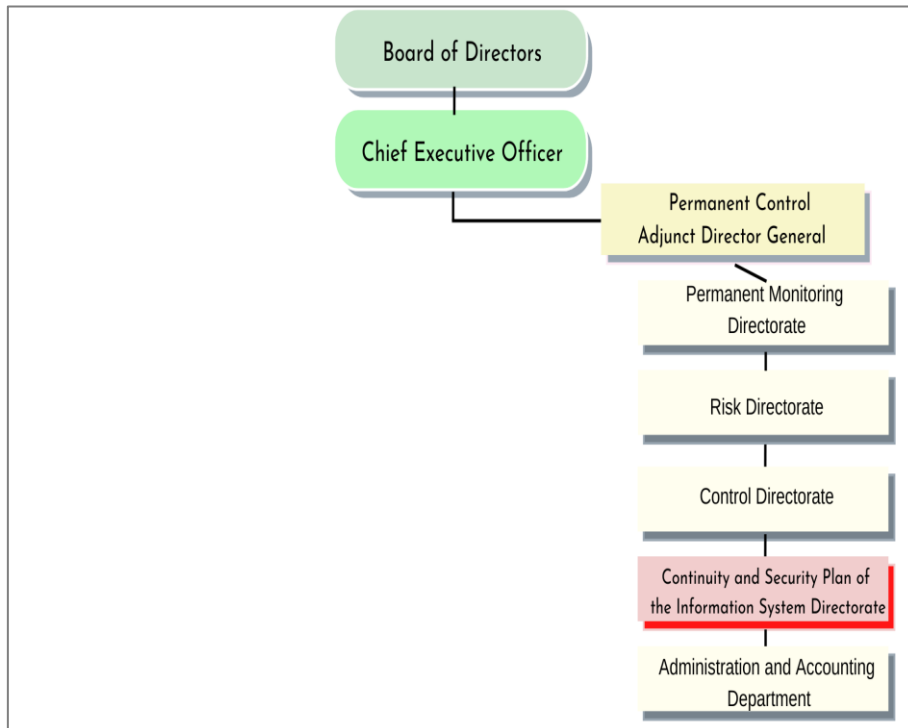


Figure 3: Part of the general chart

Our project will take place within the direction of the continuity and security plan of the information system abbreviated (DPCASSI), which is situated in Bir Mourad Rais, Algiers.

It is placed under the authority of the Adjunct Director-General of Permanent Control (DGA) and headed by a Central Director, the direction of the continuity and security plan of the information system is structured as follows (DPCASSI-ORGANIGRAMME):

- Sub-directorate of Continuity Plan: comprising 02 compartments:
 - Continuity Strategies Compartment
 - Emergency Technical Architecture Compartment
- Sub-directorate of Information System Security: comprising 02 compartments:
 - SSI Standards & Standards Compartment
 - Audit of SSI Compartment

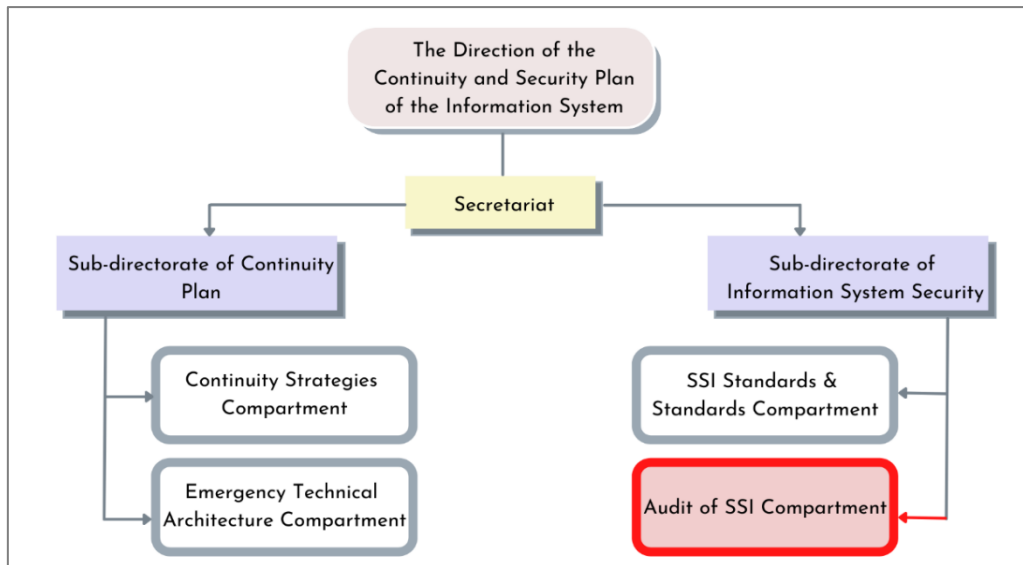


Figure 4: Organization Chart of the Direction of the Continuity and Security Plan of the Information System (DPCASSI)

We are interested in the Audit of the SSI Compartment, where they took charge of our internship.

I.2.2 Audit of SSI Compartment general missions

- Ensure that the bank's information system maintains a consistent level of security under ISS standards.
- Ensure the conduct of the risks analysis, malfunctions, and improve margins of the bank's information system in terms of security.
- Ensure the production of dashboards on the performance of ISS risk management processes, with indicators and control points that will allow one part to verify the effective implementation of security measures and, on the other part, Investigate if corrective actions are required.
- Participate in the establishment of ISS risk prevention plans
- Participate in selecting safety measures and implementation plans following security risk analysis methods and standards.
- Ensure the conduct of regular and accurate security audits of the bank's information system to guarantee the prevention of IT and information system risks: Evaluation of IT risks, physical security, logical security, change management, contingency plan, etc.
- Audit and Control the application of the directives of the PSSI.

I.3 Criticism of the existing

After meeting the employees of the Audit of SSI Compartment and posing inquiries about the organization and its activities, we knew that the Audit SSI team focused on the penetration testing of the web applications that are related to Badr.

We had an overview of their IT system, “Badr Imtiaz”, launched at the agency Badr of Chéraga. Its goal is to provide e-banking service more efficiently at all Badr branches, including the ability to carry out more online transactions. As well, Badr Imtiaz allows the customers to conduct their banking transactions remotely and in real-time.

Mr. LADJOUZI introduces us to some tools used for penetration testing; most devices are defined in Kali Linux, a free and open-source Linux-based operating system geared at advanced penetration testing and security auditing, such as Nmap; Wireshark, Nessus...

- Nmap: Network Mapper is a free port scanner, one of the tools on Kali Linux for information gathering. It facilitates insights about the host, IP address, OS detection, and network security details like open ports.
- Wireshark: Wireshark is a free package analyzer. It is used in troubleshooting and analysis of computer networks.
- Nessus: an IT security tool. It points out potential or proven weaknesses on the machines tested.

As well as other solutions in case to maximize its chances of countering threats and reduce the occurrence of possible incidents, the security of their information system

- Firewall: A firewall is a computer network element; it can be software, hardware, or a combination of both. The firewall interconnects network security levels, and its role is to secure the IT network by defining authorized and prohibited communications according to security rules.
- Intrusion Detection System (IDS): An IDS is a mechanism that aims to detect any potentially suspicious traffic on the target being scanned. It sees abnormal activities moving away from the standard.
- Antivirus: Viruses are computer programs capable of being duplicated by themselves and more or less seriously disrupt the operation of the computer

infected: machine slowing down, deleting files, destroying data. An antivirus will continuously scan the workstations to detect, block or eradicate viruses in real-time.

- Antimalware: Malware is a generic term that includes all types of malware: viruses, Trojan horses, worms, and ransomware. The anti-malware protects systems from any kind of malware.
- Anti-spam: Spam is all emails we receive in the mailbox without requesting anything. Generally, these are large volume shipments for advertising purposes. The anti-spam software aims to sort the mail by placing all these unwanted emails in an appropriate directory.
- Vulnerability Scanner: These are programs designed to detect and find security vulnerabilities of computer and communications systems, to fix them before hackers exploit them.

After a global overview, we identified the following issues:

- Absence of a supervision tool within the bank.
- Difficulties in identifying breakdowns and managing outages services.
- Wasted time diagnosing and resolving incidents.
- The administrator is not alerted in case of abnormal operating problems.

We concluded that, due to the absence of the visualization and detection real-time tools, the system administrator could not recognize security occurrences, like unapproved access to sensitive files, modification in configurations of the security system, etc., on time, because of unfortunate administration of the logs.

I.4 Problematic

Like any other bank, BADR Bank represents a high-profile target for cybersecurity attacks, and their data about customers is precious to hackers looking to carry out an attack. Therefore, banking information security and privacy are essential, denoting one of the most critical challenges facing information security practitioners. The bank also has several information systems which contain necessary information. Hence, the need to protect these systems from unauthorized access, intrusion, and information leaks.

Implementing SIEM altogether reduces threats of computer attacks by allowing the supervision of all the IT infrastructure of BADR. However, there are a few rules to consider while picking the SIEM arrangement appropriate for BADR's IT infrastructure.

In the first place, the SIEM should make it possible to collect and gather logs from numerous sources, classify and analyze incidents and events, and send real-time alerts, dashboards, and reports to several critical services. The second critical to be considered is the cost of the solution. Most SIEM paying requires a high budget.

Finally, picking a viable SIEM with the organization's unique IT infrastructure and assets will be essential.

Conclusion

We saw that the host organization, Badr bank is one of the financial institutes in Algeria; a financial bank is always an easy track for cyberattacks, and securing its information system is the mission of the IT information security team; as SSI students, we will reach at the end of this project, the demand objectives which include designing and implementing a monitoring system.

Chapter II



Theoretical background

Introduction

IT supervision is becoming indispensable for any information system; the supervision varies from one environment to another. Therefore, information system monitoring is integral to organizational continuous monitoring and incident response programs.

In this chapter, we introduce the basics of IT supervision and security monitoring, including the use of machine learning in the security field, passing by knowing the utility of log files and giving examples of Syslog, also a definition of security information and event management tools, counting the existing types.

II.1 IT supervision and security monitoring

II.1.1 IT supervision definition

In computer science, supervision is a monitoring technique to analyze, report, and abnormal alert operations of computer systems.

Information Technology supervision involves gathering measurements on equipment activity and programming in an IT climate to guarantee that applications and administrations are supported as expected.

Monitoring aims to capture all the critical actions and performance indicators at the different levels of the systems (hardware and application). The objective is threefold:

- Detect anomalies and problems in real-time and be able to act.
- Analyze system load to infer trends, identify correlations, and forecast future needs (processor, RAM, disks, network, etc.).
- Retrieve and analyze all relevant system and application logs to detect and correct errors.

Two control modes are used to supervise machines:

- **Passive mode:** the device must provide and send the information to the supervision server in case of a particular event.

- The supervised machine verifies its status and independently transmits the result to the supervision server.
- The supervision server receives the alert and processes it.
- **Active mode:** the supervision server has the initiative to query the machine to get the information
 - The supervision server sends a request to the supervised machine.
 - The machine answers the server's request.
 - The server analyzes the information and determines the status of the machine.

There are several computer-monitoring platforms and software. The choice of the most suitable software for the enterprise depends on various factors, such as the specific requirements of the individual infrastructure, the type of monitoring preferred, or the notification function desired. In addition, software should not burden a system despite regular and thorough checks.

- *Physical monitoring:* is the proper functioning of the technical part (the state of servers, workstations, hard drives, and memory).
- *Application monitoring:* is the user experience and the user's feeling when he works on his job (position too slow, problems of access to the company's resources).
- *The monitoring of the quality of service:* is the compliance of the contract between the IT provider and the customer (response time of hotlines, time of restoration of internet access).
- *The monitoring of business applications:* is the accessibility to the applications necessary for developing the company's activity (databases, CRM).

II.1.2 Security monitoring mechanism

Security monitoring involves collecting and analyzing information to detect suspicious behavior or unauthorized system changes on your network, defining which types of behavior should trigger alerts, and taking action on signals as needed. (Liang et al., n.d.)

This supervision allows mastering the various elements related to the security of the server fleet: accounts and passwords, configurations, and network monitoring. This mechanism

must enable or guarantee consistency and compliance. From an operational point of view, it is a matter of implementing a technical security policy and then detecting any discrepancies from a central point (centralization of logs and analysis of incidents).

Supervision will make it possible to facilitate these actions according to several axes:

- Facilitate the measurement of the gap between the security policy master plan (SDSSI) and the configuration of the servers.
- Validate the implementation of new security actions.
- Facilitate the implementation and deployment of this security.
- Check this security over time (compliance).

Organizations generally use a SIEM information and event or log management tool. This solution can help with the survival of security information as well; it has compliance and response to incidents by:

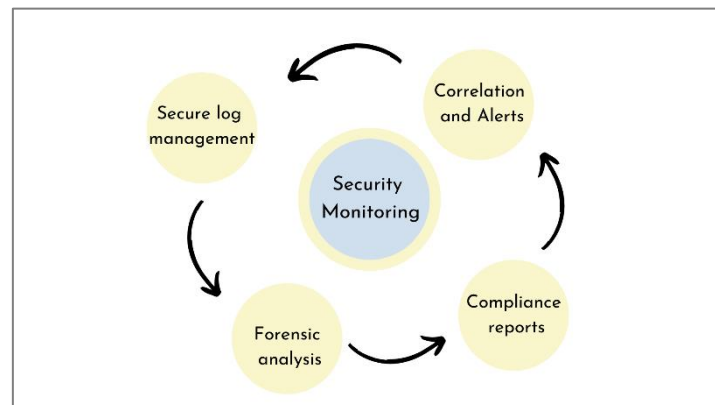


Figure 5: Security monitoring

- Collecting and aggregating log data generated throughout the information system, from applications to network and security devices, such as firewalls and antimalware detections;
- Analyze and correlate the collected information to identify attacks as quickly as possible and help respond to intrusions faster;
- Create custom reports for better visualization of organizational security;
- Improve efficiency and help focus IT risk management staff on essential events;
- Securely store all event data on a network for long-term retention and allow instant access to archived data;

Monitoring meets two main objectives:

- **Provide reports on security incidents and events:** successful and unsuccessful connections, malware activity, and other possible malicious activity.
- **Send alerts:** To monitor the entire infrastructure from one place, it needs to receive notifications when critical changes occur; it actively checks the metrics, integrations availability, network endpoints, and more.

Machine learning is also applied in monitoring mechanisms to analyze data and find patterns for better malware detection.

II.1.3 Security supervision interest

Security supervision makes it easier for enterprises to manage security by filtering massive amounts of security data and prioritizing the security alerts the software generates:

- **More effective detection of security incidents:** Making links between the defined and the coming events, such as the hack via the network, manipulation of specific equipment, etc. It also allows the detection of an incident more quickly, knowing the potential extent from the beginning and assessing the damage.
- **Faster Defensive Response:** This allows for a more immediate defensive response and avoids wasted time in research and diagnosis.
- **Incident Management:** communicating with most systems generating event logs such as desktops, firewalls, anti-virus systems, etc. In case of identifying an incident that causes an alert will generate to investigate the problem further.
- **Meeting legal compliance requirements:** Compliance reporting is automated and centralized. This process accelerates the identification and analysis of security events and recovery.
- **Global vision of the network at all times:** verification of employees' activities on the web and notification in the event of anomalies. It also identifies software and device activities.

II.2 Machine learning in security

II.2.1 Machine learning definition

Machine learning is the discipline of Artificial Intelligence (AI)'s underlying relationships between information and data. (10th International Conference on Cloud Computing & Institute of Electrical and Electronics Engineers, 2020) It refers to algorithms and processes that “learn” in generalizing past data and experiences to predict future outcomes.

At its core, machine learning is a set of mathematical techniques implemented on computer systems that enables a process of information mining, pattern discovery, and drawing inferences from data. (Clarence Chio et al., 2018).

Machine learning technologies can be applied in many applications, such as outlier detection, character recognition, robotics, natural language processing, and automatic speech recognition to exchange audio information into strings of words.

It is also used in general routines like web searching, advertisement post, stock market predictions, behavioral study, predicting climate, big data analytics, and other applications. It plays a significant role in application development. Machine learning simplifies training experience and hypothetical results that estimate to fulfill the objective of targeted output. Machine learning allows the systems to act on anonymous information and correctly forecast the following information. Machine learning aims to predict future events. (Rajendra Kumar Dwivedi, Arun Kumar Rai, et al., 2020).

II.2.2 Machine learning for anomaly detection

In security, machine learning learns by analyzing data to find patterns for better malware detection in encrypted traffic, find threats, predict unsafe browsing, or protect data in the cloud by uncovering suspicious user behavior (Dorsey et al., 2020).

- *Find threats on a network:* Machine learning detects threats by monitoring the behavior of the network for anomalies. Machine learning engines process massive amounts of data in real-time to discover critical incidents. These techniques allow for the detection of insider threats, unknown malware, and policy violations.
- *Keep people safe when browsing:* Machine learning can predict “bad neighborhoods” online to help prevent people from connecting to malicious websites. In addition,

machine learning automatically analyzes Internet activity to identify attack infrastructures staged for current and emergent threats.

- *Provide endpoint malware protection:* Algorithms can detect malware trying to run on endpoints. It identifies new malicious files and activities based on the attributes and behaviors of known malware.
- *Protect data in the cloud:* Machine learning can protect productivity by analyzing suspicious cloud app login activity, detecting location-based anomalies, and conducting IP reputation analysis to identify threats and risks in cloud apps and platforms.
- *Detect malware in encrypted traffic:* Machine learning can detect malware by analyzing encrypted traffic data elements in standard network telemetry. Machine learning algorithms pinpoint malicious patterns to find threats hidden with encryption.
- *Anomaly detection:* Machine learning helps prevent fraud, adversary attacks, and network intrusions that can compromise the organization's future by finding and identifying outliers.

What we are interested in for our project is machine learning in anomaly detection. '*Anomaly detection refers to identifying unexpected intruders or breaches.*' (Clarence Chio et al., 2018).

II.2.3 Machine learning techniques

There are various types of machine learning techniques used for outlier detection. These techniques are categorized into unsupervised machine learning and supervised machine learning. (Rajendra Kumar Dwivedi et al., 2020).

- **Supervised learning** consists of input variables (x) and an output variable (Y). It uses an algorithm to learn the mapping function from input to output. $Y = f(X)$
It uses tagged or labeled training information to produce models for simplifying the relationship between a feature vector (inputs) and a supervisory signal or class label (outputs). The objective of learning models is to reduce the errors for a specific set of information.

Supervised learning is generally classified into two types: Linear Regression and Classification. Classification is further categorized into decision trees and Support Vector Machines (SVM).

a) *Decision Tree:* A *decision tree is a learning technique that* uses a treelike structure. It shows the decisions and their significance. A decision tree consists of the resource's costs, utility, and outcomes. It is a mechanism to present an algorithm that only contains conditional statements.

b) *Support Vector Machine (SVM):* SVM is a method that analyzes the statistics for predictions. Support vector machines are points close to the hyper-plane, creating different classes. SVM is mainly used for anomaly detection, regressions analysis, and classification.

- **Unsupervised Learning** It is used to find structures from unlabeled data sets where output is unknown. This learning is generally used in data compressions, anomaly detection, etc. The technique uses probabilistic data models for training. The most common models of this learning are clustering and visualization. Here we are discussing K-Means Clustering and Principal Component Analysis (PCA) as follows:

a) *K-Means Clustering:* K-Means is a method whose implementation is easy on large datasets. It is used in various applications such as computer vision, image processing, etc. This method is frequently used as a pre-processing phase for different methods to find the initial pattern.

b) *Principal Component Analysis (PCA):* PCA is a learning algorithm commonly used for dimensionality reduction. This method aims to convert higher dimensional data into a lower dimension to minimize the computing cost. This is the most frequently used multivariate system for finding anomalies, new informative features, and uncorrelated structures.

The data used in this phase is a file named log files generated by any source that contains information about usage patterns, activities, and operations within an operating system, application, server, or another device.

II.3 Log files

Logs are unstructured plain text printed by logging statements to detect application debug, availability, and security logging. A log message is generated by a system or a device to inform a specific event; it is recorded with a set of fields such as a timestamp (the occurrence time of the event), user information, verbosity level (the severity level of the event), and the event description.

```
2008-11-09 20:46:55,556 INFO dfs.DataNode$PacketResponder:
Received block blk_3587508140051953248 of size 67108864 fr
om /10.251.42.84
```

Figure 6: Log message example (He et al., 2020)

Log files are generated by applications, containers, databases, firewalls, endpoints, IoT devices, servers, web services, etc.



Figure 7: Log files sources

The information system generates many log files to gather the necessary information. Thus, many kinds of logs exist, including:

- **Event logs:** records data about network traffic and users, for example, login attempts, failed password attempts, and application events.

Firewall, Antivirus, Intrusion Detection System IDS

- **Server logs:** contains movements of every kind of a particular server in a given period. The server provides a detailed understanding of the site or the application accessed.

DNS Servers, Routers, Switches

- **System logs:** holds a record of the operating system events that specify the process of the system. In addition, it shows instructions, errors, and events connected with the computer OS.

Windows, Linux, and macOS all those operating systems generate system logs.

The supervision of the IT infrastructures allows real-time anomaly detection and interaction in a short time. The most sensitive information is stored in the servers; in that case, they use a standard message format to exchange data to communicate between the logging servers in a network device.

II.3.1 Logs collection protocol (Syslog)

Syslog is a protocol characterizing an event log service of a computer system as well as the name of the format that allows these exchanges. (Tsunoda & Keeni, 2014).

Its primary purpose is to transfer by network the different system logs generated by the network devices, such as the firewalls, routers, etc.).

As a protocol, Syslog consists of client and server parts. The client party sends the data to the organization through port UDP 514. It is feasible to utilize TCP. The servers gather data and create logs.

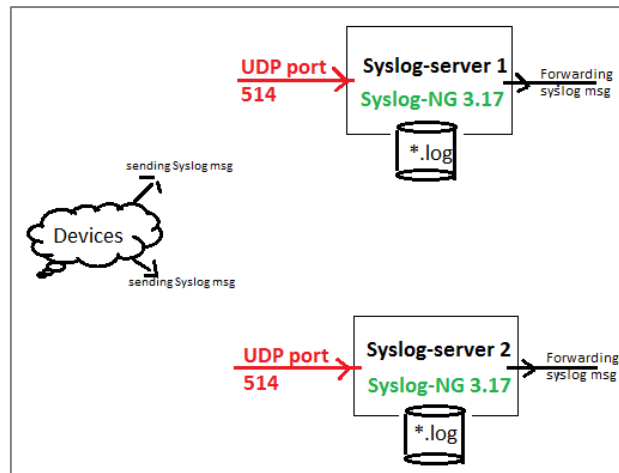


Figure 8: Deployment between devices and Syslog server

- **Syslog Format:** The Syslog message consists of three parts: PRI (a calculated priority value), HEADER (with identifying information), and MSG (the message itself). It contains the following information:
 - Priority
 - Version
 - Timestamp
 - Hostname
 - Application name
 - Process ID
 - Message ID
 - Structured Data
 - Message

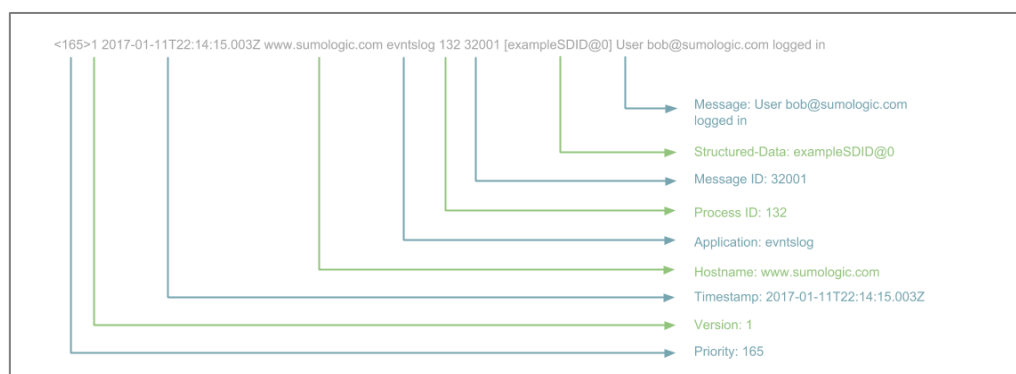


Figure 9: Syslog information format

The priority value is calculated using the formula (Priority = Facility * 8 + Level).

As figure 8 describes, a Syslog message includes a priority value at the beginning of the text. The priority value ranges from 0 to 191. (*Syslog - Definition and Details*, n.d.)

The facility represents the machine process that created the Syslog event. Therefore, the facility means an instructing SMS to the remote Syslog Server only for those events whose facility matches the one defined in the field from 0 to 23.

So, by changing the facility number and the severity level, the number of alerts (messages) sent to the remote Syslog server will also change.

The Facility value is a way of determining which process of the machine created the message. Since the Syslog protocol was initially written on BSD Unix, the Facilities reflect the names of UNIX processes and Daemons. (*Case Solution*, n.d.). The following list represents the available Facilities as per RFC5424:

Table 1: Facilities numbers Syslog

Facility Number	Facility Description	Facility Number	Facility Description
0	kernel messages	12	NTP subsystem
1	user-level messages	13	log audit
2	mail system	14	log alert
3	system daemons	15	clock daemon
4	**security/authorization messages	16	local use 0 (local0)
5	messages generated internally by Syslog	17	local use 1 (local1)
6	line printer subsystem	18	local use 2 (local2)
7	network news subsystem	19	local use 3 (local3)
8	UUCP subsystem	20	local use 4 (local4)
9	clock daemon	21	local use 5 (local5)
10	security/authorization messages	22	local use 6 (local6)
11	FTP daemon	23	local use 7 (local7)
** SMS default			
Note: Items in yellow are the facility numbers available on the SMS.			
Note: These Facility Numbers are only applicable to the I/TX products. The N/NX will not use these values and will always report 0.			

The following table highlight the 08 severity levels of Syslog:

Table 2: Severity levels Syslog

SEVERITY LEVEL	EXPLANATION
**	SEVERITY IN EVENT Default SMS setting for Syslog Security option. This setting will send all events to the remote Syslog system
0	EMERGENCY A "panic" condition - notify all tech staff on call? (Earthquake? Tornado?) - affects multiple apps/servers/sites.
1	ALERT Should be corrected immediately - notify staff who can fix the problem - an example is loss of backup ISP connection.
2	CRITICAL Should be corrected immediately, but indicates failure in a primary system - fix CRITICAL problems before ALERT - an example is loss of primary ISP connection.
3	ERROR Non-urgent failures - these should be relayed to developers or admins; each item must be resolved within a given time.
4	WARNING Warning messages - not an error, but indicated that an error will occur if action is not taken, e.g. file system 85% full - each item must be resolved within a given time.
5	NOTICE Events that are unusual but not error conditions - might be summarized in an email to developers or admins to spot potential problems - no immediate action required.
6	INFORMATIONAL Normal operational messages - may be harvested for reporting, measuring throughput, etc. - no action required.
7	DEBUG Info is useful to developers for debugging the app, not useful during operations.
** SMS default	

II.4 SIEM

II.4.1 SIEM Definition

SIEM, Security information and event management, is a security solution that assists organizations with perceiving potential security dangers and vulnerabilities before they get an opportunity to upset business tasks. It surfaces user behavior anomalies and utilizes artificial intelligence to computerize many manual processes. As a result, it has become a staple in advanced security operation centers (SOCs) for security and the consistency of management use cases. (González-Granadillo et al., 2021).

‘Security information and event management (SIEM) systems are increasingly used to cope with the security challenges involved in critical infrastructure protection. However, these systems have several limitations.’ (di Sarno et al., 2016).

SIEM systems are broadly used to perform real-time monitoring and control of foundation resources. A SIEM system incorporates two previously heterogeneous systems - a security information management (SIM) system that focuses on the analysis of historical data to improve the long-term effectiveness and efficiency of cyber security mechanisms and a

security event management (SEM) system that aggregates data into a manageable amount of information to enable the rapid handling of the security incident. (D. Carr, 2005).

These systems are designed to analyze security information from the monitored infrastructures to discover security breaches.

II.4.2 History and evolution of SIEM

The first notion of Security Information and Event Management is attributed to a report by Gartner Inc. (Williams and Nicolett 2005).

The analysts have distinguished three generations during the evolution of SIEM systems

- The first generation of SIEMs, presented in 2005, provides primary log aggregation for different systems and basic event correlation techniques. This generation consolidates log management and event management, which were already isolated; SEM analyzes log and event data in real-time, providing threat monitoring, event correlation, and incident response, with SIM, which collects, analyzes, and generates log data reports.

They are restricted in the size of data they can process and in refining alerts and visualizations they generate.

- The second generation of SIEM solutions was better equipped to handle big data. For example, these SIEMs can correlate historical log data with real-time events and data from threat intelligence sources.
- Gartner proposed a new generation of SIEM, called "SIEM Analytics," in 2017. The latter integrates advanced techniques such as user and entity behavior analytics ("UEBA") based on machine learning to establish behavioral baselines of users or computer systems and identify anomalies. This includes security automation, orchestration, and response "SOAR," which can help analysts quickly investigate incidents and activate security tools to respond to an incident automatically.

II.4.3 Theory and process flow of SIEM

According to the previous research, SIEM collects data from different IT equipment such as servers, network devices, etc. Then, SIEM stores, normalizes, aggregates, and analyzes this data to discover anomalies, detect threats and allow organizations to investigate alerts.

The figure below displays the entire process flow of the SIEM system :

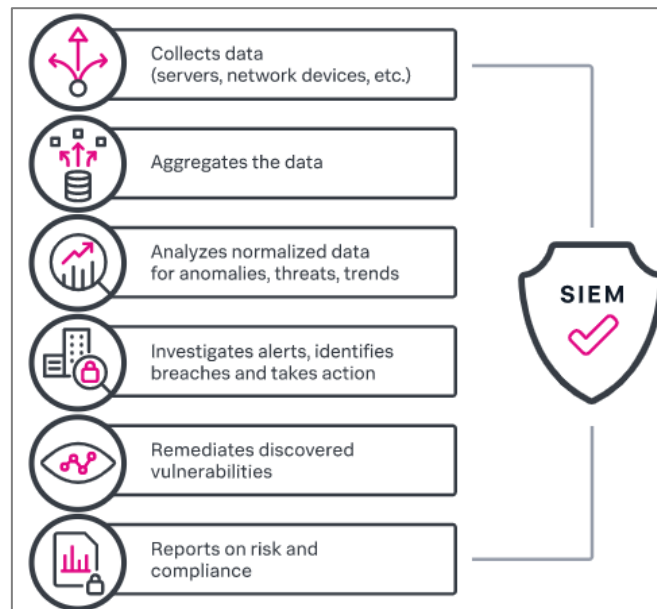


Figure 10: SIEM steps

To allow security managers to monitor events, the SIEM system needs much data to begin its process; the data required is the logs, so the SIEM starts by collecting logs from various sources of information.

Information sources refer to all devices (router, switch, server, etc.) and applications that can generate logs (Figure log file source) that the SIEM system can subsequently collect and process. Operating systems are also sources of information, as they generate logs. These logs display all the statistics of your system: who logged in, who did what on the system, and everything users are doing or what the operating system itself is doing.

II.4.3.1 Data collection

Data collection in a SIEM system relies on logs and events from hundreds of organizational log source systems. Each system generates an event with every new action that happens.

The first step is for the SIEM system to collect the logs generated by the above sources. In general, there are two methods of collecting logs (the Push method and the Pull method), although the actual mechanisms for retrieving logs vary depending on the specific SIEM used.

The push method can facilitate installation and configuration at the SIEM level. Usually, a SIEM solution that uses this method sets up a receiver and then points the source device to that receiver.

Example with Syslog: When setting up the source device using Syslog, set the IP address or DNS name of a Syslog server on the network, and the device will automatically start sending its logs via Syslog to the Syslog receiver.

Unlike the push method, in which the source device sends logs to the SIEM without any SIEM interaction, the pull method requires the SIEM to initiate a connection to the source device and actively retrieve the logs from the source device.

For example: if the logs are stored in flat text files on a network share, SIEM must first establish a connection to the network shares using the stored credentials before it can read the flat text file for the source device logs.

Because each time the SIEM system has to establish a connection, which is not the case with the push method, the pull method has the advantage of being more secure.

II.4.3.2 Data normalization

At this stage, logs from the various sources of information are already transmitted to the SIEM system. However, they are still in their native format and therefore unusable for the SIEM system.

For these logs to be helpful, they must first be reformatted into a single standard format that SIEM can use. Standardization is turning all these different log files into a single format. Every kind of SIEM will handle the standardization act in different ways, but the result is that all logs, regardless of device type or manufacturer, look the same in SIEM.

The following figure shows an example of an event ‘Connection rejection’, the log source is a firewall:

1. Raw event stream received log by a collector.
2. Collector finds which log source type besides event and load parser or take it from the cache.
3. For each event, applied parser. A parser is a set of regex. Each regex is used to find the event's field(source IP, destination port, username, etc.).
4. Event normalized and categorized.

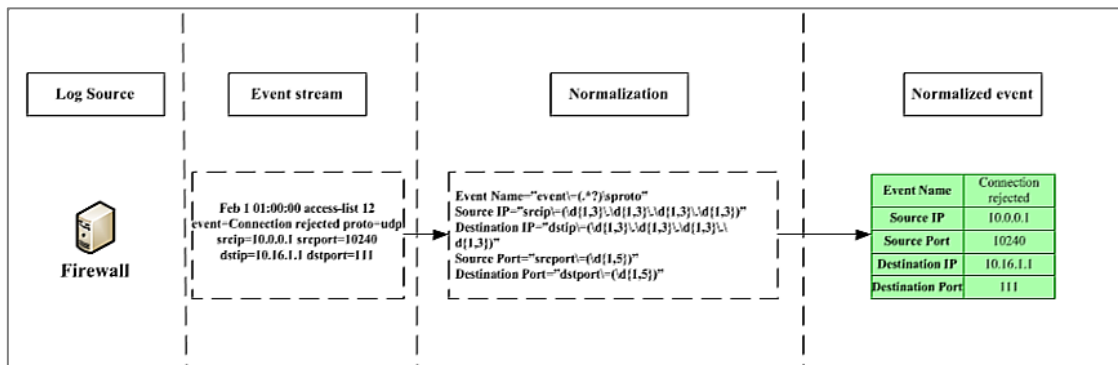


Figure 11: Event normalization example

II.4.3.3 Aggregation

Data aggregation is a handy feature that reduces the amount of redundant information collected in logs. It means that several similar events may be reduced to one based on conditions; these conditions are simple rules based on normalized events.

- **Aggregation types**

1. Simple aggregation: When two events are identical, they are aggregated.
2. Field-based aggregation: When selected event fields are similar in defined threshold and time interval.

The figure below shows an example of an event ‘Connection rejection’, but this time two servers access requests from the same IP address within 10 seconds detected by the firewall would generally create two lines in a log; event aggregation will make it only one line:

1. Event stream of raw events coming for normalization and categorization.
2. Normalization parsers are applied to raw events.
3. Normalized event processed by aggregation engine.

4. Aggregated event coming to next step.

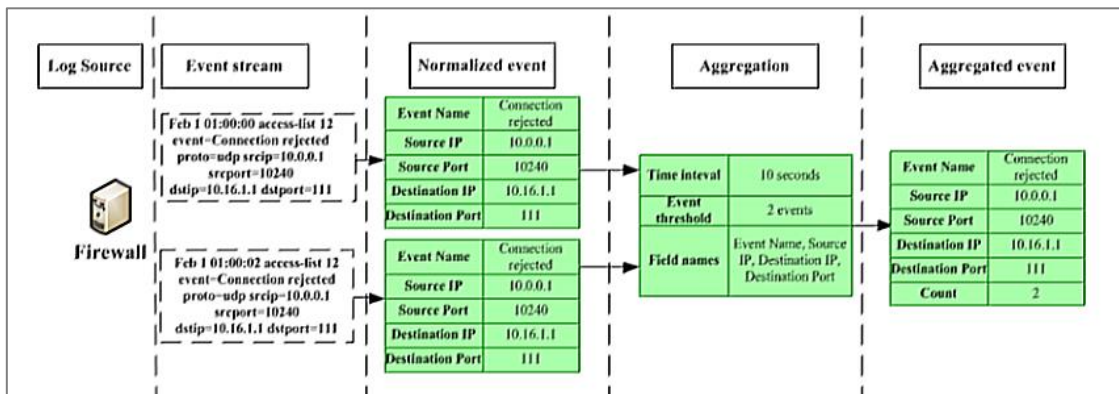


Figure 12: Event aggregation example

II.4.3.4 Correlation

The correlation of formal events to a correlated event is performed to simplify incident response procedures in real-time on incoming normalized log data taken from log sources and then analyzes the data to identify relationships by displaying a single event triggered by multiple events from various sources.

The correlation types are resumed into two categories (Event Correlation | Vladimir Potapov Blog):

- Incident detection – contains detection conditions;
- Auxiliary – routine with lists for incident detection rules, reports, and dashboards.

Using event correlation tools can perform actions based on user-defined rules, such as sending alerts for hardware or application failures.

II.4.3.5 Alerting and automatic response

The SIEM solution must also be able to respond to a given event or set of circumstances automatically. For example, responses may include alerting by sending an email or a network modification directive, such as adjusting a firewall or setting up a switch.

SIEM platforms associate alerts with rules, even integrating alert definitions into the policy management system. This way, when the rules are created, their criticality and the appropriate response are defined at the same time.

II.4.3.6 Archiving

SIEM solutions are also used for legal and regulatory reasons. Probative archiving ensures the integrity of the traces. Solutions can use RAID disks, calculate the footprint, and use encryption or other means to ensure the integrity of the traces.

II.4.4 Benefits of SIEM

The implementation of a security control within an organization is necessary for the IT equipment protection; SIEM system benefits are resuming in (The SIEM Advantage):

- **Data aggregation and visibility:** Logs are normalization and correlation in a SIEM system, providing good IT visibility. A massive amount of data must be supervised; that is why the SIEM capabilities related to data aggregation and normalization are so beneficial. The tool also analyzes and correlates this data, finding connections that can quickly help IT staff detect security incidents.
- **Streamline compliance reporting:** SIEM server receives log data from many hosts and can generate one report that addresses all of the relevant logged security events among these hosts.
- **Threat Detection:** SIEMs have a assortment of features and functionality that incorporates security monitoring: the entire collection, normalization, correlation, and analysis of logs. SIEM also alerts security analysts when suspicious events are detected.
- **Increased efficiency:** Due to the ability of SIEM to collect event logs from multiple devices, IT staff can quickly identify potential issues, as well; as a quick check on activities and analysis files; SIEM systems can also improve reporting processes across the business.
- **Better overall handling of security breaches and events:** SIEM software can reduce the impact of a security breach by providing a fast response to the detected events, such as the financial cost of a breach and the damage caused to the business and any IT systems in place.

II.4.5 SIEM solutions type

In the category of SIEM (Security Information and Event Management), we can roughly distinguish three types of solutions:

- The “blank” products of any original configuration for which everything must be put in place. This is the case of Splunk, which relies on adaptability and flexibility to ensure an effective solution whatever the need.
- Solutions based on existing and scalable bases to ensure correlation, alerting, reporting, and offer parsers adapted to the majority of the equipment on the market.
- Ready-to-use solutions that are easy to use and spontaneously meet a specific need. These, if they can give satisfaction at the moment, generally fish in terms of scalability and functionality.

SIEM solution aims to respond to the companies’ real-time need to analyze security events concerning internal and external threat management. This solution allows monitoring applications, user behaviors, and data access. Through the functionalities provided by the solution, it is, therefore, possible to collect, standardize, aggregate, correlate and analyze data from events from machines, systems, and applications (firewall, IDS/ISP, Network Machines, Security Machines, Applications, Databases, Servers, Directories).






II.4.5.1 Commercialized solutions

The SIEM solutions used by the large companies are generally commercialized and built according to their needs; however, it costs fortunes, making it impractical for some companies due to financial issues. Enlisted below are the trending Security Information and Event Management commercialized tools available in the market.

All the solutions are compatible with the width of the business, small, medium, and large; however, not all the solutions are adapted to the operating systems; according to the table, ArcSight is adaptable only with Windows OS. Moreover, SIEM solutions deployment varies from one system to another, like SolarWinds is used on-premise and cloud. Finally, not all the solutions offer a free trial, but the price varies depending on the functionalities offered. We will not go into more detail and make comparisons because the purpose of our

project is not to make a demonstration about the commercialized solutions or define any of these.

Table 3: Most popular SIEM tools 2022

SIEM	Best for	OS Platform	Deployment	Free Trial	Price
 SolarWinds	Small, Medium, and Large businesses.	Windows, Linux, Mac, Solaris.	On-premise & Cloud	30 days	Starts at \$4665.
 Datadog	Small, Medium, & Large businesses.	Windows, Mac, Linux, Debian, Ubuntu, CentOS, RedHat.	On-premise and SaaS.	Available	Security Monitoring price starts at \$0.20 per GB of analyzed logs per month.
 Splunk	Small, Medium, and Large businesses.	Windows, Linux, Mac, Solaris.	On-premises & SaaS	Splunk Enterprise: 60 days Splunk Cloud: 15 days Splunk Light: 30 days Splunk Free: Free sample for core enterprise platform.	Get a quote.
 McAfee ESM	Small, Medium, and Large businesses.	Windows & Mac.	On-premises, Cloud, or Hybrid	Available	Get a quote.
 ArcSight	Small, Medium, and Large businesses.	Windows.	Appliance, Software, Cloud (AWS & Azure)	Available	Based on data ingested and security events correlated per second.

II.4.5.2 Open-source solution

Several open-source SIEM solutions have become robust and competitive. As a result, there are many benefits to using an open-source SIEM solution, and not just for small businesses. Unfortunately, existing solutions either lack core SIEM capabilities, such as event correlation and reporting or require combining with other tools. As always, though, there are some excellent contenders. (Vazão et al., 2019). We define the four most famous open-source SIEMs solutions:

- **Elasticsearch**



Figure 13: Elasticsearch Logo

Elasticsearch, formerly ELK Search, is a distributed open-source data search and analysis engine based on Apache Lucene and developed in Java. The project began as an extensible version of the Lucene open-source search framework. It is a comprehensive ecosystem of open-source tools for data ingestion, enrichment, storage, analysis, and visualization (ELK stands for Elasticsearch, Logstash, and Kibana components.) Logstash and Beats provide the logs. Beats bring together senders and data collectors, while Logstash filters this data and activates many custom plug-ins. Elasticsearch is the engine that feeds data mining, and Kibana provides visualization. (Stoleriu et al., 2021).

Elasticsearch is a powerful and versatile suite, but it lacks some essential features; with some additional important features, Elasticsearch would be a complete SIEM tool. In particular, it is insufficient in correlation and does not provide ready-to-use alerts or autonomous incident management. However, with its robust architecture, customization, and open-source nature, Elasticsearch is unsurprisingly very powerful and provides the basics for many other choices in this list. The following table demonstrates the pros and cons of Elasticsearch:

Table 4: Pros and cons of Elasticsearch

Pros of Elasticsearch	Cons of Elasticsearch
<ul style="list-style-type: none"> - Compatible to run on every platform because it is developed in Java. - In a real-time search engine, any added document is searchable in this engine. - It offers the concept of a gateway, allowing full backups to be created quickly. - A distributed document-oriented that makes it easy to scale up in a large organization. - It supports all document types except those that do not help text rendering. - Offers the most effective full-text search property. It performs searches based on language and returns those documents that match the search condition. - Generates high productivity with parallel processing by allocating primary and replica shards across all available nodes. 	<ul style="list-style-type: none"> - Sometimes, the problem of split-brain situations occurs in Elasticsearch. - Does not have multi-language support for handling request and response data. Supports only a JSON format. - Not a suitable data store as other options such as MongoDB, Hadoop, etc. It performs well for small use cases. - It is a flexible and powerful data storage search engine but challenging to learn. Especially in terms of enterprise search usage, it is not as uncomplicated as the box search.

- **OSSIM:**



Figure 18: AlienVault OSSIM logo

OSSIM is one of the most powerful open source options developed by AlienVault within infrastructure with complete security supervision. The framework in the sense of OSSIM aims to centralize, organize and improve the detection and display for monitoring system security events information from a company. OSSIM performs event collection and normalization. It has short-term logging and monitoring capabilities, long-term threat assessment, and built-in automated responses.(Phillipe Martinet, 2006)

However, OSSIM is neither flexible nor manageable. System administrators complain about the cumbersome configurations, especially on Windows, and the time-consuming investments needed to customize the software.

The following table demonstrates the pros and cons of OSSIM:

Table 5: Pros and cons of OSSIM

Pros of OSSIM	Cons of OSSIM
<ul style="list-style-type: none"> - It can be operated on-premise and virtually. - Requires only a single server - There is community support via its product forum. - Developers provide ongoing development increasing its value to users. - Learning function that allows the solution to increase the reliability of its feedback. - Intuitive interface because of the modularity of the control panel that adapts to the customer's needs. 	<ul style="list-style-type: none"> - Limited flexibility makes its customization a long process. - Implementing a complex solution requires a relevant audit and risk assessment process in configuring the desired security policy.

- **Wazuh :**



Figure 14: Wazuh Logo

Wazuh is an open-source based on OSSEC (open source and free Host Intrusion Detection System (HIDS)) and a common choice among organizations because of its capabilities in threat detection, integrity monitoring, and compliance as an incident management tool. Wazuh collects, aggregates, indexes, and analyzes security data, enabling organizations to detect intrusions and identify threats and behavioral anomalies. The following table demonstrates the pros and cons of Wazuh (Wazuh · The Open Source Security Platform):

Table 6: Pros and cons of Wazuh

Pros of Wazuh	Cons of Wazuh
<ul style="list-style-type: none"> - Based on (and compatible with) OSSEC. - Supports the cloud infrastructure monitoring, including AWS (Amazon Web services) and Microsoft Azure (Cloud computing services). - Integrates with Splunk to visualize alerts and API data. 	<ul style="list-style-type: none"> - Complicated architecture: requires a full Elastic Stack deployment in addition to the Wazuh server components.

- **Quadrant Sagan:**



Figure 15: Quadrant Sagan Logo

Quadrant Sagan is an open-source real time log analysis and correlation engine with high performance; it runs under *nix operating systems, written in C, and uses a multi-threaded architecture to deliver high-performance log and event analysis.

Sagan is compatible with Snort, it detects the threat, and Snort is used to prevent threat damage. Sagan was designed to be lightweight, so it is also beneficial for companies that do not require several features and focus only on their application's performance. It is

perfect for businesses that use Snort or plan to implement an IPS in addition to a SIEM. (Sagan User Guide Documentation Release 1.2.2 Champ Clark III, 2022). The following table demonstrates the pros and cons of Quadrant Sagan:

Table 7: Pros and cons of Quadrant Sagan

Pros of Sagan	Cons of Sagan
<ul style="list-style-type: none"> - Fully compatible with Snort databases, rules, and user interfaces. - Multi-threaded architecture is designed for high performance. 	<ul style="list-style-type: none"> - Relatively young project with a small community. - The difficult installation process can involve building the entire SIEM from the source.

II.5 Solution proposed

The previous comparison shows ELK stack and Wazuh are more responsive to our needs. The next chapter contains a detailed study of the two tools and explains the reasons behind choosing these tools as a SIEM solution.

Conclusion

Security of information systems and monitoring of events are essential requirements in all the IT infrastructure, especially against suspicious acts supervision varies from one need to another. Likewise, supervision tools vary from one need to another.

In this chapter, we define the basics of IT supervision and security monitoring, including the use of machine learning in the security field, passing by knowing the utility of log files and giving examples of Syslog, also a definition of security information and event management tools. And the integration of machine learning in anomaly detection, we also mention the SIEM solutions and their functioning. Finally, from the previous comparison, we conclude that ELK stack and Wazuh are more responsive to our needs.

The next chapter contains a detailed study of the two tools and explains the reasons behind choosing these tools as a SIEM solution.

Chapter III



Design and implementation

Introduction

Based on the comparison we made on open source tools, we deduced that for our solution, it is appropriate to use ELK Stack and Wazuh.

In this chapter, we focus on the conceptual aspect of our project; then, we give an observation of the proposed solution via its architecture. Also, present the different interfaces illustrating the most interesting options.

III.1 Conceptual study

III.1.1 Context definition

To describe the content of this solution, we need to know further how this solution will serve the bank's needs. Thus in our discussion with our supervisor Mr. LADJOUZI, we concluded the following clarifications. First, information system analysts must be able to:

- Detect and stop attempts to compromise user identification information (Brute force, Pass the hash, etc.). In a successful compromise, it is essential to identify the affected entities to investigate the impact and prevent further damage.
- Detect Privilege Escalations: Once an opponent receives a higher authorization level, the risk of damage and information leakage increases.
- Data Loss Protection: It is essential to protect all sensitive information within the organization and prevent users from sharing this data outside the infrastructure. We need to monitor all endpoints to detect an abnormal data output volume (also on the channels) and set alerts on any anomaly based on past behavior. We must not forget the critical endpoints, the accounts on the list of monitoring, or users who have recently been or are about to be licensed.
- Detect and control system changes: appropriate rules must be created and applied to report critical events, such as unauthorized changes to configuration files, removal of audit trails, or falsification of audit logs.

- Ensure high availability of services by reacting against back attacks (denial of service) by using network traffic logs to issue alerts in case of malicious traffic peaks or deviations from the baseline of regular traffic, such as an abnormal number of requests from multiple ports or the same IP address.
- Detects malware entering the network.
- Monitor server and device performance (CPU usage, temperature, RAM, disk, etc.).

Information system managers must be able to have a global overview of all events in the form of a dashboard.

We can deduce the actors who will use our solution from the previous points.

III.1.2 Actors

- IT Security Analyst: - Type: Human
- Role: Secure the information system and ensure the high availability of services.
- Assigned to a team of 3 people
- IT Security Manager: - Type: human.
- Role: To have a global view of the state of the IT security system under dashboard shape.
- Assigned to Information Security System manager of BADR Bank

III.1.3 Context diagram

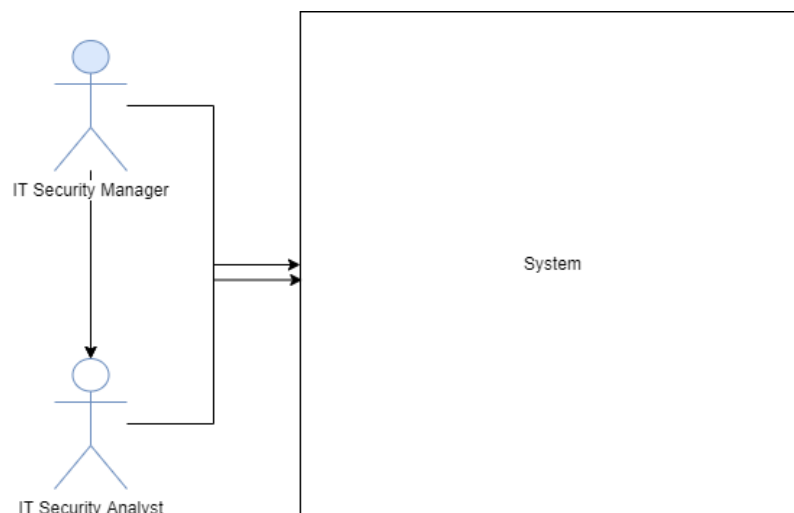


Figure 16: Context diagram

III.1.4 Modeling

- **Package diagram :**

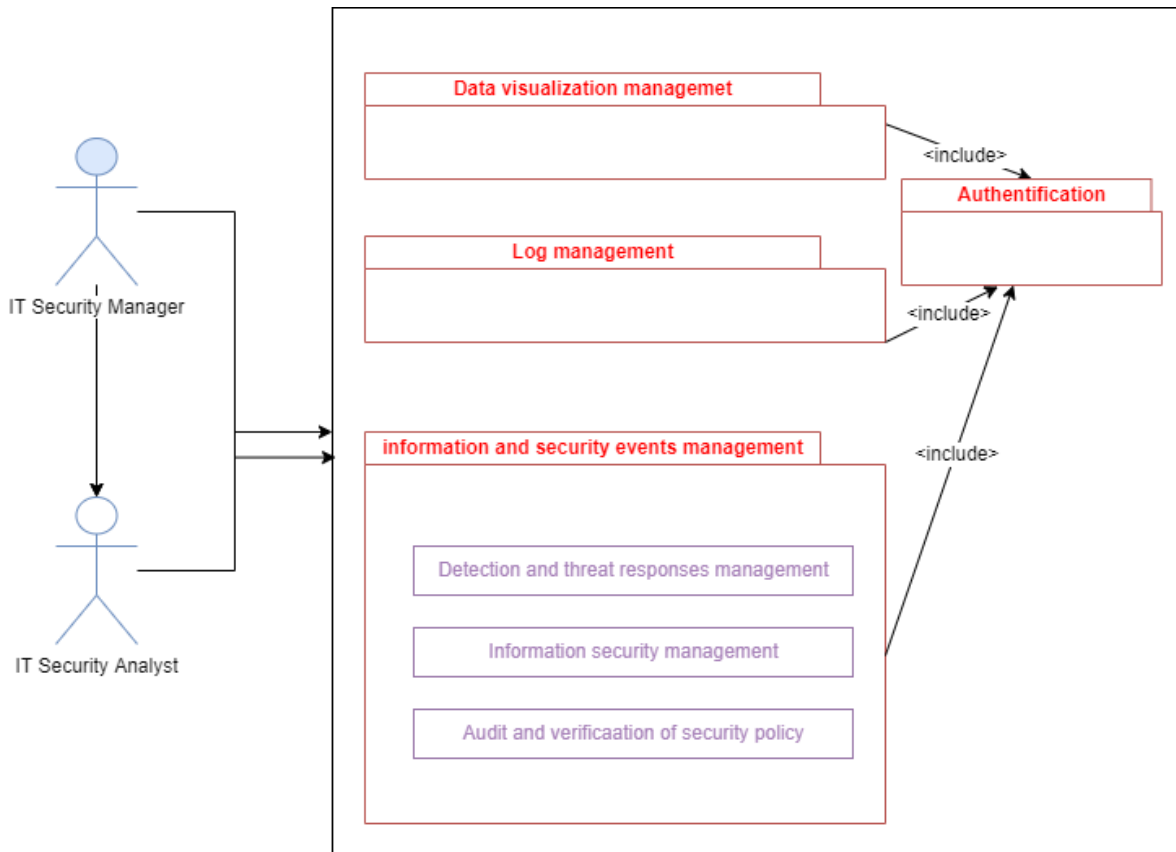


Figure 17: Package diagram

Figure 17 shows the package diagram, and each package contains cases that will be explained below. The Information Management and Security Events themselves consist of three packages.

- **Use case diagrams:**

Thus, we have identified the actors and the significant families of functionalities (packages). Now, the task is to define the needs of each actor in these packages in more detail.

- A use case diagram for the log management package:

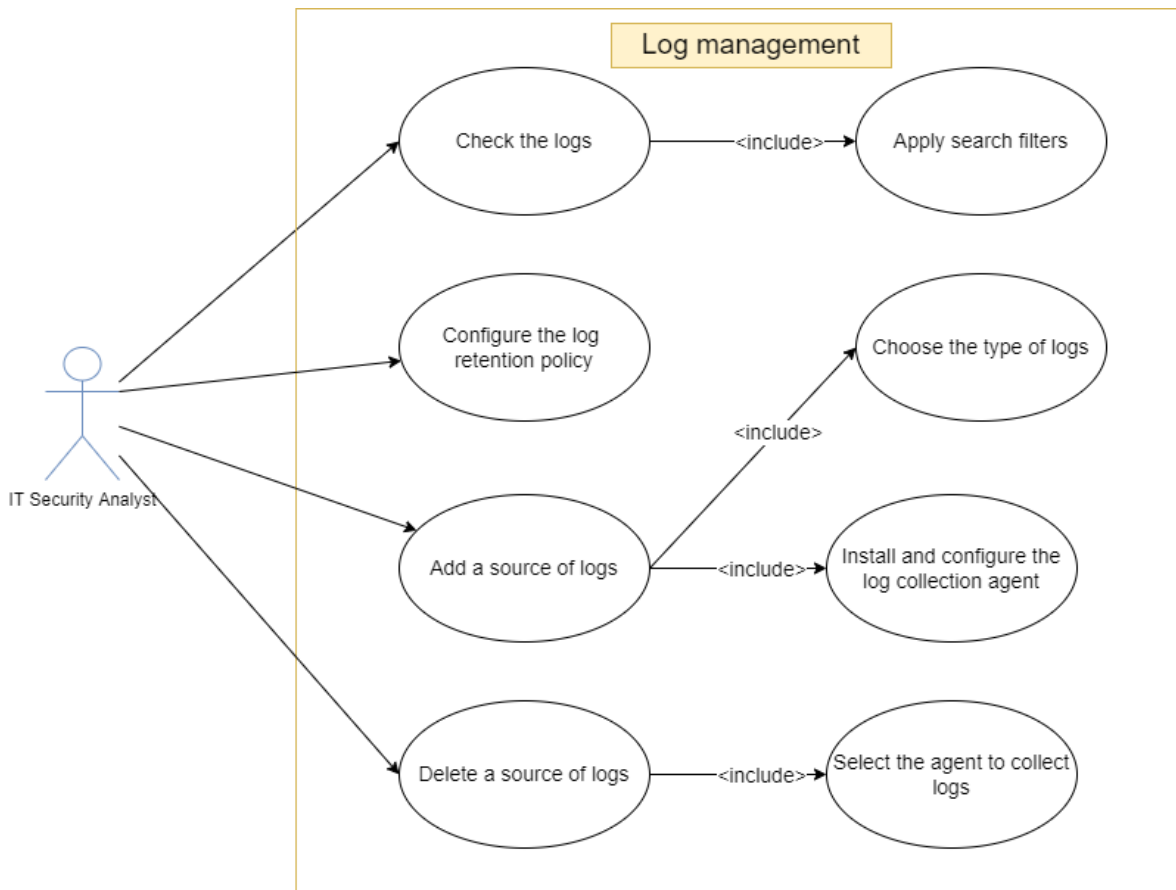


Figure 18: Log management use case diagram

Description of use cases:

IDENTIFICATION
<p>Case Number: 1</p> <p>Name: Consult the logs</p> <p>Actor(s): Analyst</p> <p>Description: Analysts need access to the logs.</p> <p style="text-align: center;">Prerequisites: The user is required to be authenticated as an analyst.</p>
<p>The nominal scenario (It is here to describe the ideal progress of actions, where everything goes for the best.)</p> <ol style="list-style-type: none"> 1. The system displays a page that includes filtering fields to retrieve logs (e.g., logs of the last 15 minutes from the source whose IP address is 192.168.1.177 and the index is: logs. Apache). 2. The system collects and displays the logs corresponding to a default filter. <p>The alternative scenarios (here, it is a matter of describing the possible different steps related to the user's choices; for example. This is the case for actions related to</p>

«extended» conditions.)

3. The analyst can change the search filter in (1.) and click search; in this case:

III.1. The system retrieves logs associated with the filter and displays them instead of logs by default.

IDENTIFICATION

Case Number: 2

Name: Add a log source

Actor(s): Analyst

Description: Adding a log source must be possible for analysts

Preconditions: The user must be authenticated as an analyst

The nominal scenario:

1. The system displays a page that contains all possible log sources with their logo (Windows, Cisco, Apache, ...).

2. The analyst makes a selection.

3. The analyst installs the collection agent on the source device.

4. The analyst configures the agent from the source device:

- Specification of the destination address (collection server)

- Activation of the module associated with the type of logs chosen in (1-) and that the analyst wants to recover (e.g., apache log recovery module, Windows application log recovery module, Cisco security log recovery module, ...).

- Provide the authentication (key) information needed to authenticate the agent on the server.

5. The system authenticates the agent, accepts the received logs, and associates them with the type chosen in (1-), so the system uses the associated normalization class to standardize the logs.

IDENTIFICATION

Case Number: 3

Name: Configure log retention time

Actor(s): Analyst

Description: Programming the retention time of logs must be possible for analysts.

Preconditions: The user must be authenticated as an analyst
<p>The nominal scenario</p> <ol style="list-style-type: none"> 1. The system displays a page containing the current retention policy configuration. 2. The user makes the necessary changes. 3. The user saves the changes and restarts the services for the changes to be applied.

IDENTIFICATION
<p>Case Number: 4</p> <p>Name: delete a source of logs</p> <p>Actor(s): Analyst</p> <p>Description: Deactivation of a log source must be possible for analysts</p> <p style="padding-left: 40px;">Preconditions: agent installed in the source</p>
<p>The nominal scenario:</p> <ol style="list-style-type: none"> 1. The user changes the agent configuration from remote configuration files (SSH, Telnet) or by physical presence. 2. The user can delete the agent from the source, and the logs will no longer be received from the source.

A use case diagram for the data visualization package:

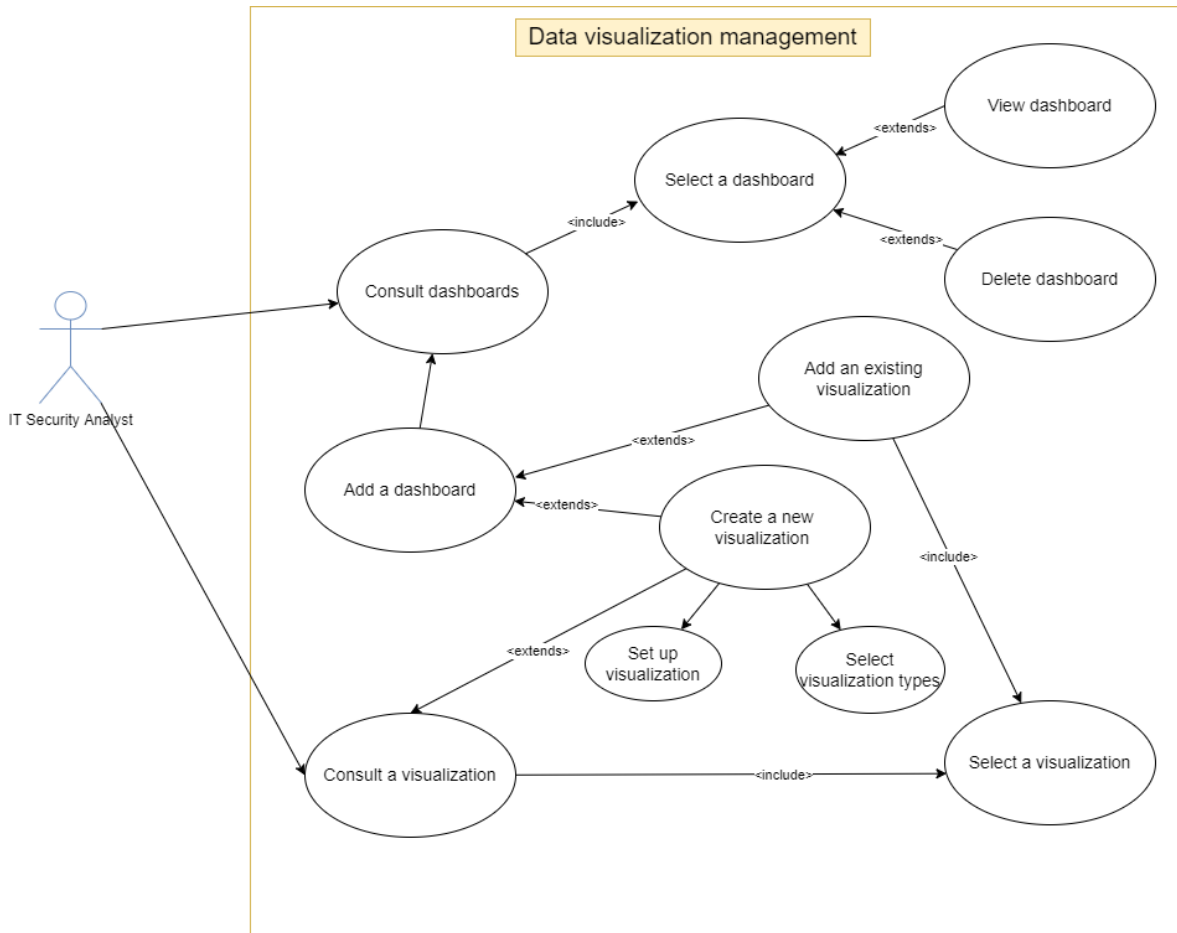


Figure 19: Data visualization use case diagram

Description of use cases:

IDENTIFICATION
<p>Case Number: 5</p> <p>Name: View Dashboards</p> <p>Actor(s): Analyst, Manager</p> <p>Description: Visualization of dashboards is essential for analysts and manager</p> <p>Preconditions: The user must be authenticated as an analyst or as a manager</p>
<p>The nominal scenario</p> <ol style="list-style-type: none"> 1. The system displays a list of all the dashboards present. 2. User selects a dashboard. 3. The system displays the dashboard (which contains one or more views). <p>The alternative scenarios:</p> <ol style="list-style-type: none"> 4. In step (1.), the user can choose to add (create) a dashboard in this case:

- 4.1. The system displays a blank page with the added existing viewing options and creates a new viewing.
- 4.2. If the user clicks on add existing review, then:
 - 4.2.1. The system displays the list of views.
 - 4.2.2. The user selects one or more visualizations and confirms the action.
 - 4.2.3. The system adds the visualization(s) to the dashboard
 - 4.2.4. This operation (4.) can be repeated in a loop.
- 4.3 If the user clicks on create a new review, then:
 - 4.3.1. The system displays a list of possible viewing types.
 - 4.3.2. User selects a view type (example: horizontal bar)
 - 4.3.3. The user selects the indicator and configures the graph (ex X-axes: data1, Y-axes: data2)
 - 4.3.4. The user gives a name to identify this visualization and saves the visualization.
 - 4.3.5. The system records the review.
- 4.4. User has completed dashboard creation, names the dashboard, and saves the changes.
5. In step (1.), the user can select a dashboard and click on the button (delete dashboard); in this case, the dashboard will be deleted after confirmation of the action.
6. In step (3.), the user can delete a dashboard view, add a new view, change the configuration of a view, and save the changes.

IDENTIFICATION

Case Number: 6

Name: view visualizations

Actor(s): Analyst, Manager

Description: Creating custom dashboards requires creating a graph (visualization object) and is essential for analysts and the manager to create custom dashboards

Preconditions: The user must be authenticated as an analyst or as a manager.

The nominal scenario:

1. The system displays a list of all graphs (visualization) present.
 2. User selects a view.
 3. The system displays the review (graph).
- The alternative scenarios:
4. In step (1.), the user can choose to add (create) a view in this case:
 - 4.1. The system displays a list of possible viewing types.
 - 4.2. User selects a view type (example: horizontal bar)
 - 4.3. The user selects the indicator and configures the graph (ex X-axes: data 1, Y-axes: data 2)
 - 4.4. The user gives a name to identify this visualization and saves the visualization.
 - 4.5. The system records the review.
 5. In step (1.), the user can choose a view and click on the button (delete view); in this case, the view will be deleted after confirmation of the action. In this case, this view will be removed from all the dashboards it was added to.

- A use case diagram for the information security package

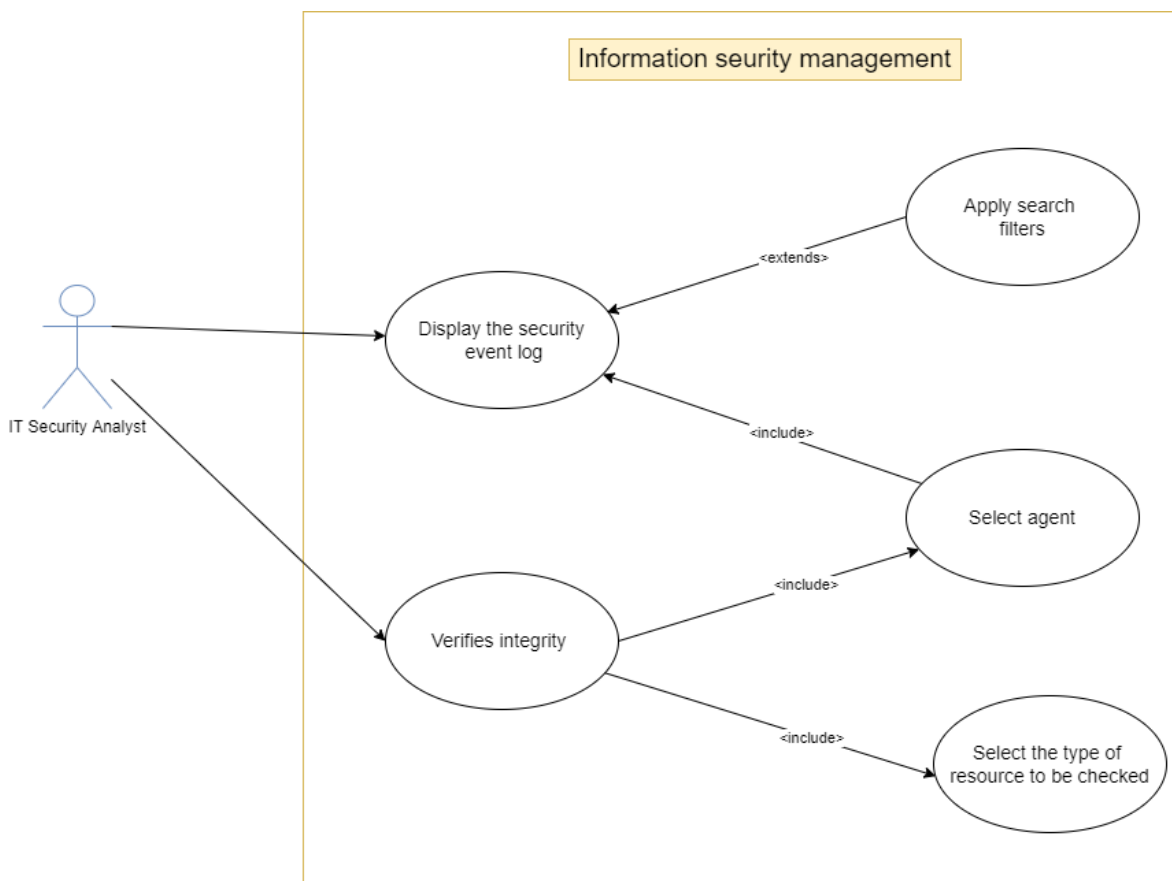


Figure 20: Information security use diagram

Description of use cases:

IDENTIFICATION

Case Number: 7

Name: Display the security event log.

Actor(s): Analyst

Description: Security event logs must be available for the analysts

Preconditions: The user must be authenticated as an analyst

The nominal scenario:

1. The system displays a list of all agents.
2. The user selects an agent.
3. The system displays the security logs for the selected agent.

The alternative scenarios:

3. The analyst can change the search filter in (1.) and click search; in this case:

III.1. The system collects the security logs associated with the filter and displays them instead of the default logs.

IDENTIFICATION

Case Number: 8

Name: Verifies integrity

Actor(s): Analyst

Description: Integrity verification of an agent must be possible for analysts

Preconditions: agent installed in the source

The nominal scenario:

1. The analyst must select the agent he wants to verify the integrity of these resources from the list of agents.
2. The analyst selects the type of resource checked (examples: files, registry key, etc.).
3. The system spear a check and displays whether the integrity of this resource in the selected agent has been compromised or not.

- A use case diagram for the threat detection and response package:

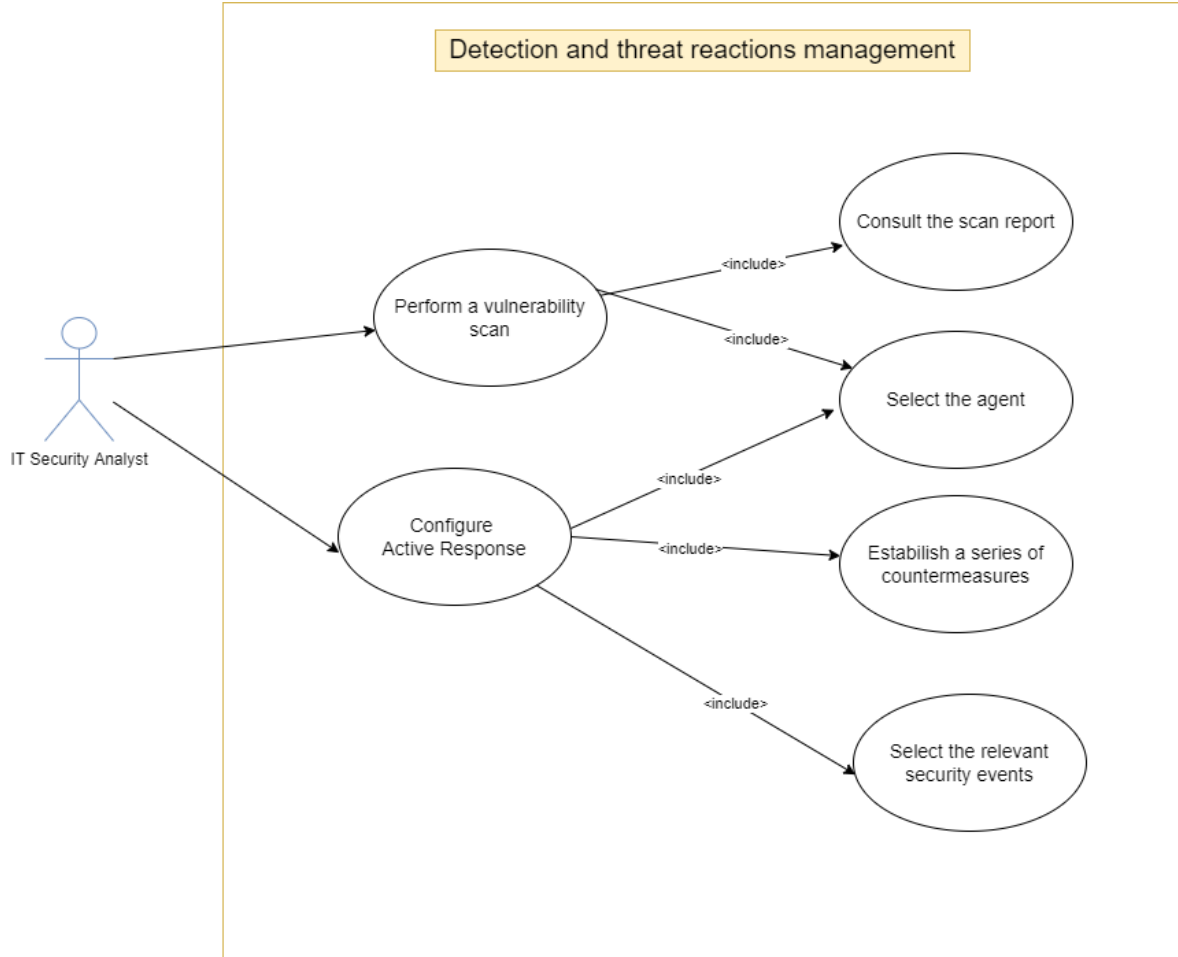


Figure 21: Detection and response to threats use case diagram

IDENTIFICATION

Case Number: 9

Name: Perform a vulnerability scan

Actor(s): Analyst

Description: An agent's vulnerability scan must be possible for analysts

Preconditions: The user must be authenticated as an analyst

The nominal scenario

1. The analyst must select the suspicious agent to perform a vulnerability scan.
2. The system initiates the scan on the affected agent.
3. A report containing the scan result will be generated at the end of the scan, and the user can consult.

IDENTIFICATION

Case Number: 10

Name: Configure Active Response

Actor(s): Analyst

Description: Configuration of an active agent response must be possible for analysts

Preconditions: The user must be authenticated as an analyst

The nominal scenario:

1. The analyst must select the agent.
2. The analyst selects the event concerned by the configuration of the active response.
3. The analyst must specify the actions to be established (in the form of a series of countermeasures) as soon as the event in question appears.

- A use case diagram for the threat detection and response package:

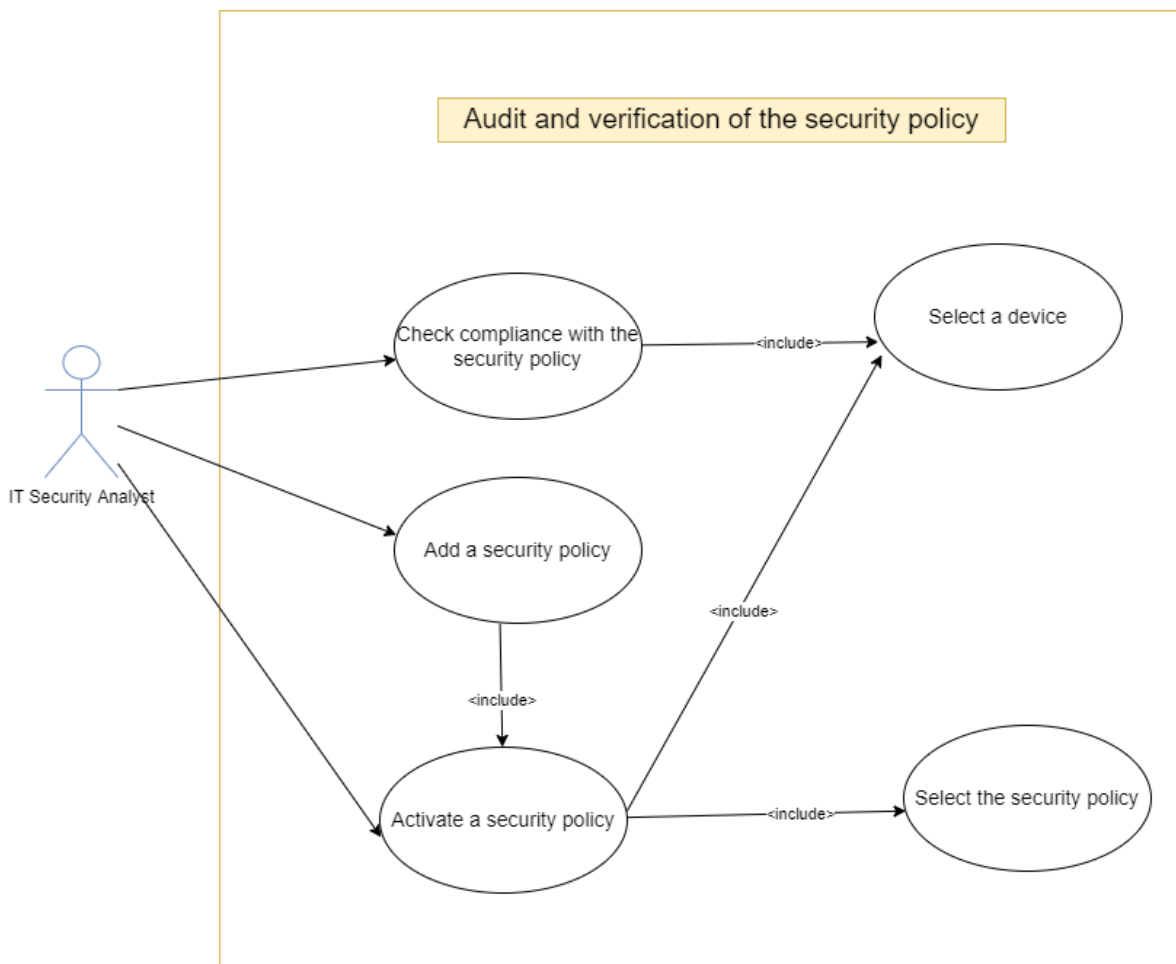


Figure 22: Threat detection and response use case diagram

Description of use cases:

IDENTIFICATION

Case Number: 11

Name: Verify compliance with security policy

Actor(s): Analyst

Description: The verification of compliance with the security policy of a device must be possible for analysts

Preconditions: The user must be authenticated as an analyst

The nominal scenario:

1. The analyst selects a device.
2. The analyst spears the security policy check option on the device.
3. The system performs checks.
4. The analyst will be informed if the security policy is being followed on this device or not.

IDENTIFICATION

Case Number: 12

Name: Add Security Policy

Actor(s): Analyst

Description: The addition of a security policy must be possible for analysts

Preconditions: The user must be authenticated as an analyst

The nominal scenario:

1. The analyst writes the new security policy.
2. The new security policy must respect the model recognized by the system.
3. The system checks whether or not the new security policy exists.
4. The analyst must select a device where they want to activate this policy.

IDENTIFICATION

Case Number: 13

Name: Activate a security policy

Actor(s): Analyst

Description: Activation of a security policy on a device must be possible for analysts

Preconditions: The user must be authenticated as an analyst

The nominal scenario:

1. The analyst selects a device.
2. The analyst selects the security policy to be activated (the selected policy must exist on the list of security policies).
3. Activate the specified agent policy.

III.1.5 Conceptual solution

Developing a real SIEM solution that meets the expectations of a bank like BADR requires a lot of effort and technical knowledge. Therefore, it is a task assigned to Cyber and Information Security teams, and the project can last from one to two years.

Figure 8 is the block diagram representing our system's overall architecture (modules and interaction between modules).

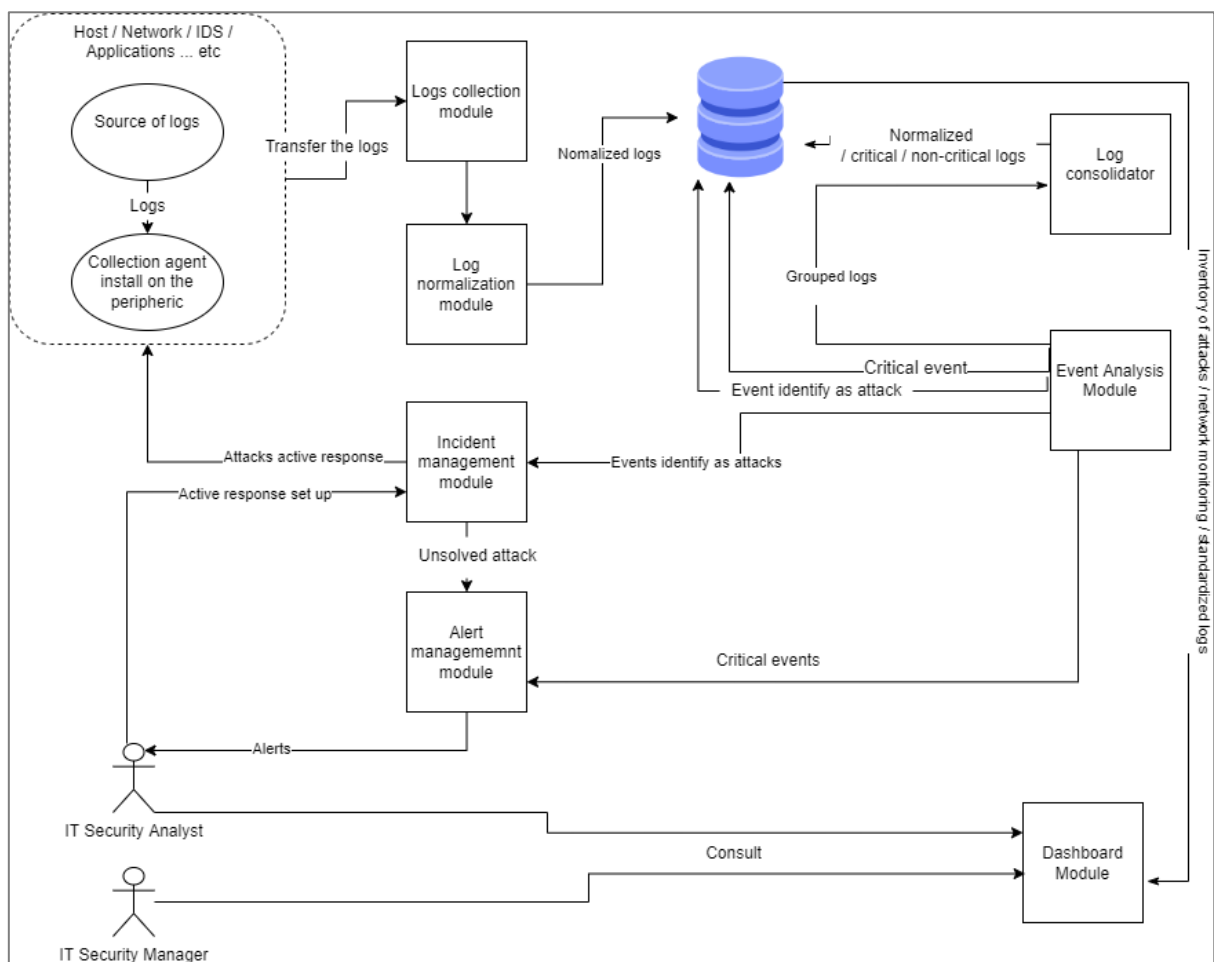


Figure 23: Block diagram

The block diagram shows that the system must contain at least seven modules.

The modules: log collection module, log normalization module, log deconsolidation module, event analysis module, incident management module, alert management module, and dashboard module.

It also has a database, which contains the agents' information (secret key, information about the system where the agent is installed, the status of the agent (active – blocked, etc.), the standardized logs of each device, inventory of attacks, etc. In addition, the system accepts logs from different devices, standardizes logs in a standard format, groups them according to consolidation schemes, performs a classification (normal attack, critical or non-critical event), and generates reports and recommendations.

Below are the flowcharts that explain how each module works.

- **Log collection module**

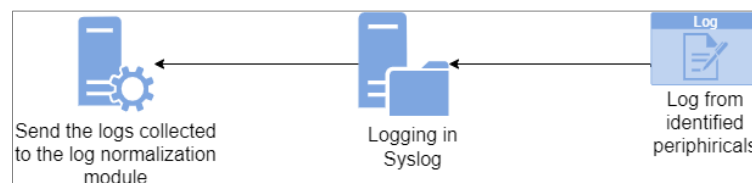


Figure 24: Log collection module

As shown in the figure, logs from the identified devices are collected. Each field is then filled in automatically in their respective domains. Then, the logs are imported into the Normalization module. The collection module uses Syslog Listener as its implementation. The (RFC) 3164 is an IETF memo stating that logs must be standardized to facilitate collection in the network.

As a rule, logs are in the following format based on Syslog software. It triggers the parameters Date/ Time, IP address, installation (kernel, user-level, mail system, Daemon system, etc.), severity (emergency, alert, critical, error, warning, notification, etc.), and the Message. (More information in the Syslog section).

Different devices can fill all Syslog fields while others do not. However, this module does not recognize devices not listed in the identified device database. (Back to theoretical background).

- **Log normalization module:**

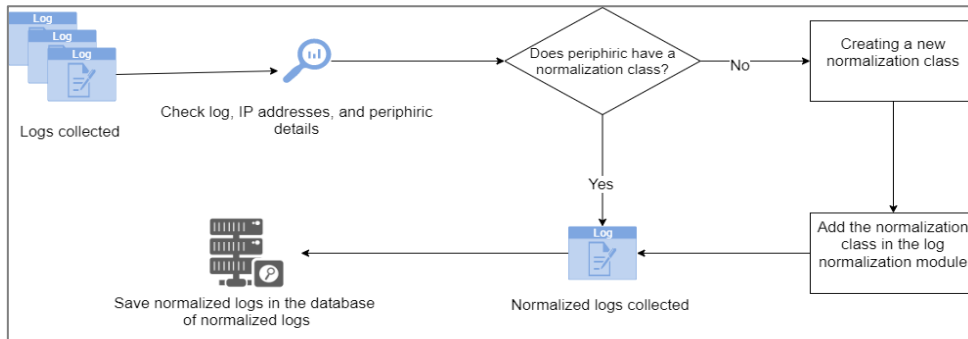


Figure 25: Log normalization module

The collected logs are sent to the log normalization module, which checks the logs and searches for the IP address and device details.

If the device has a normalization class, the logs are immediately normalized and stored in the database of the normalized logs. A new class must be added if the device does not have a normalization class. Normalization includes content fields, removing unnecessary fields in device-generated logs, and translating fields to similar formats.

The standardization module also contributes to the consolidation of logs. The devices are already classified from the device identification module; when the logs of these devices are passed to the normalizer, it is possible to predict their module's destination. For example, suppose the normalization module detects that the log is from an IDS. In that case, there is an excellent possibility that this is an attack; therefore, this log should be an intended Incident Management Module. All normalized logs are placed in the normalized log database before being forwarded to the log consolidator.

- **Log consolidation module:**

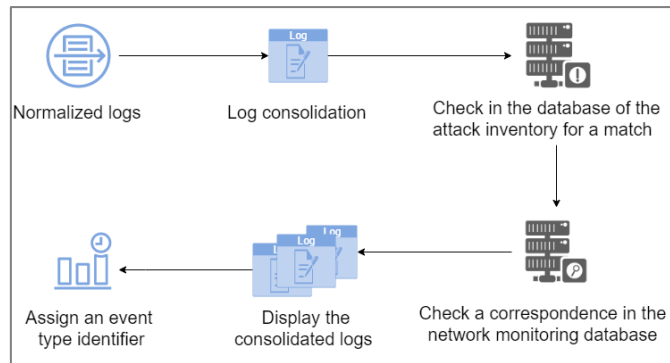


Figure 26: Log consolidation module

The consolidation module must be implemented using correlation techniques, and there are three particularly effective correlation techniques already used by several SIEM like OSSIM Alien Vault:

- *Cross-Correlation*: Compares information from IDSs and vulnerability scanners and prioritizes events in case the data is vulnerable or not to a particular attack. For example, suppose attack A exploits only vulnerability X by correlation. In that case, the vulnerability scanner scans the network and reports network vulnerabilities. Suppose the vulnerability scanner reports that vulnerability A is on the network. The IDS detects that attack A is occurring through cross-correlation. In that case, this event has a higher priority than the various events that do not interfere with the network.
- *Inventory Correlation*: Checks if the attack affects a particular service and operating system and a particular system version and checks if the host attacked a system operating/active service. The event is eliminated if the conditions are not together. For example, attack B only runs a specific operating system/software installed on the terminals. If attack B occurs and the PC 1 is known for OS/software it exploits, the attack concerning PC1 is a priority to events that do not affect the network or a specific terminal.
- *Logical Correlation*: Refers to the logical rules used to join different small events to fit a new model.

The consolidation module also uses two databases (attack inventory and network surveillance (monitoring)).

- **Event Analysis Module:**

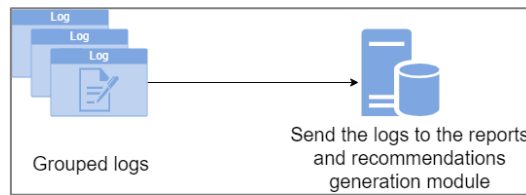


Figure 27: Event Analysis Module

Once logs are grouped and classified as events, problem logs must be resolved and sent to their respective modules. For example, the Event Analysis module passes all events to the Report/Recommendation Generator module, and those regardless of the (id) of event.

- **Incident Management Module:**

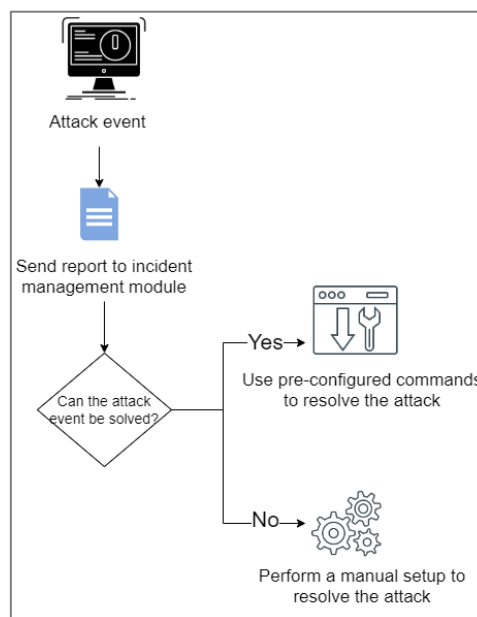


Figure 28: Incident Management Module

Attack events originate primarily from an IDS/IPS. Unresolved attacks are first brought to the Alert Management module, which will warn the user by sending alerts.

To implement this module, we must use two techniques (sniping and shunning). Event sniping, or session sniping, is a direct intervention to disrupt the victim's connection. The action is done by injecting forged packets to reset the link (bit RST in the TCP protocol). The port, source IP, and sequence numbers must be synchronized with the traffic that triggered the event for the reset.

Shunning is the denial of access to a host suspected of an attack. In the implementation, one solution is to stop an attacker's IP access to reduce the possibility of extending the attack to other targets in the protected environment.

The attacks that cannot be resolved automatically are determined manually by the user.

- **Alert Management Module:**

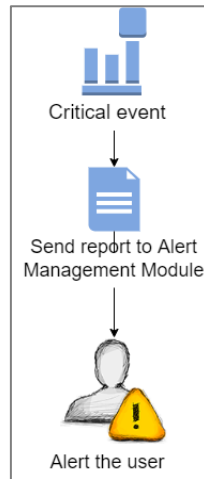


Figure 29: Alert Management Module

Critical events are sent to the alert management module that will handle to alert the analyst.

- **Dashboard Module :**

This module sends queries to the database, retrieves the information needed to calculate the indicators, and displays the result as a graph.

At this stage of our study, we began to have a clearer idea about our objective and how to achieve it and especially about the degree of difficulty of the project, knowing that we must respect the duration of the internship and especially that we do not have the adequate equipment. For example, in the normalization module, we have to study each possible log source and create a normalization class; here is the class diagram of the log normalization module; it is necessary to consider the possibility of extension in case there is a new source of logs:

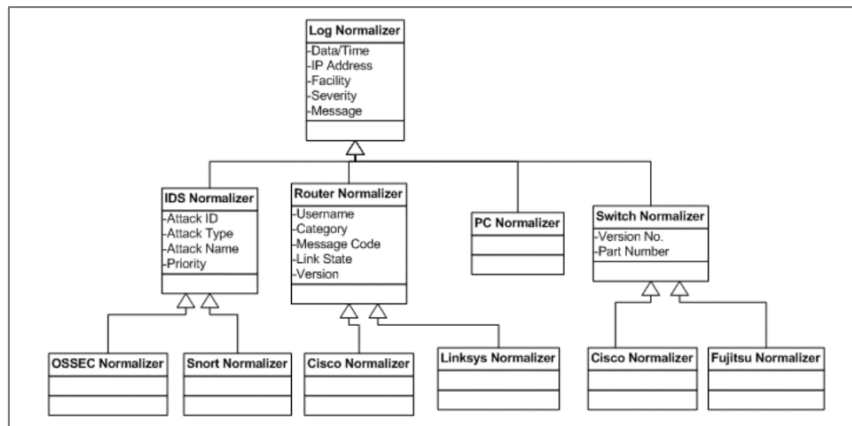


Figure 30: Class diagram of the log normalization module

The problem is that we do not have the right to use the company’s resources because they are in production (in operation); in this case, we have considered using a virtual environment to carry out the project. However, we quickly realized we didn’t have enough resources to simulate a whole network with the different sources of logs. Therefore, to overcome all these problems, we decided to: choose and then study and modify an open-source solution so that it is adapted to our system design and thus meets the needs of the company.

The solution chosen should make it possible to centralize and analyze the logs and generate alerts and recommendations, as well as provide a global overview of the system state in the form of a dashboard. To solve the problem of (collection, research, and visualization), we decided to use the ELK suite. Elasticsearch for search, Logstash for collection, and Kibana for visualization.

III.2 ELK Stack

III.2.1 ELK definition

ELK is an open-source suite comprising three main components: Elasticsearch, Logstash, and Kibana. Beats were later added to form the ELK stack. The ELK Suite allows the logs aggregation from all the systems and applications, analyzing them and creating visualizations for application and infrastructure monitoring. (INTRODUCTION TO ELK STACK).

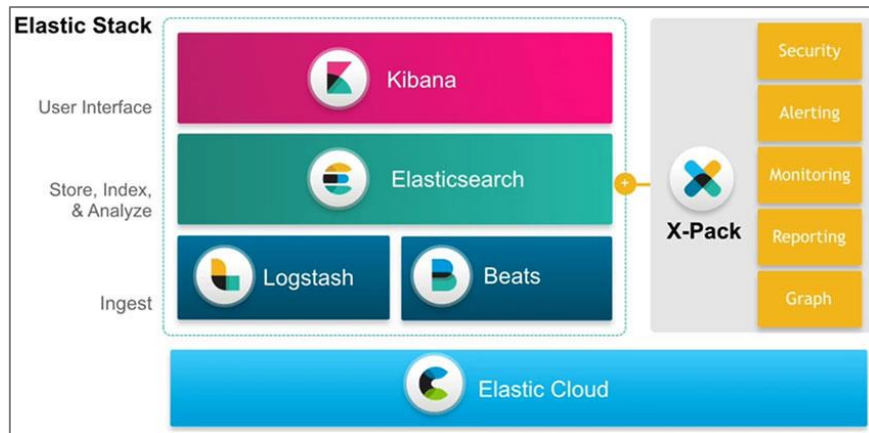


Figure 31 : ELK Stack

Elasticsearch is the main component, which centralizes information and accesses it via a RESTful API, Logstash allows data aggregation in Elasticsearch, Kibana enables the creation of dashboards and the visualization of data in Elasticsearch, and Beats can be installed on the machines to be monitored as an agent to send you the logs.

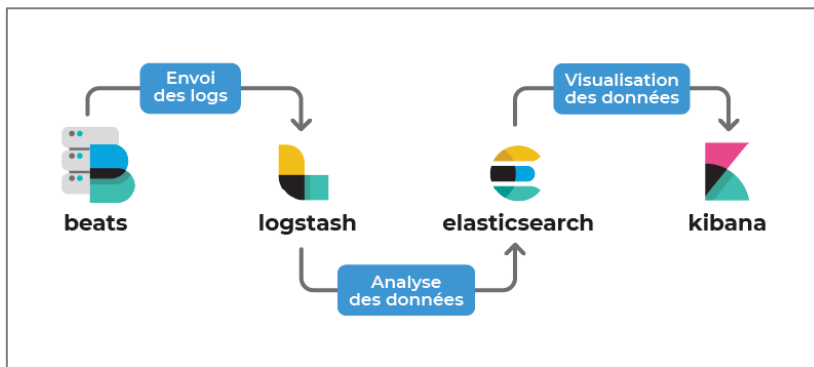


Figure 32: ELK Stack component

III.2.2 ELK Architecture

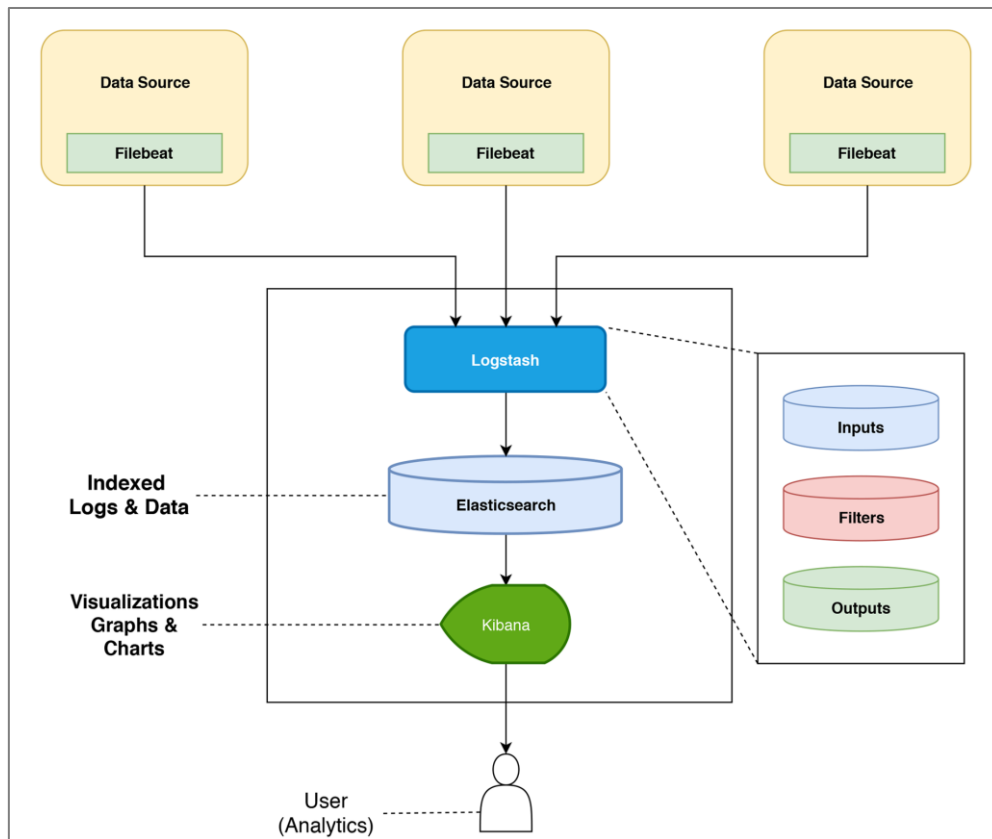


Figure 33: ELK Stack architecture

III.2.3 ELK components

➤ Elasticsearch

Elasticsearch is a Java distributed search and analytics engine based on Apache Lucene. It was created by Shay Banon in 2004, released in 2010, and became the most popular search engine. It is used for log analytics, full-text search, security intelligence, business analytics, and operational intelligence. This system gives it flexibility, scalability, and speed that other search engines of this type do not have. Thanks to this, one can carry out all kinds of searches, whether with a simple keyword or even an entire phrase. (Ioannis Voulgaris, 2020). In addition, the requests made to find information on ElasticSearch are sent using REST APIs, making it perfectly adaptable to all circumstances.

- *Apache Lucene:*

Apache Lucene is a free, open-source Java-based search library providing Application Programming Interfaces for performing standard search and search-related tasks like indexing, querying, highlighting, language analysis, and many others. (Białecki et al.).

Lucene is a full-text search, meaning a program searches for one or more user-defined terms in a series of text documents. This shows that Lucene is not only used in the context of the World Wide Web, even though search functions are ubiquitous on the web. Lucene can also be used for archives, libraries, or even the home PC. Furthermore, Lucene not only searches HTML documents and works with emails or PDF files. (Apache Lucene : Tutoriel - IONOS.).

The core of Elasticsearch is the Apache Lucene library, which includes features for indexing, searching, retrieving, and updating documents, and text analysis. The figure defines the integration between Elasticsearch and Lucene and their external interactions :

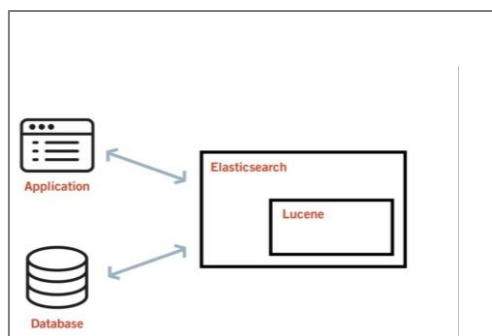


Figure 34: Elasticsearch and Lucene integration

- **Key concepts of Elasticsearch:**

ElasticSearch has 6 fundamental concepts (Interrogez Efficacement Vos Bases de Données Avec ElasticSearch):

- **The *node*:** a node refers to a running instance of the ElasticSearch software. In other words, it is an application process executing on a machine. A server can run one or more ElasticSearch instances depending on its resources. Nodes can also run on a cluster of machines when high availability is needed;
- **The *cluster*:** an ElasticSearch cluster is a set of one or more nodes. The cluster provides collective indexing and distribution of search queries across all cluster nodes;

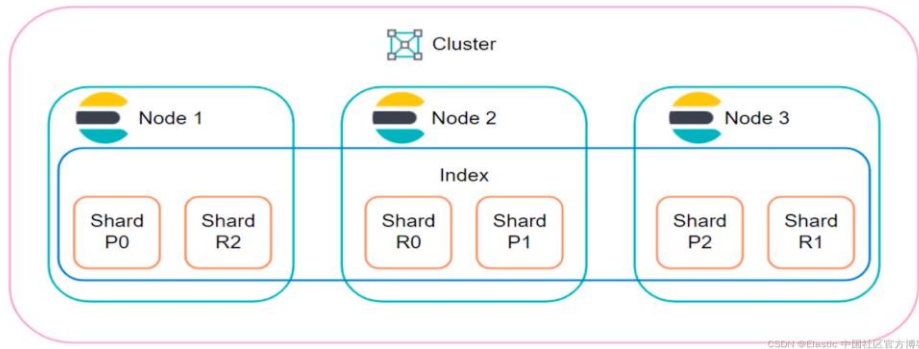


Figure 35: Elasticsearch cluster

- **Index:** an index is a collection of JSON documents. Elasticsearch server can store unlimited index due to its distribution system and clusters;

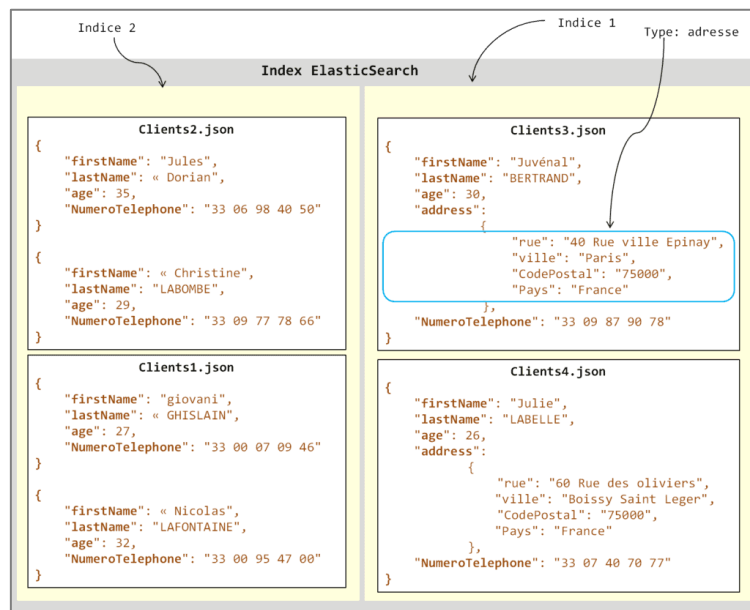


Figure 36: Elasticsearch index

- **Mapping:** refers to documents that share a standard set of fields in the same index;
- **The document:** is a set of fields or properties defined in a specific way in the JSON format. Each document belongs to a type and resides in an index;
- **The score (Shard):** is a part of the index. Indexes are horizontally partitioned into *shards*, and these partitions are distributed in the nodes of the ElasticSearch cluster. Each partition contains all the properties (fields) of the document. The distribution of partitions in the cluster distinguishes between the notion of *primary* and *secondary partitions*;

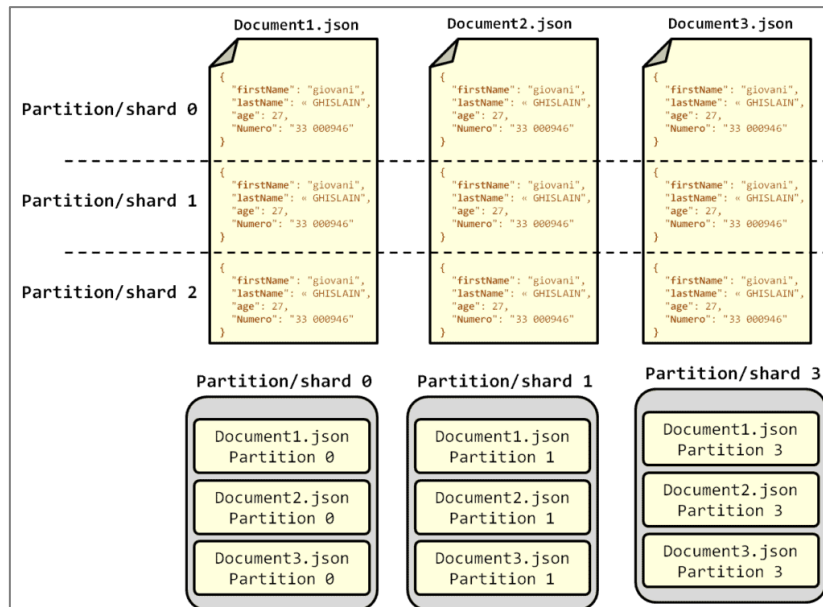


Figure 37: Elasticsearch partition and shard

- The **replica**: indexes and partitions are replicated across the cluster nodes to increase the high availability of processing in the event of a failure. In addition, replicas make it possible to parallelize content search processing in the cluster, which increases the performance of Elasticsearch;

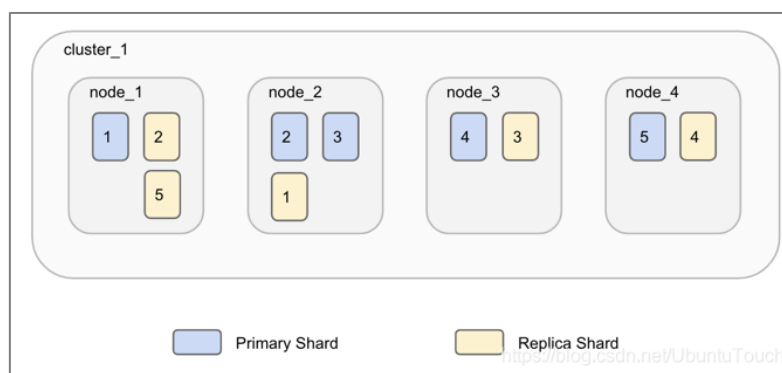


Figure 38: Elasticsearch replica

Elasticsearch has a highly benefits, especially when using Big Data:

- **Distributed architecture**: Elasticsearch could be installed in two different ways (Defined in detail in the next chapter). Installing instances (nodes) of Elasticsearch on the same machine or installing it on multiple machines in a cluster. The documents stored are divided into partitions with different sizes, then distributed and replicated on each node of an Elasticsearch cluster. Due to the processing

method, the data has high availability, and the searches will not be interrupted, even in the event of a malfunction on one or more nodes of the cluster.

- *Document-oriented structure:* Elasticsearch is also a NoSQL database, which allows it to process text documents, logs, objects, and even an existing database. It facilitates the information management, analyses, and searches better than a relational database such as MySQL . Data is stored in JSON format and aggregated to form an index. The documents of identical format and type are put in the same index.
- *Processing speed:* Elasticsearch is a real-time search engine due to the features below: The consultation of the data inserted into Elasticsearch is made only in a few seconds after storing and indexing it in the database. Since it has a distributed architecture and processing is done through an index, the searching time for information on the system database is deficient, depending on the complexity and scope of the search.
- *Ease of communication:* Sending a request for data or information to the Elasticsearch database is passed through the REST API, which is managed by the software. When a request is made via the Elasticsearch REST API, it is first sent to the cluster's central node and then distributed to the other nodes called "slave" nodes, where a data partition is stored. Finally, the result is synthesized and then returned to the requestor. This query system is used parallel with all programming languages such as Java, Python, R, etc.
- *Open source:* Elasticsearch is distributed under an Elastic license and SSPL (or Server Side Public License). The SSPL license is a medium between an open-source license and a proprietary license that allows developers to use it as a back-end for their solutions. In addition, the documentation is available in several languages, making it easier for users and developers worldwide to learn about. Elasticsearch is free; however, there are a few paid features in recent versions of Elasticsearch.
- *Ease of cloud integration:* Elasticsearch provides an unlimited volume of storage space. Elasticsearch can be used on all existing cloud solutions, whether on a private or public cloud. Clouds give the integrity of this software among the tools they offer to manage all the clusters centrally.

➤ Logstash

Logstash is an open-source server-side data processing pipeline that collects data from various sources, transforms it into formatted documents, and sends it to the desired destination. It is a data pipeline that collects and stores them centrally for Elasticsearch, open-source analytics, and search engines. Logstash is a popular choice for loading data into Elasticsearch for its powerful log processing capabilities, and it's integrated into over 200 open-source plugins.

Indexers like Lucene are used to index the logs for better search performance, and then the output is stored in Elasticsearch or other output destination. The data in output storage is available for Kibana and other visualization software. (Logstash - Internal Architecture).

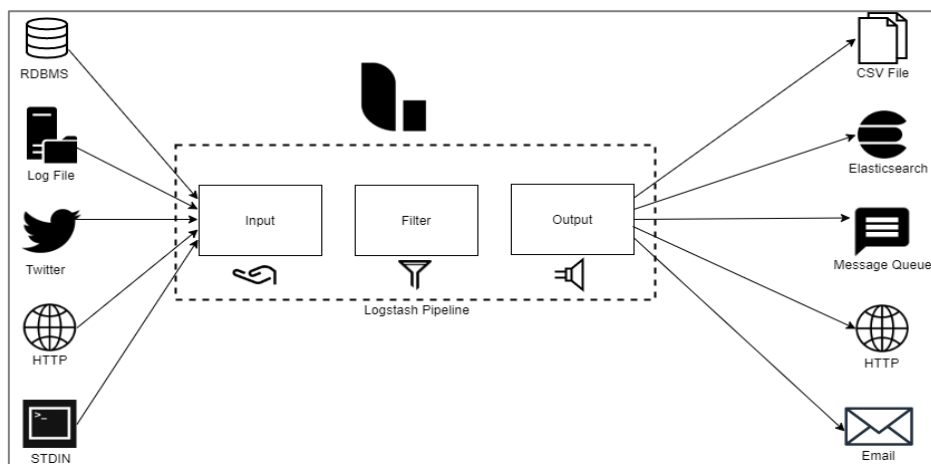


Figure 39: Logstash architecture

A Logstash pipeline is composed of three elements: input, filter, and output. It consists of choosing the inputs, configuring the filters, and extracting the relevant data from the logs.

- The input plugins capture data from different sources

Example of Logstash input plugin:

- File
 - Beats
 - MongoDB
 - Elasticsearch
 - Http
- The filter plugins modify the input data according to the specification, such as removing a specific field or converting the unstructured data into structured data.
Example of Logstash filter plugins:

- GeoIP
 - CSV
 - JDBC
 - Date
 - JSON
 - XML
- The output plugins send data from Logstash to single or multiple sources. Examples of Logstash output plugins are as follows:
 - File
 - MongoDB
 - Elasticsearch
 - Http
 - Nagios

Logstash filters plugins are available on the official elastic website¹

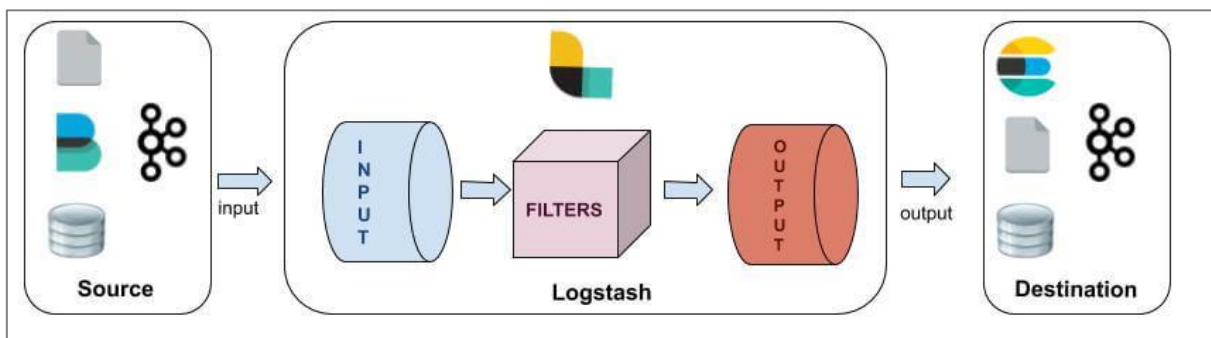


Figure 40: Logstash pipeline

Depending on the configuration file, a Logstash agent can act in different roles:

- Shipper: send the collected events to another Logstash instance or another software.
- Broker and Indexer: receives and indexes the events.
- Search and Storage: searching and storing the events.
- Web Interface: different options available: native one, based on Elasticsearch Shipper, Indexer, Broker, Searching/Storage, Web interface.

¹ <https://www.elastic.co/guide/en/logstash/current/filter-plugins.html>

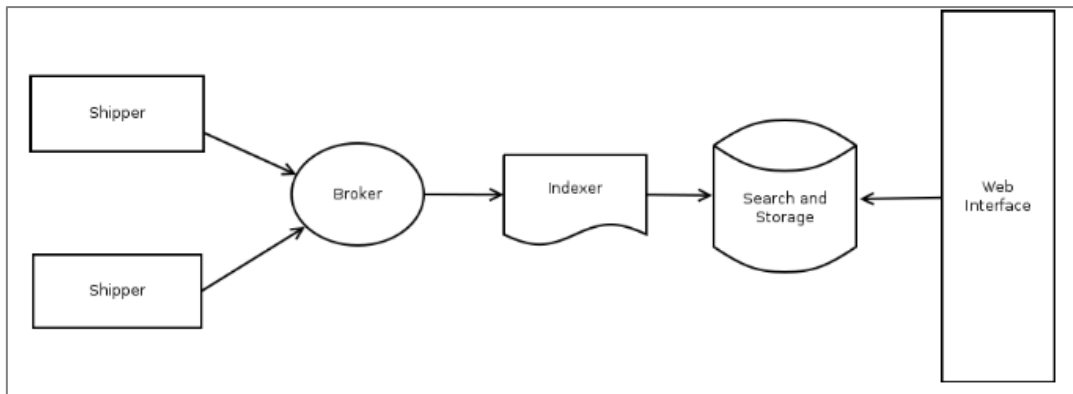


Figure 41: Logstash internal architecture

Logstash is used for:

- *Log analysis*: Logstash ingests unstructured and semi-structured logs from a variety of data sources.
- *IT Operations management*: The clustering algorithm groups the data using Elasticsearch, and Kibana provides a visualization.
- *Pre-integrated filters*: Logstash filters allow quickly transforming the most common data types, indexing them into Elasticsearch, and running queries without creating custom data transformation pipelines.

➤ **Beats**

Beats are open-source data shippers that install as agents on the servers to send operational data to Elasticsearch. Its role is to complement Logstash. Logstash is a server-side component, while Beats has a role client-side. Beats consist of a basic library, libbeat, which provides an API for shipping data from the source, setting input options, and implementing logging. Beats is installed on machines not part of server-side components such as Elasticsearch, Logstash, and Kibana. These resident agents on nodes that are not part of the cluster can also be called nodes of edge nodes. (What Are Beats? | Beats Platform Reference [8.2] | Elastic).

There are several types of beats, are described in the table below:

Table 8: Beats types

<p>Auditbeat Audit Data</p>	<p>Lightweight agent designed for audit data transfer Collect data from the Linux audit framework and monitor the integrity of files. Auditbeat forwards these events in real-time to the Elastic Stack for further analysis.</p>
<p>Filebeat Log files and journals</p>	<p>Lightweight transfer agent designed for logs. Filebeat is a lightweight transfer agent that allows centralizing the logs and files collected from any source.</p>
<p>Functionbeat Cloud data</p>	<p>Serverless transfer agent designed for cloud data. Deploy Functionbeat as a function on the cloud provider's FaaS platform. Data can be collected, transferred, and monitored from cloud services.</p>
<p>Heartbeat Heart Beat Availability</p>	<p>Lightweight transfer agent designed for availability monitoring. Monitor service availability with active detection. From a list of URLs, Heartbeat asks this question: "are you active?" Heartbeat then passes this information and response time to the Elastic Stack for further analysis.</p>
<p>Metricbeat Metrics</p>	<p>Lightweight transfer agent designed for indicators. Collect metrics from the systems and services. Statistics related to CPU, memory, Redis, or NGINX... Metricbeat takes care of the lightweight statistics transfer for the systems and services.</p>
<p>Packetbeat Network traffic</p>	<p>Lightweight agent designed for network data transfer. Packetbeat is a lightweight network packet analyzer that sends data from the hosts and containers to Logstash or Elasticsearch.</p>
<p>Winlogbeat Windows event log</p>	<p>Lightweight agent designed for forwarding Windows event logs. Send Windows event log streams to Elasticsearch and Logstash with Winlogbeat.</p>

- **Beats Architecture:**

Beats can send data directly to Elasticsearch or via Logstash, where it can further process and enhance the data before visualizing it in Kibana.

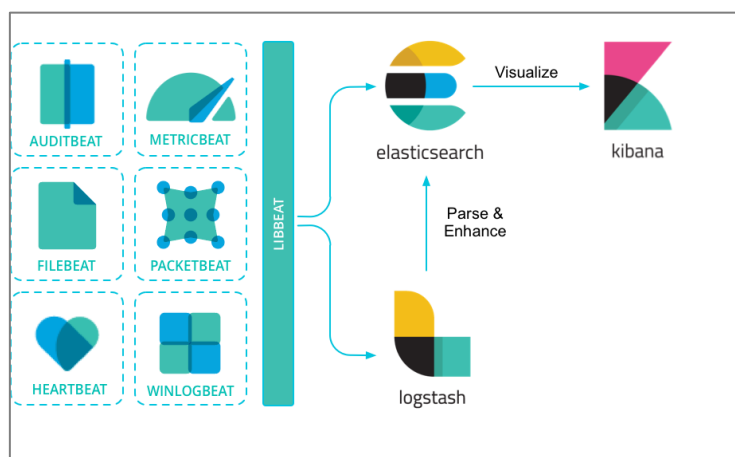


Figure 42: Beats architecture

➤ Kibana

Kibana is a real-time data visualization and exploration tool for log and time-series analytics, application monitoring, and operational intelligence. It offers features such as histograms, line charts, pie charts, heatmaps, and built-in geospatial support, making Kibana the default choice for viewing data stored in Elasticsearch. (Ioannis Voulgaris, 2020). Kibana connects to the browser and refers to a gateway to the Elastic cluster and the data it hosts.

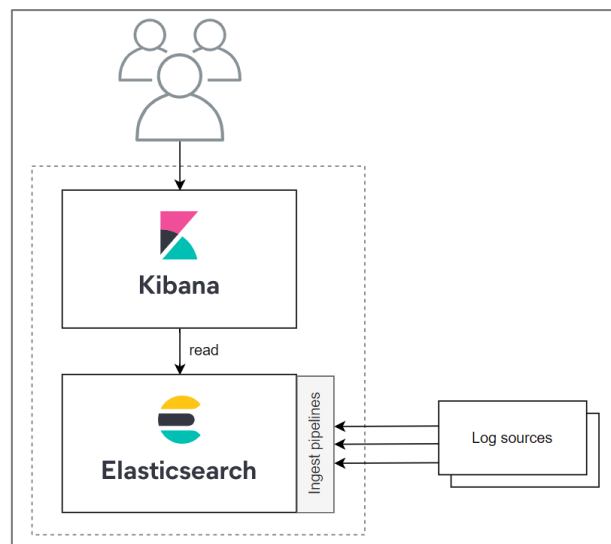


Figure 43: Kibana architecture

- **Benefits of Kibana:**

- *Mapping support:* Kibana has powerful geospatial features that seamlessly overlay geographic information on the data and visualize the results on maps.
- *Pre-integrated aggregations and filters:* With Kibana's pre-built aggregations and filters, it can run various analytics such as histograms, top-N queries, etc.
- *Easy-to-access dashboards:* The dashboards and reports are easily set up and shareable. It needs only a browser to view and explore the data.

➤ X-Pack

X-Pack is an Elastic Stack extension that integrates across the entire Elastic Stack. It provides many capabilities such as security, monitoring, and reporting and includes essential premium features like Alerting, Machine Learning, and Monitoring. (Set up X-Pack | Elasticsearch Guide [7.17] | Elastic).

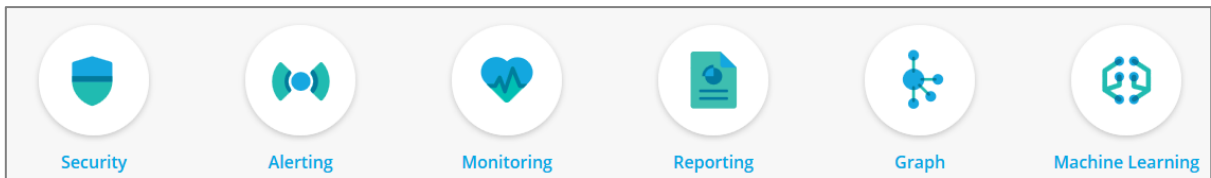


Figure 44: X-Pack components

- **The Elastic X-Pack features:**

- *Security (formerly Shield):* X-Pack fully integrates with authentication systems like Active Directory and LDAP, custom-created realms with home-grown identity management systems and built-in native authentication. X-Pack Security grants Elastic Stack the power to manage users and roles, delivering specific access to the right users. Multitenancy support grants users access to specific Elasticsearch indices. SSL/TLS encryption can be used to secure node-to-node, HTTP, and transport client traffic across your Elastic Stack. Protect access to cluster from unapproved hosts with IP Filtering. Auditing logs let you maintain a complete record of all system and user activity.
- *Alerting (formerly Watcher):* Alerting features in X-Pack give the full power of the Elasticsearch query language to identify changes in the data that are interesting to the infrastructure team. Get alerts in communications, like Slack, email, etc., to integrate with the existing monitoring infrastructure. Common examples of alerts could be:
 - ✓ Logging from multiple locations for the same user
 - ✓ A component of a system is nearing the end of life
 - ✓ Elasticsearch indexing rate has plummeted
- *Monitoring (formerly Marvel):* The X-Pack monitoring features help keep a pulse on Elastic Stack performance... Its dashboards help assess their status at various levels, providing the user with all the needed information to keep Elastic Stack optimized. Analyze the performance of the current cluster against historical data to help with future capacity planning. Multiclustert monitoring support helps to simplify the workflow.
- *Reporting:* The Reporting feature can handle large volumes of reporting requests, such as Kibana visualization or dashboard, on-demand, scheduled for

later, or triggered by conditions. These reports are easily shareable with stakeholders and can be scheduled to be delivered at specific times.

- *Graph*: Graph offers a relationship-orientated approach that lets the user explore the connections in data using the relevance capabilities of Elasticsearch. The charts distinguish between popularity and relevance and explore the Elasticsearch indices to uncover hidden relationships.
- *Machine Learning*: X-Pack machine learning features automatically model the behavior of Elasticsearch data — trends, periodicity, and more — in real-time to identify issues faster, streamline root cause analysis, and reduce false positives.

With X-Pack, the Elastic Stack experience can be infinitely enriched. Here's how:

- **Secure it**: Activate authentication for the cluster and define roles and permissions.
- **Monitor it**: Maintain a pulse on Elastic Stack to keep it firing on all cylinders.
- **Report it**: Easily generate and share reports of your Kibana visualizations.

III.2.4 Employing ELK Stack for SIEM solution

As we concluded in the previous part, The ELK stack is likely a log analysis and management platform.

The ELKStack provides collection, processing, normalization, enhancement, and storage of log data from various sources. These processes are grouped under the term “log management.”

Log management is known as a necessary part of any Security Information and Event Management (SIEM) solution (According to the SIEM definition); however, log management by itself is insufficient to consider a fully SIEM solution.

To build a SIEM solution (as we saw in the theoretical background chapter), there are some characteristics that the system should have to call it a SIEM.

- *Log collection*: Aggregate data from multiple data sources, including applications. Using a combination of Beats and Logstash.
- *Log processing*: The logs from different sources are generated in other formats and must be normalized. The normalization process involves transforming the various log messages into meaningful field names, mapping the field types correctly, and

enriching specific fields where necessary. Logstash supports many different filter plugins for log parsing and can break up the logs, enrich specific fields with geographic information, drop fields, and add fields, for instance.

- *Storage and retention:* ELK doesn't perform log archiving.
- *Querying:* It is a phase after collecting, parsing, and indexing the data in Elasticsearch. Log queries enable the user to conduct a forensic investigation into previous security incidents by using Kibana for visualization.
- *Dashboards:* Kibana supports a wide array of different visualization types such as pie charts, graphs, geographical maps, single metrics, data tables, etc.
- *Correlation:* A correlation rule defines the specific sequence of events. No such thing exists within the free ELK stack.
- *Alerts:* The ELK Stack does not ship with a built-in mechanism for alerting on suspicious activity. But this capability will be enabled using other tools like X-pack (paid version).
- *Incident Management:* The ELK stack does not have many capabilities for incident management.

The following comparative table is a resume of ELK Stack features available in the open-source version.

The whole table is available on their official website²:

² <https://www.elastic.co/fr/subscriptions>

Table 9: Elastic Stack features subscriptions

	FREE		GOLD	PLATINUM
	OPEN SOURCE	BASIC		
	Download		Request Info	Request Info
ELASTIC STACK				
Elasticsearch				
✓ Scalability & Resiliency	✓	✓	✓	✓
✓ Query & Analytics	✓	✓	✓	✓
✓ Data Enrichment	✓	✓	✓	✓
✓ Management & Tooling	✓	✓	✓	✓
✓ Security		✓	✓	✓
✓ Alerting			✓	✓
✓ Machine Learning				✓
Kibana				

The ELK Stack can not form a complete SIEM solution. The table below sums up the previous reasons:

Table 10: ELK Features

	Yes/No	The but...
Log collection	Yes	One of the ELK Stack's core capabilities. Large and diversified data pipelines involve more than the ELK Stack however.
Log processing	Yes	Complex processing requires close monitoring and architectural considerations in designing the pipeline.
Storage	Yes	Requires organizational engineering commitment and expertise for guaranteeing HA and scalability. For historical data, long-term retention is required.
Querying	Yes	Depends on accurate parsing. Requires a certain amount of expertise.
Correlation	No	ELK does not provide correlation rules. Correlation will depend heavily on analytical work.
Dashboards	Yes	Extremely powerful, requires expertise.
Alerts	No	Is not provided out-of-the-box. Can use hosted ELK, commercial or open source add-ons.
Incident management	No	Totally out of scope for the ELK Stack. Requires hooking in additional tools.

ELK Stack is a powerful tool for centralized logging. However, it cannot be used as-is for SIEM, which requires a security analyst. The missing features in ELK Stack free open source are alerting correlation rules and incident management.

The ELK Stack can be augmented with other platforms and services as a SIEM tool. The next point defines the platform that we integrate into the ELK stack.

III.3 Wazuh

III.3.1 Wazuh definition

Wazuh is an open-source host-based intrusion detection system (HIDS) for threat prevention, recognition, and reaction. It gets on-premises, virtualized, containerized, and cloud workplaces. Wazuh is broadly used by many organizations, from small businesses to large enterprises; it provides comprehensive intrusion detection service for most operating systems, including Linux, OpenBSD, FreeBSD, macOS, Solaris, and Windows.

The Wazuh solution comprises a few endpoint security agents sent on the monitored systems and an administration server, which gathers and analyzes the information collected by the agents. Moreover, Wazuh has been completely coordinated with Elastic Stack, providing a web search tool and information visualization device that permits users to explore through security alerts. (Mattia Incoronato, 2020).

III.3.2 Wazuh features

The main capabilities that Wazuh provides are as follows:

- *Intrusion detection:* agents filter the monitored systems are seeking out malware, rootkits, and doubtful anomalies. They can distinguish hidden records, cloaked processes or unregistered network audience members, and irregular system call responses.
- The server component uses a signature-based approach to intrusion detection, utilizing its regular expression engine to analyze collected log information and explore for pointers of compromise.
- *Log data analysis:* peruses operating system and application logs and safely forwards them to a central manager for rule-based analysis and capacity. When no agent is sent, the server can receive data from network devices or applications using Syslog.
- *File integrity monitoring:* monitors the record system, recognizing changes in content, authorizations, ownership, and properties of records they got, to keep an

eye on. It natively recognizes users and applications utilized to create or modify files in expansion. File integrity monitoring capabilities can be combined with threat insights to distinguish threats or compromised hosts. In addition, a few administrative compliance measures, such as PCI DSS, require it.

- *Vulnerability detection*: drag computer program inventory data and send it to the server, correlated with continuously upgraded CVE (Common Vulnerabilities and Exposure) databases, to recognize well-known vulnerable software.
- *Configuration assessment*: monitors system and application setup settings to guarantee compliance with security policies, measures, and hardening guides. Agents perform periodic scans to identify applications that are known to be vulnerable, unpatched, or insecurely configured.
- *Incident response*: out-of-the-box active responses to perform different countermeasures to address active threats, such as blocking access to a system from the danger source when specific criteria are met. Regulatory compliance – gives a few essential security controls to comply with industry guidelines and rules. These features, combined with its adaptability and multi-platform support, offer assistance to organizations to meet specialized compromise.
- *Regulatory compliance*: gives a few basic security controls to comply with industry guidelines and rules. These features, adaptability, and multi-platform support assist organizations that meet specialized compliance requirements.
- *Cloud security*: helps monitor cloud infrastructure at an API level, utilizing integration modules that can drag security information from well-known cloud suppliers, such as Amazon AWS, Azure, or Google Cloud. In expansion, Wazuh provides rules to evaluate the configuration of the cloud environment, quickly spotting weaknesses.
- *Containers security*: provides security visibility into the Docker hosts and holders, checking their behavior and recognizing dangers, vulnerabilities, and anomalies. The Wazuh agent has local integration with the Docker engine permitting users to monitor pictures, volumes, network settings, and running containers.
- *Wazuh WUI*: gives an effective user interface for information visualization and analysis. This interface can manage the Wazuh setup and monitor its status.

III.3.3 Wazuh components

The Wazuh solution consists of several endpoint security agents deployed on the monitored systems and a management server, which collects and analyzes the data collected by the agents. Additionally, Wazuh has been fully integrated with Elastic Stack, providing a search engine and data visualization tool that allows users to navigate their security alerts.

The following figure represents Wazuh architecture composition.

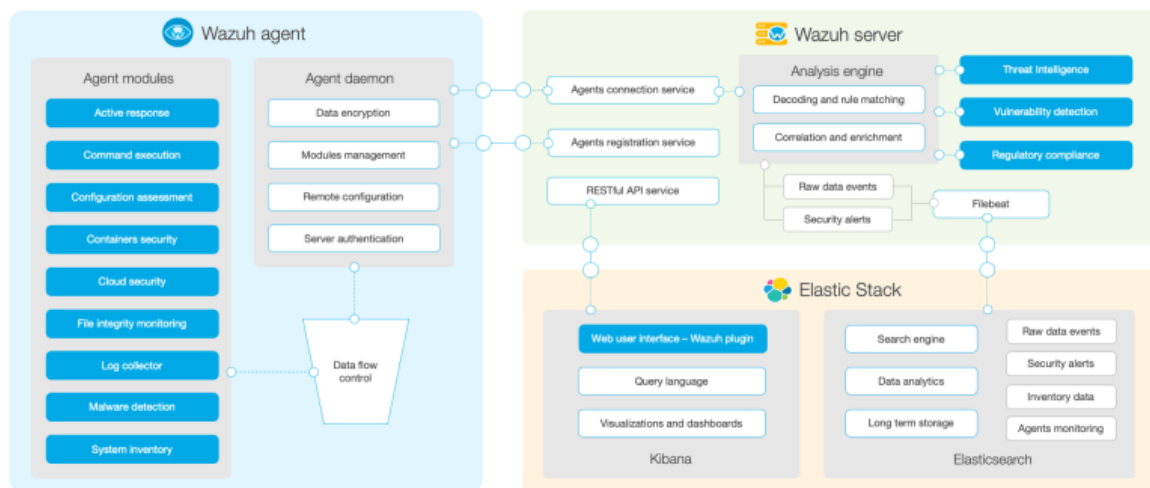


Figure 45: Wazuh components architecture

III.3.3.1 Wazuh agent

The Wazuh agent: is multi-platform and installed on endpoints such as laptops, desktops, servers, cloud instances, or virtual machines; it provides prevention, detection, and response capabilities, and it communicates with the Wazuh server through an encrypted and authenticated channel, and send data in real-time. It supports Windows, Linux, macOS, HP-UX, Solaris, and AIX platforms. (Wazuh Agent - Components · Wazuh Documentation).

a) Agent architecture

The Wazuh agent includes a modular architecture. Each component is responsible for its tasks, counting, monitoring the file system, reading log messages, collecting inventory information, scanning the system setup, and searching for malware. In addition, users can

oversee agent modules through configuration settings, adjusting the solution to their specific use cases. The chart below represents the agent architecture and components:

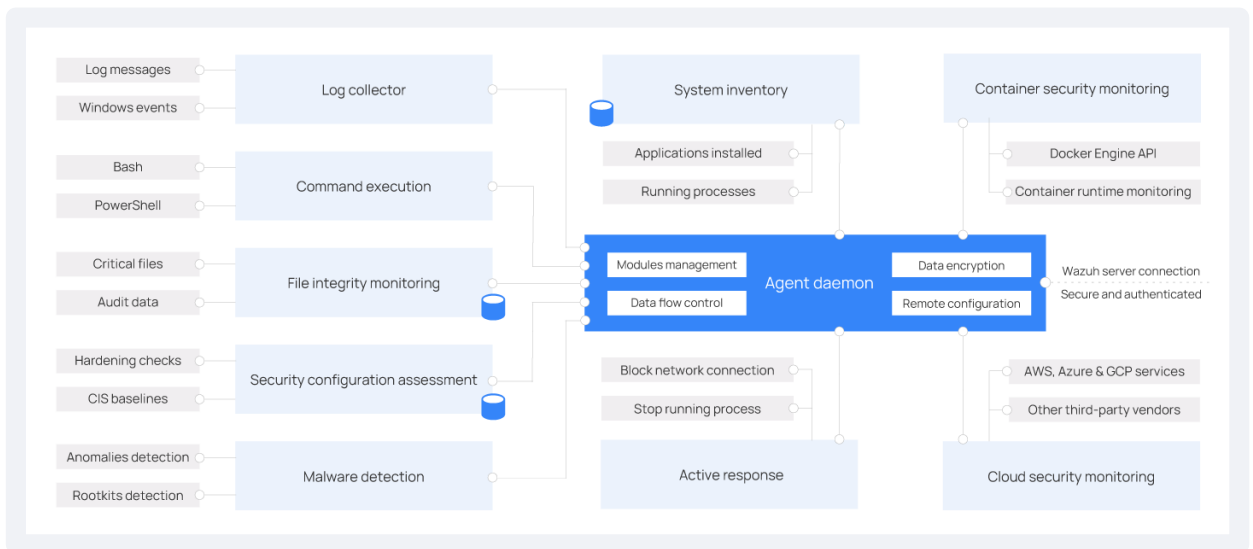


Figure 46: Wazuh Agent architecture

b) Agent modules

All agent modules are configurable and perform diverse security tasks. This modular design enables or disables each component according to the security needs. Below is a description of the various purposes of all the agent modules. (Wazuh Agent - Components · Wazuh Documentation.).

- **Log collector:** This agent component can examine log files and Windows events, collecting operating system and application log messages. It supports XPath filters for Windows events and recognizes multi-line formats like Linux Audit logs. In addition, it can enhance JSON events with extra metadata.
- **Command execution:** Agents occasionally run authorized commands, collecting their output and detailing it to the Wazuh server for advance analysis. This module is utilized for various purposes, such as monitoring hard disk space cleared out or getting a list of the final logged-in users.
- **File integrity monitoring (FIM):** This module monitors the file system, detailing when files are created, erased, or altered. It keeps track of file attributes, authorizations, ownership, and content changes. When an event happens, it captures the details of who, what, and when in real-time. Also, the FIM module builds and

maintains a database with the state of the monitored files, permitting queries to be run remotely.

- *Security configuration assessment (SCA)*: This component provides continuous configuration assessment, utilizing out-of-the-box checks based on the Center of Internet Security (CIS) benchmarks. Users can claim SCA checks to monitor and enforce their security policies.
- *System inventory*: This agent module periodically scans, collecting stock information such as operating system version, network interfaces, running processes, installed applications, and a list of open ports. Filter results are stored in local SQLite databases that can be queried remotely.
- *Malware discovery*: Employing a non-signature-based approach, this component is capable of detecting anomalies and the possible presence of rootkits. Too, it looks for hidden processes, files, and covered-up ports, whereas the monitoring system calls.
- *Active response*: This module runs automatic actions when threats are recognized, triggering responses to block a network connection, stop a running prepare, or erase a malicious file. Users can moreover make custom responses.
- *Container security monitoring*: This agent module is integrated with the Docker Engine API to monitor changes in a containerized environment. For example, it detects changes to network configuration or data volumes. It also alerts about containers running in privileged mode and around users executing commands in a running container.
- *Cloud security monitoring*: This component monitors cloud suppliers such as Amazon AWS, Microsoft Azure, or Google GCP. It natively communicates with its APIs. It is capable of identifying changes to the cloud infrastructure and collecting cloud services log information.

III.3.3.2 Wazuh server

Wazuh Server: Analyzes data received from agents, processes it through decoders and rules, and triggers alerts using the MITRE ATT&CK framework and regulatory compliance requirements such as PCI DSS, GDPR, HIPAA, CIS, and NIST 800-53 when threats or anomalies are detected. A single server can analyze data from hundreds or thousands of agents and scale horizontally when configured in a cluster. The server is also

used to manage agents, configuring and upgrading them remotely if necessary. (Components - Getting Started with Wazuh · Wazuh Documentation).

a) Server architecture

The Wazuh server runs the analysis engine, the Wazuh RESTful API, the agent enrollment service, the agent connection service, the Wazuh cluster daemon, and Filebeat. The server is installed on a Linux operating system and usually runs on a stand-alone physical machine, virtual machine, docker container, or cloud instance.

The diagram below represents the server architecture and components:

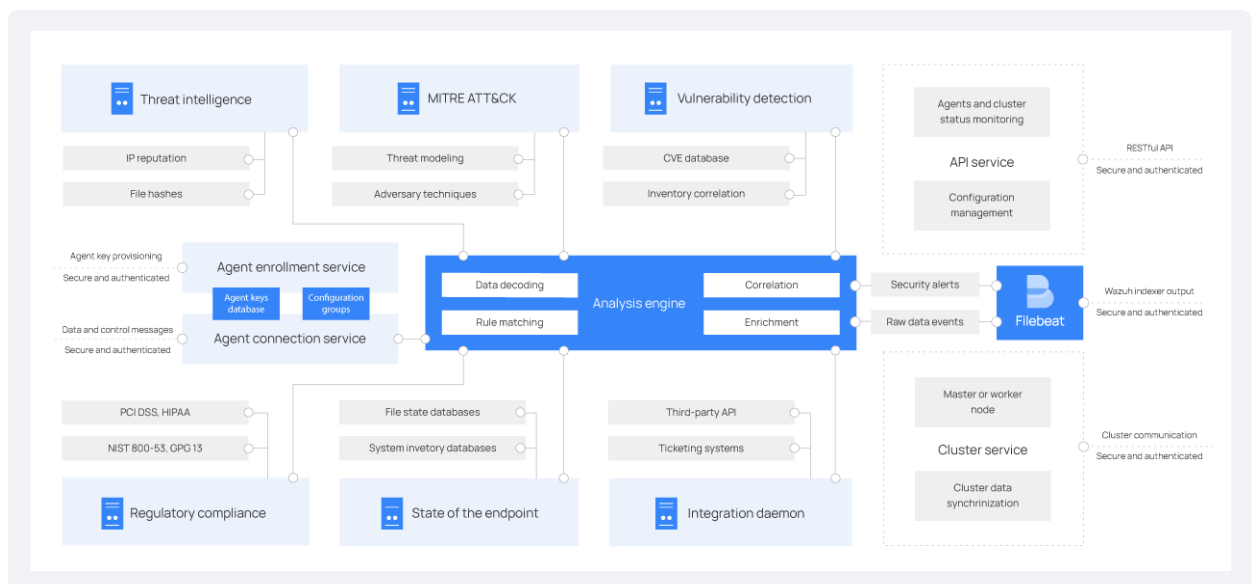


Figure 47: Wazuh server architecture

b) Server components

The Wazuh server comprises a few components listed below that have diverse functions:

- *Agent enrollment service*: provides and distributes unique authentication keys to each agent. The process runs as a network service and supports authentication via TLS/SSL certificates or by providing a fixed password.
- *Agent connection service*: receives data from the agents. It uses the keys shared by the enrollment service to validate each agent's identity and encrypt the communications between the Wazuh agent and the Wazuh server. Additionally, this service provides centralized configuration management, enabling you to push new agent settings remotely.

- *Analysis engine*: This is the server component that performs the data analysis. It uses decoders to identify the type of information being processed (Windows events, SSH logs, web server logs, and others). These decoders extract relevant data elements from the log messages, such as source IP address, event ID, or username. Then, by using rules, the engine identifies specific patterns in the decoded events that could trigger alerts and possibly even call for automated countermeasures (e.g., banning an IP address, stopping a running process, or removing a malware artifact).
- *Wazuh RESTful API*: provides an interface to interact with the Wazuh infrastructure. It is used to manage the configuration settings of agents and servers, monitor the infrastructure status and overall health, manage and edit Wazuh decoders and rules, and query the state of the monitored endpoints.
- *Wazuh cluster daemon*: used to scale Wazuh servers horizontally, deploying them as a cluster. This configuration, combined with a network load balancer, provides high availability and load balancing. In addition, the Wazuh cluster daemon is what Wazuh servers use to communicate and keep synchronized.
- *Filebeat*: It sends events and alerts to the Wazuh indexer. It reads the output of the Wazuh analysis engine and ships events in real-time. It also provides load balancing when connected to a multi-node Wazuh indexer cluster.

III.3.4 Wazuh agents and Wazuh server communication

The Wazuh agent communicates with the Wazuh server to transport collected data and security-related events. The agent also sends operational data, announcing its configuration and status. Once connected, the agent can be updated, monitored, and configured remotely from the Wazuh server.

The agent communicates with the server through a secure channel (TCP or UDP), giving data encryption and compression in real-time. Moreover, it includes flow control components to avoid flooding, queueing events when essential, and ensuring the network bandwidth.

The agent must be enrolled before connecting it to the server for the first time. This process gives the agent a unique key utilized for authentication.(Wazuh Agent - Components · Wazuh Documentation).

III.4 Wazuh and ELK Stack communication

As above was, mentioned The ELK stack only doesn't form a SIEM solution; however, while integrating Wazuh, which offers additional security functionalities, Wazuh and Elk is a complete SIEM solution; each tool completes the other; the following table resumes the features of each tool:

Table 11: Comparison table for ELK and Wazuh features

	ELK Stack	Wazuh
Log collection	Yes	Yes
Log processing	Yes	Yes
Storage	Yes	Yes
Querying	Yes	Yes
Correlation	No	Yes
Threat detection	No	Yes
Dashboards	Yes	No (With using ELK, we install the Wazuh plugin for Kibana)
Alerts	No	Yes (email or Slack platform configuration)
Incident response	No	Yes

The Wazuh server decodes and checks the rules of the events received with the analysis engine. Events can be spooled to one or both of the following files, depending on whether or not a rule is tripped:

- The file `/var/ossec/logs/archives/archives.json` contains all events whether they tripped a rule or not.
- The file `/var/ossec/logs/alerts/alerts.json` contains only events that tripped a rule with high enough priority (the threshold is configurable).

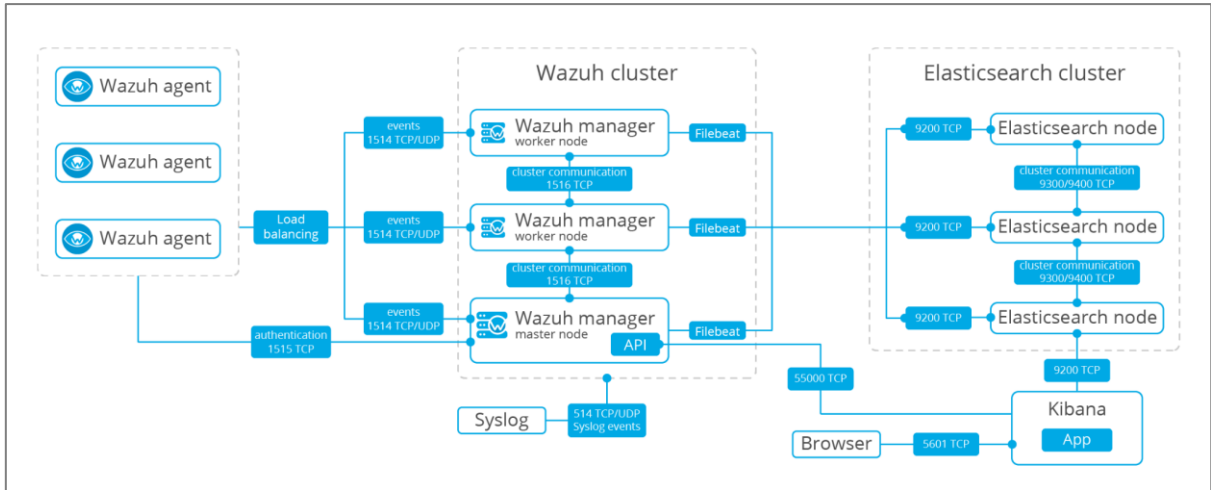


Figure 48: Wazuh and Elasticsearch communication

Below is the list of default ports and protocols used by the component to assure their communication:

Table 12: Wazuh and ELK communication protocols

Component	Port	Protocol	Purpose
Wazuh agent	1514	TCP/UDP	Send the collected events
	1515	TCP	Authenticate with the Wazuh manager
Wazuh manager	1516	TCP	Communication between the wazuh clusters
	514	TCP/UDP	Receive events from Syslog
Wazuh API	55000	TCP	HTTP request to Kibana
Elasticsearch	9300-9400	TCP	Elasticsearch clusters communication
Kibana	5601	TCP	Kibana start server

III.5 Implementation

In computer science, implementation refers to the realization, so the objective of this section includes:

Presentation of the techniques and the tools used for the implementation of the project: we used open-source tools (ELK Stack and Wazuh, that has been already mentioned and explained in detail in the design chapter), we demonstrated our lab environment, the installation steps as well as the chosen deployment.

Demonstrate the interactions of the different actors (Director, analyst): in this part, we present the other interfaces illustrating the most important features of our thematic.

III.5.1 Simplified Network Architecture (DPCASSI - Badr bank)

The figure below shows the architecture of the Badr Bank - DPCASSI network. For confidentiality, some details were deliberately hidden/ changed/ simplified to not divulge information that may compromise the company's security.

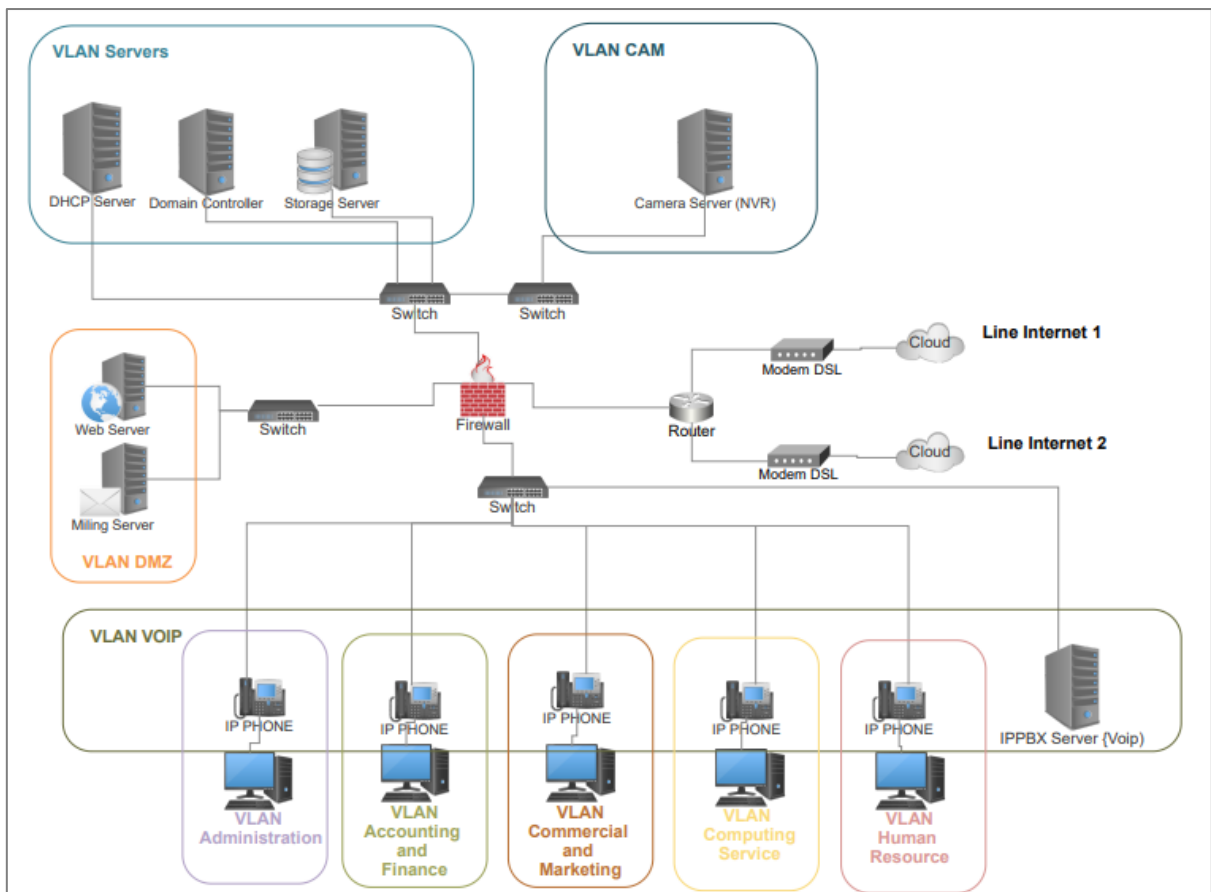


Figure 49: DPCASSI Badr bank network architecture

Description of the architecture and different technologies used: The architecture of the figure is straightforward; the network is segmented into several VLANs:

- *The DMZ VLAN:* contains servers that can be accessed from the internet, including the mail server and the web server.
- *The VLAN administration, accounting, and finance, sales and marketing, IT service, and human resources:* represent the different departments of the company.

- *VoIP VLAN (Voice over IP):* is the VLAN used to transmit the voice. The IPPBX (Internet Protocol Private Branch Exchange) server is in this same VLAN and contains several RJ11 ports for connection with the RTCP (Public Switched Telephone Network); when a telephone call arrives on the analog line connected with the IPPBX, it will convert the received analog signal to a digital signal and will transmit the call to the associated telephone number using the SIP (Session Initiation Protocol, layer: application) protocol.
- *The server VLAN:* contains the enterprise internet servers, including the domain controller, the DHCP server, the DNS server, and the file storage server.
- *The camera VLAN:* contains the network video recorder (NVR) server connected to a switch. The company's IP cameras are all connected to the same switch.
- *The firewall:* allows Controller the flow of data between different VLAN and servers, and it also includes a web filtering service to restrict access to pornographic sites, video games, or social networks (Facebook, YouTube, ...etc.).
- *Two outputs to the Internet: the router uses PBR (Policy Based Routing) routing:* This technique allows routing based on the source IP address and not the destination; the objective is to provide good connectivity to employees by giving access to an internet line with certain VLAN only and access to the other line with the remaining VLAN.

III.5.2 Architecture deployment

To deploy the Elastic Stack (Elasticsearch, Logstash, Beats, Kibana) and the Wazuh plugin for Kibana, there are two types of deployment:

III.5.2.1 All-in-one deployment

It's generally used for testing or in small working environments. It supports around 100 agents. Both Elastic Stack and Wazuh server are installed on the same host. The Wazuh server, including the Wazuh manager as a single-node cluster and Filebeat. The communication is encrypted using certificates.

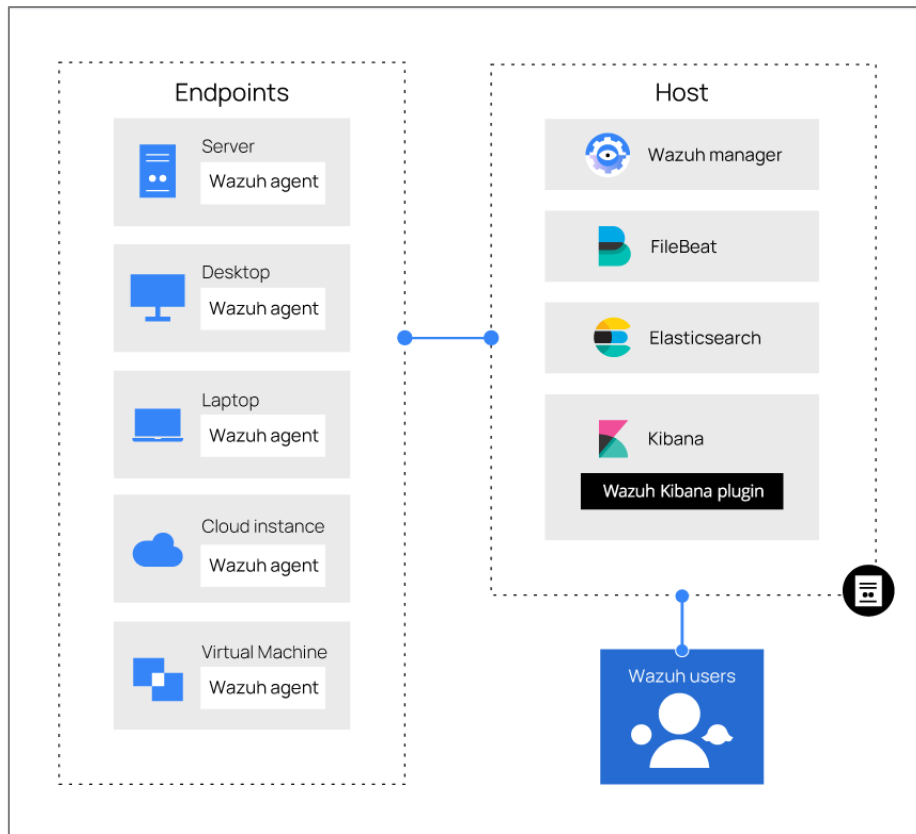


Figure 50: SIEM ALL-in-one deployment architecture

The minimum requirements for the all-in-one deployment are 4 GB of RAM and 2 CPU cores, and the recommended are 16 GB of RAM and 8 CPU cores. In addition, a 64-bit operating system is required.

Table 13: All-in-one deployment hardware requirements

Component	Minimum		Recommended	
	RAM (GB)	CPU (cores)	RAM (GB)	CPU (cores)
Wazuh server + Elastic Stack	4	2	16	8

III.5.2.2 Distributed deployment

It is used in large working environments and allows high availability and scalability of the product. Both Elastic Stack and Wazuh server are installed in a separate host as a single-node or multi-node cluster. However, Kibana could be installed in the same server as the Elasticsearch node or on a separate one.

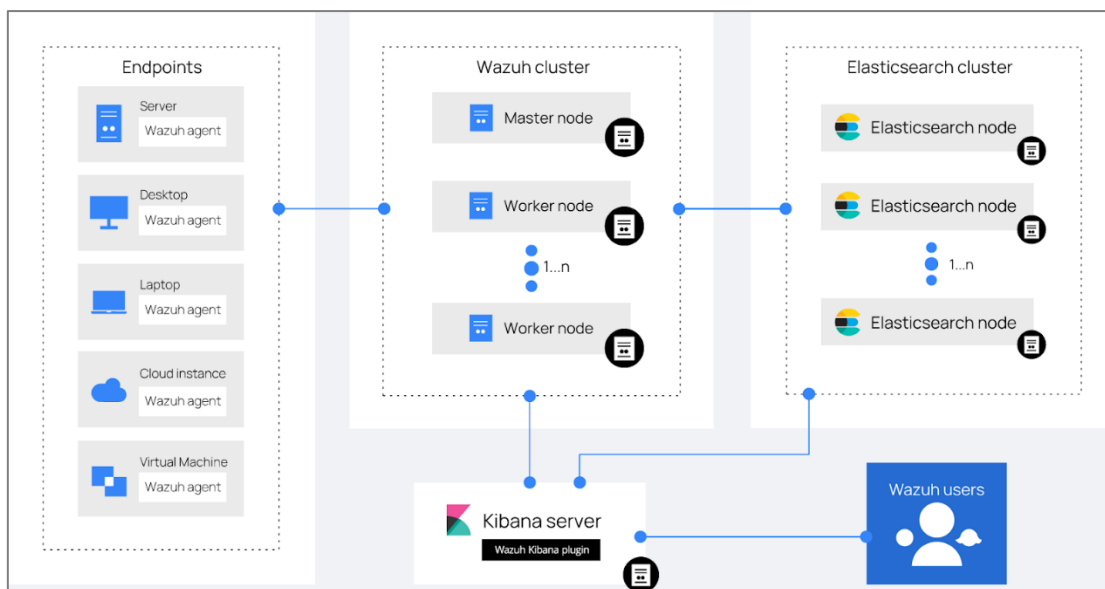


Figure 51: SIEM Distributed deployment architecture

The minimum requirements for distributed deployment are :

- For the Wazuh server: 2 GB of RAM and 2 CPU cores; the recommended are 8 GB of RAM and 4 CPU cores.
- For Elastic Stack: 4 GB of RAM and 2 CPU cores; the recommended are 16 GB of RAM and 8 CPU cores.

Table 14: Distributed deployment hardware deployment

Component	Minimum		Recommended	
	RAM (GB)	CPU (cores)	RAM (GB)	CPU (cores)
Wazuh server	2	2	8	4
Elastic Stack	4	2	16	8

Disk space in the two deployment requirements depends on the alerts per second generated.

III.5.2.3 ELK Stack and Wazuh versions compatibility

The last version of Elastic is 8.2, and this version is not compatible with the previous version of Wazuh, which is 4.3. The following table represents Elastic Stack versions that are compatible with the Wazuh manager 4.3.1 using the Wazuh Kibana plugin:

Table 15: ELK Stack versions compatibility

Elastic Stack version
7.10.2
7.16.0–7.16.3
7.17.0–7.17.3

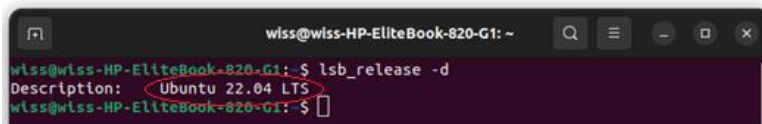
III.5.3 Lab environment

III.5.3.1 Lab environment development

This section presents the development environment, consisting of the hardware and the software: The centralization server and Test servers. Since the organization data, especially security data, are confidential, our work will be done under our helpful software. For our implementation, we choose All-in-one deployment because it doesn't require many hardware sources.

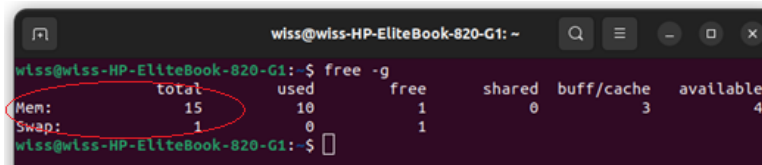
a) **Centralisation server:**

- Operating system:



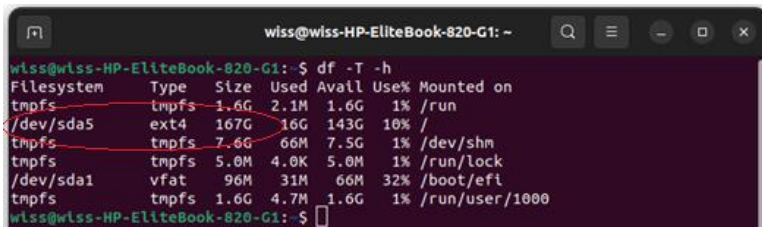
```
wlss@wlss-HP-EliteBook-820-G1: ~  
wlss@wlss-HP-EliteBook-820-G1:~$ lsb_release -d  
Description: Ubuntu 22.04 LTS  
wlss@wlss-HP-EliteBook-820-G1:~$
```

- 16GB RAM



```
wlss@wlss-HP-EliteBook-820-G1: ~  
wlss@wlss-HP-EliteBook-820-G1:~$ free -g  
              total        used         free   shared  buff/cache   available  
Mem:           15            10             1         0          3           4  
Swap:           1             0             1  
wlss@wlss-HP-EliteBook-820-G1:~$
```

- 167 GB of disk



```
wlss@wlss-HP-EliteBook-820-G1: ~  
wlss@wlss-HP-EliteBook-820-G1:~$ df -T -h  
Filesystem      Type      Size  Used Avail Use% Mounted on  
tmpfs           tmpfs     1.6G  2.1M  1.6G   1% /run  
/dev/sda5       ext4     167G  16G  143G  10% /  
tmpfs           tmpfs     7.6G   66M  7.5G   1% /dev/shm  
tmpfs           tmpfs     5.0M   4.0K  5.0M   1% /run/lock  
/dev/sda1       vfat     96M   31M   66M  32% /boot/efi  
tmpfs           tmpfs     1.6G  4.7M  1.6G   1% /run/user/1000  
wlss@wlss-HP-EliteBook-820-G1:~$
```

- Elasticsearch
- Logstash

- Kibana
- Wazuh 4.3.1

b) Test servers:

Server 1 : Ubuntu 20.04 TLS , RAM 4G

Server 2 : Windows 10 , RAM 4G

III.5.3.2 Lab environment architecture:

The architecture includes two wazuh agents: the Ubuntu agent allows the collection of logs to the processing tool, and the Windows agent captures security events and sends them to the wazuh manager. The wazuh manager sends the collected data to Filebeat, which transfers all the log lines to Logstash, the processing tool of Elasticsearch; it converts the received data into indexed data that will be stored in the BD (Elasticsearch).

To consult the logs and the event, the user connects to the visualization tool Kibana, retrieves the data from the comic, and exposes it. Kibana also could send alerts via the Slack platform that the employees communicate on it.

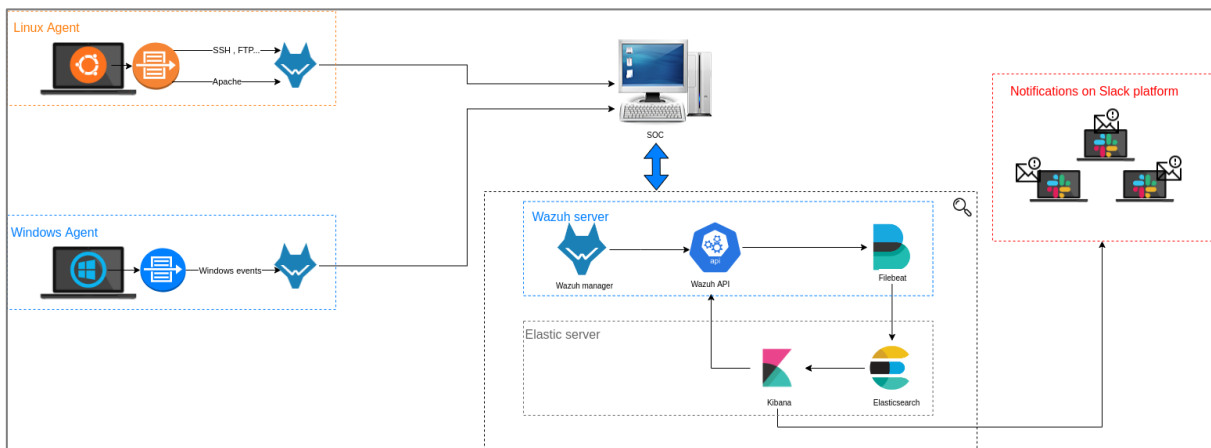


Figure 52: Lab environment architecture

III.5.4 Installation

III.5.4.1 Installing Elasticsearch

- *Installing the GPG key:*
- The Debian packages are all signed with the Elasticsearch Signing Key (PGP key D88E42B4)³.

³ <https://pgp.mit.edu>

```
root@w1ss-HP-EliteBook-820-G1:/home/w1ss# curl -s https://artifacts.elastic.co/GPG-KEY-elasticsearch | apt-key add -
Warning: apt-key is deprecated. Manage keyring files in trusted.gpg.d instead (see apt-key(8)).
OK
```

- ***Adding the Elastic stack repository repository :***

```
root@w1ss-HP-EliteBook-820-G1:/home/w1ss# echo "deb https://artifacts.elastic.co/packages/7.x/apt stable main" | tee /etc/apt/sources.list.d/elastic-7.x.list
deb https://artifacts.elastic.co/packages/7.x/apt stable main
root@w1ss-HP-EliteBook-820-G1:/home/w1ss# apt-get update
Hit:1 http://dz.archive.ubuntu.com/ubuntu jammy InRelease
Get:2 http://security.ubuntu.com/ubuntu jammy-security InRelease [110 kB]
Get:3 http://dz.archive.ubuntu.com/ubuntu jammy-updates InRelease [109 kB]
Hit:4 https://dl.winehq.org/wine-builds/debian bullseye InRelease
Get:5 https://artifacts.elastic.co/packages/7.x/apt stable InRelease [13.7 kB]
Hit:6 https://dl.google.com/linux/chrome/deb stable InRelease
Get:7 https://artifacts.elastic.co/packages/7.x/apt stable/main i386 Packages [73.8 kB]
Hit:8 http://dz.archive.ubuntu.com/ubuntu jammy-backports InRelease
Get:9 https://artifacts.elastic.co/packages/7.x/apt stable/main amd64 Packages [98.8 kB]
Fetched 406 kB in 2s (204 kB/s)
Reading package lists... Done
W: https://artifacts.elastic.co/packages/7.x/apt/dists/stable/InRelease: Key is stored in legacy trusted.gpg keyring (/etc/apt/trusted.gpg), see the DEPRECATION section in apt-key(8) for details.
```

- ***Installing the Elasticsearch package :***

```
root@w1ss-HP-EliteBook-820-G1:/home/w1ss# apt-get install elasticsearch=7.17.3
+ apt-get install elasticsearch=7.17.3
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following NEW packages will be installed:
  elasticsearch
0 upgraded, 1 newly installed, 0 to remove and 0 not upgraded.
Need to get 0 B/312 MB of archives.
After this operation, 518 MB of additional disk space will be used.
Selecting previously unselected package elasticsearch.
(Reading database ... 197241 files and directories currently installed.)
Preparing to unpack ../elasticsearch_7.17.3_amd64.deb ...
Creating elasticsearch group... OK
Creating elasticsearch user... OK
Unpacking elasticsearch (7.17.3) ...
Setting up elasticsearch (7.17.3) ...
### NOT starting on installation, please execute the following statements to configure elasticsearch service to start automatically using syst
end
  sudo systemctl daemon-reload
  sudo systemctl enable elasticsearch.service
### You can start elasticsearch service by executing
  sudo systemctl start elasticsearch.service
Created elasticsearch keystore in /etc/elasticsearch/elasticsearch.keystore
root@w1ss-HP-EliteBook-820-G1:/home/w1ss#
```

- ***Downloading the configuration file /etc/elasticsearch/elasticsearch.yml :***

```
root@w1ss-HP-EliteBook-820-G1:/home/w1ss# curl -s -o /etc/elasticsearch/elasticsearch.yml https://packages.wazuh.com/4.3/tpl/elastic-basic/elasticsearch_all_in_one.yml
+ curl -s -o /etc/elasticsearch/elasticsearch.yml https://packages.wazuh.com/4.3/tpl/elastic-basic/elasticsearch_all_in_one.yml
```

- ***Downloading the configuration file for creating the certificates:***

```
root@w1ss-HP-EliteBook-820-G1:/home/w1ss# curl -s -o /usr/share/elasticsearch/instances.yml https://packages.wazuh.com/4.3/tpl/elastic-basic/instances_aito.yml
+ curl -s -o /usr/share/elasticsearch/instances.yml https://packages.wazuh.com/4.3/tpl/elastic-basic/instances_aito.yml
```

- ***Generating credentials for all the Elastic Stack pre-built roles and users:***

The `elasticsearch-setup-passwords` command sets the passwords for the built-in users.

```

root@w1ss-HP-EliteBook-820-G1:/home/w1ss# /usr/share/elasticsearch/bin/elasticsearch-setup-passwords auto
+ /usr/share/elasticsearch/bin/elasticsearch-setup-passwords auto
Initiating the setup of passwords for reserved users elastic,apm_system,kibana,kibana_system,logstash_system,beats_system,remote_monitoring_us
er.
The passwords will be randomly generated and printed to the console.
Please confirm that you would like to continue [y/N]y

Changed password for user apm_system
PASSWORD apm_system = 4rwuEsASS2nKdHsfanS3

Changed password for user kibana_system
PASSWORD kibana_system = u45JywCuUHSbTbSDPcz9

Changed password for user kibana
PASSWORD kibana = u45JywCuUHSbTbSDPcz9

Changed password for user logstash_system
PASSWORD logstash_system = 2RbQPmM9xEzF1wLntg86

Changed password for user beats_system
PASSWORD beats_system = m4ZPClTkEft89jjsNmM

Changed password for user remote_monitoring_user
PASSWORD remote_monitoring_user = TrODm0ozUggy2EtoJ5RG

Changed password for user elastic
PASSWORD elastic = IAGhFpyHLThIq9XxChXH

root@w1ss-HP-EliteBook-820-G1:/home/w1ss#

```

- The command `curl -XGET https://localhost:9200 -u elastic:<elastic_password> -k` verifies if the installation was successfully made.

Running the command by replacing `elastic_password` with the password generated in the previous step for *the elastic* user:

```

root@w1ss-HP-EliteBook-820-G1:/home/w1ss# curl -XGET https://localhost:9200 -u elastic:IAGhFpyHLThIq9XxChXH -k
+ curl -XGET https://localhost:9200 -u elastic:IAGhFpyHLThIq9XxChXH -k
{
  "name": "elasticsearch",
  "cluster_name": "elasticsearch",
  "cluster_uuid": "m4bMANM2Sne9S9r8K7sIEKA",
  "version": {
    "number": "7.17.3",
    "build_flavor": "default",
    "build_type": "deb",
    "build_hash": "5ad023604c8d7416c9eb6c0eadb62b14e766caff",
    "build_date": "2022-04-19T08:11:19.070913226Z",
    "build_snapshot": false,
    "lucene_version": "8.11.1",
    "minimum_wire_compatibility_version": "6.8.0",
    "minimum_index_compatibility_version": "6.0.0-beta1"
  },
  "tagline": "You Know, for Search"
}
root@w1ss-HP-EliteBook-820-G1:/home/w1ss#

```

III.5.4.2 Installing wazuh server

- *Installing the GPG key:*

```

root@w1ss-HP-EliteBook-820-G1:/home/w1ss# curl -s https://packages.wazuh.com/key/GPG-KEY-WAZUH | apt-key add -
+ curl -s https://packages.wazuh.com/key/GPG-KEY-WAZUH | apt-key add -
+ apt-key add -
Warning: apt-key is deprecated. Manage keyring files in trusted.gpg.d instead (see apt-key(8)).
OK

```

- *Adding the Wazuh repository:*

```

root@w1ss-HP-EliteBook-820-G1:/home/w1ss# curl -s https://packages.wazuh.com/key/GPG-KEY-WAZUH | apt-key add -
+ curl -s https://packages.wazuh.com/key/GPG-KEY-WAZUH | apt-key add -
+ apt-key add -
Warning: apt-key is deprecated. Manage keyring files in trusted.gpg.d instead (see apt-key(8)).
OK
root@w1ss-HP-EliteBook-820-G1:/home/w1ss# echo "deb https://packages.wazuh.com/4.x/apt/ stable main" | tee -a /etc/apt/sources.list.d/wazuh.l
ist
+ tee -a /etc/apt/sources.list.d/wazuh.list
+ echo 'deb https://packages.wazuh.com/4.x/apt/ stable main'
deb https://packages.wazuh.com/4.x/apt/ stable main

```


III.5.4.3 Installing the Wazuh manager

- *Installing the Wazuh manager package:*

```
root@wiss-HP-EliteBook-820-G1:/home/wiss# apt-get install wazuh-manager
+ apt-get install wazuh-manager
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
Suggested packages:
  expect
The following NEW packages will be installed:
  wazuh-manager
0 upgraded, 1 newly installed, 0 to remove and 0 not upgraded.
Need to get 119 MB of archives.
After this operation, 455 MB of additional disk space will be used.
Get:1 https://packages.wazuh.com/4.x/apt/stable/main amd64 wazuh-manager amd64 4.3.0-1 [119 MB]
Fetched 119 MB in 2min 35s (766 kB/s)
Selecting previously unselected package wazuh-manager.
(Reading database ... 198367 files and directories currently installed.)
Preparing to unpack .../wazuh-manager_4.3.0-1_amd64.deb ...
Unpacking wazuh-manager (4.3.0-1) ...
Setting up wazuh-manager (4.3.0-1) ...
root@wiss-HP-EliteBook-820-G1:/home/wiss#
```

- *Enable and start the Wazuh manager service:*

```
root@wiss-HP-EliteBook-820-G1:/home/wiss# systemctl daemon-reload
+ systemctl daemon-reload
root@wiss-HP-EliteBook-820-G1:/home/wiss# systemctl enable wazuh-manager
+ systemctl enable wazuh-manager
Synchronizing state of wazuh-manager.service with SysV service script with /lib/systemd/systemd-sysv-install.
Executing: /lib/systemd/systemd-sysv-install enable wazuh-manager
Created symlink /etc/systemd/system/multi-user.target.wants/wazuh-manager.service → /lib/systemd/system/wazuh-manager.service.
root@wiss-HP-EliteBook-820-G1:/home/wiss# systemctl start wazuh-manager
+ systemctl start wazuh-manager
```

- *Checking if the Wazuh manager is active :*

```
root@wiss-HP-EliteBook-820-G1:/home/wiss# systemctl status wazuh-manager
+ systemctl status wazuh-manager
● wazuh-manager.service - Wazuh manager
   Loaded: loaded (/lib/systemd/system/wazuh-manager.service; enabled; vendor preset: enabled)
   Active: active (running) since Tue 2022-05-17 14:55:38 CET; 16s ago
     Process: 40255 ExecStart=/usr/bin/env /var/ossec/bin/wazuh-control start (code=exited, status=0/SUCCESS)
    Tasks: 133 (limit: 18432)
   Memory: 656.2M
      CPU: 34.265s
   CGroup: /system.slice/wazuh-manager.service
           └─48415 /var/ossec/framework/python/bin/python3 /var/ossec/api/scripts/wazuh-apid.py
           └─48454 /var/ossec/bin/wazuh-authd
           └─48467 /var/ossec/bin/wazuh-db
           └─48490 /var/ossec/bin/wazuh-execd
           └─48501 /var/ossec/bin/wazuh-analysisd
           └─48510 /var/ossec/bin/wazuh-syscheckd
           └─48522 /var/ossec/bin/wazuh-remoted
           └─48555 /var/ossec/bin/wazuh-logcollector
           └─48565 /var/ossec/bin/wazuh-monitord
           └─48574 /var/ossec/bin/wazuh-modulesd
           └─48722 /var/ossec/framework/python/bin/python3 /var/ossec/api/scripts/wazuh-apid.py
           └─48725 /var/ossec/framework/python/bin/python3 /var/ossec/api/scripts/wazuh-apid.py

May 17 14:55:35 wiss-HP-EliteBook-820-G1 env[48355]: Started wazuh-db...
May 17 14:55:35 wiss-HP-EliteBook-820-G1 env[48355]: Started wazuh-execd...
May 17 14:55:35 wiss-HP-EliteBook-820-G1 env[48355]: Started wazuh-analysisd...
May 17 14:55:35 wiss-HP-EliteBook-820-G1 env[48355]: Started wazuh-syscheckd...
May 17 14:55:35 wiss-HP-EliteBook-820-G1 env[48355]: Started wazuh-remoted...
May 17 14:55:36 wiss-HP-EliteBook-820-G1 env[48355]: Started wazuh-logcollector...
May 17 14:55:36 wiss-HP-EliteBook-820-G1 env[48355]: Started wazuh-monitord...
May 17 14:55:36 wiss-HP-EliteBook-820-G1 env[48355]: Started wazuh-modulesd...
May 17 14:55:38 wiss-HP-EliteBook-820-G1 env[48355]: Completed.
May 17 14:55:38 wiss-HP-EliteBook-820-G1 systemd[1]: Started Wazuh manager.
root@wiss-HP-EliteBook-820-G1:/home/wiss#
```


III.5.4.4 Installing Filebeat

- *Installing the Filebeat package :*

```
root@w1ss-HP-ElliteBook-820-G1:/home/w1ss# apt-get install filebeat=7.17.3
+ apt-get install filebeat=7.17.3
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following NEW packages will be installed:
  filebeat
0 upgraded, 1 newly installed, 0 to remove and 0 not upgraded.
Need to get 36.3 MB of archives.
After this operation, 144 MB of additional disk space will be used.
Get:1 https://artifacts.elastic.co/packages/7.x/apt/stable/main/amd64/filebeat amd64 7.17.3 [36.3 MB]
Fetched 36.3 MB in 56s (651 kB/s)
Selecting previously unselected package filebeat.
(Reading database ... 217038 files and directories currently installed.)
Preparing to unpack .../filebeat_7.17.3_amd64.deb ...
Unpacking filebeat (7.17.3) ...
Setting up filebeat (7.17.3) ...
```

- *Downloading the pre-configured Filebeat config file used to forward Wazuh alerts to Elasticsearch:*

```
root@w1ss-HP-ElliteBook-820-G1:/home/w1ss# curl -s /etc/filebeat/filebeat.yml https://packages.wazuh.com/4.3/tpl/elastic-basic/filebeat_all_in_one.yml
+ curl -s /etc/filebeat/filebeat.yml https://packages.wazuh.com/4.3/tpl/elastic-basic/filebeat_all_in_one.yml
```

- *Downloading the alerts template for Elasticsearch:*

```
root@w1ss-HP-ElliteBook-820-G1:/home/w1ss# curl -s /etc/filebeat/wazuh-template.json https://raw.githubusercontent.com/wazuh/wazuh/4.3/extensions/elasticsearch/7.x/wazuh-template.json
+ curl -s /etc/filebeat/wazuh-template.json https://raw.githubusercontent.com/wazuh/wazuh/4.3/extensions/elasticsearch/7.x/wazuh-template.json
root@w1ss-HP-ElliteBook-820-G1:/home/w1ss# chmod go+r /etc/filebeat/wazuh-template.json
+ chmod go+r /etc/filebeat/wazuh-template.json
```

- *Downloading the Wazuh module for Filebeat:*

```
root@w1ss-HP-ElliteBook-820-G1:/home/w1ss# curl -s https://packages.wazuh.com/4.x/filebeat/wazuh-filebeat-0.1.tar.gz | tar -xvz -C /usr/share/filebeat/module
+ curl -s https://packages.wazuh.com/4.x/filebeat/wazuh-filebeat-0.1.tar.gz
+ tar -xvz -C /usr/share/filebeat/module
wazuh/
wazuh/module.yml
wazuh/archives/
wazuh/archives/config/
wazuh/archives/config/archives.yml
wazuh/archives/ingest/
wazuh/archives/ingest/pipeline.json
wazuh/archives/manifest.yml
wazuh/alerts/
wazuh/alerts/config/
wazuh/alerts/config/alerts.yml
wazuh/alerts/ingest/
wazuh/alerts/ingest/pipeline.json
wazuh/alerts/manifest.yml
wazuh/_meta/
wazuh/_meta/config.yml
wazuh/_meta/fields.yml
wazuh/_meta/docs.asciidoc
root@w1ss-HP-ElliteBook-820-G1:/home/w1ss#
```

- *Editing the configuration file of filebeat /etc/filebeat/filebeat.yml*

```
root@w1ss-HP-ElliteBook-820-G1:/home/w1ss# nano /etc/filebeat/filebeat.yml
+ nano /etc/filebeat/filebeat.yml
root@w1ss-HP-ElliteBook-820-G1:/home/w1ss#
```

- *Replacing `elasticsearch` `password` in `output.elasticsearch.password:<elasticsearch_password>` with the previously generated password for the elastic user*

```

root@wiss-HP-EliteBook-820-G1: /home/wiss
GNU nano 6.2 /etc/filebeat/filebeat.yml *
# wazuh - Filebeat configuration file
output.elasticsearch.hosts: ["127.0.0.1:9200"]
output.elasticsearch.password: IAGhFpyHlThI9XxChXH

filebeat.modules:
- module: wazuh
  alerts:
    enabled: true
  archives:
    enabled: false

setup.template.json.enabled: true
setup.template.json.path: /etc/filebeat/wazuh-template.json
setup.template.json.name: wazuh
setup.template.overwrite: true
setup.iln.enabled: false

output.elasticsearch.protocol: https
output.elasticsearch.ssl.certificate: /etc/elasticsearch/certs/elasticsearch.crt
output.elasticsearch.ssl.key: /etc/elasticsearch/certs/elasticsearch.key
output.elasticsearch.ssl.certificate_authorities: /etc/elasticsearch/certs/ca/ca.crt
output.elasticsearch.ssl.verification_mode: strict
output.elasticsearch.username: elastic

```

III.5.4.5 Installing Kibana

- *Installing the Kibana package :*

```

root@wiss-HP-EliteBook-820-G1: /home/wiss
root@wiss-HP-EliteBook-820-G1:/home/wiss# apt-get install kibana=7.17.3
+ apt-get install kibana=7.17.3
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following NEW packages will be installed:
  kibana
0 upgraded, 1 newly installed, 0 to remove and 0 not upgraded.
Need to get 267 MB of archives.
After this operation, 677 MB of additional disk space will be used.
Get:1 https://artifacts.elastic.co/packages/7.x/apt/stable/main amd64 kibana amd64 7.17.3 [267 MB]
Fetched 267 MB in 6min 35s (676 kB/s)
Selecting previously unselected package kibana.
(Reading database ... 219114 files and directories currently installed.)
Preparing to unpack .../kibana_7.17.3_amd64.deb ...
Unpacking kibana (7.17.3) ...
Setting up kibana (7.17.3) ...

```

- *Downloading the Kibana configuration file:*

```

Terminal
root@wiss-HP-EliteBook-820-G1: /home/wiss
root@wiss-HP-EliteBook-820-G1:/home/wiss# curl -s -o /etc/kibana/kibana.yml https://packages.wazuh.com/4.3/tpl/elastic-basic/kibana_all_in_one.yml
+ curl -s -o /etc/kibana/kibana.yml https://packages.wazuh.com/4.3/tpl/elastic-basic/kibana_all_in_one.yml
root@wiss-HP-EliteBook-820-G1:/home/wiss#

```

- *Editing the configuration file of kibana /etc/kibana/kibana.yml file:*

```

root@wiss-HP-EliteBook-820-G1: /home/wiss
root@wiss-HP-EliteBook-820-G1:/home/wiss# nano /etc/kibana/kibana.yml
+ nano /etc/kibana/kibana.yml
root@wiss-HP-EliteBook-820-G1:/home/wiss#

```

- *Replacing `elasticsearch` `password` in `output.elasticsearch.password:<elasticsearch_password>` with the previously generated password for the elastic user*

```
wiss@wiss-HP-EliteBook-820-G1: ~
GNU nano 6.2 /etc/kibana/kibana.yml
server.host: 192.168.1.38
server.port: 443
elasticsearch.hosts: https://localhost:9200
elasticsearch.password: IAGhFpyHlThlg9XxChXH

# Elasticsearch from/to Kibana
elasticsearch.ssl.certificateAuthorities: /etc/kibana/certs/ca/ca.crt
elasticsearch.ssl.certificate: /etc/kibana/certs/kibana.crt
elasticsearch.ssl.key: /etc/kibana/certs/kibana.key

# Browser from/to Kibana
server.ssl.enabled: true
server.ssl.certificate: /etc/kibana/certs/kibana.crt
server.ssl.key: /etc/kibana/certs/kibana.key

# Elasticsearch authentication
xpack.security.enabled: true
elasticsearch.username: elastic
uiSettings.overrides.defaultRoute: "/app/wazuh"
elasticsearch.ssl.verifcationMode: certificate
telemetry.banner: false
```

- **Creating the `/usr/share/kibana/data` directory:**

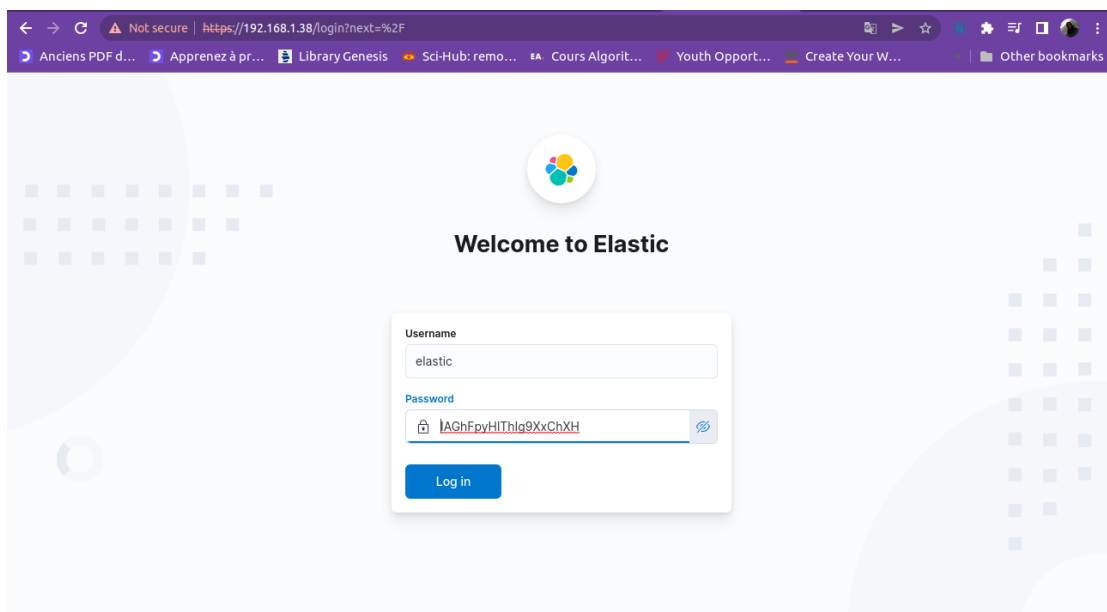
```
root@wiss-HP-EliteBook-820-G1: /home/wiss
root@wiss-HP-EliteBook-820-G1:/home/wiss# mkdir /usr/share/kibana/data
+ mkdir /usr/share/kibana/data
root@wiss-HP-EliteBook-820-G1:/home/wiss# chown -R kibana:kibana /usr/share/kibana
+ chown -R kibana:kibana /usr/share/kibana
root@wiss-HP-EliteBook-820-G1:/home/wiss#
```

- **Access the web interface using the password generated during the Elasticsearch installation process:**

URL: https://<wazuh_server_ip> (192.168.1.38)

user: elastic

password: <PASSWORD_elastic>



III.6 Wazuh user interface

III.6.1 Home page

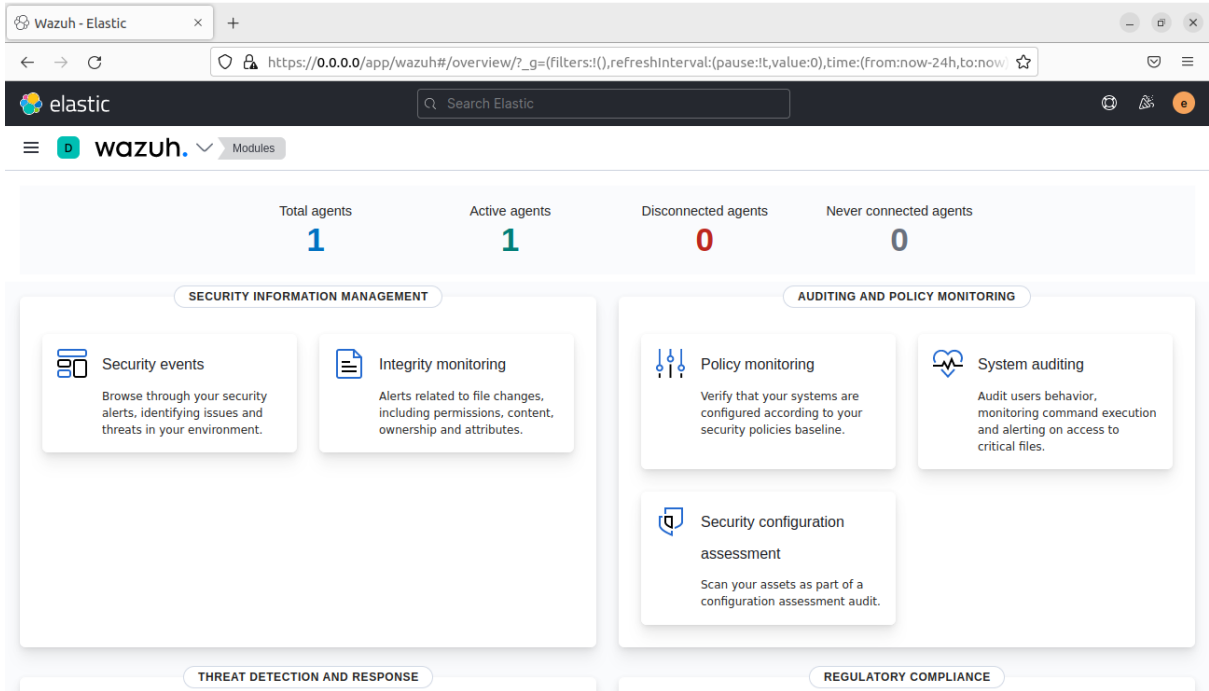


Figure 53: Wazuh home page

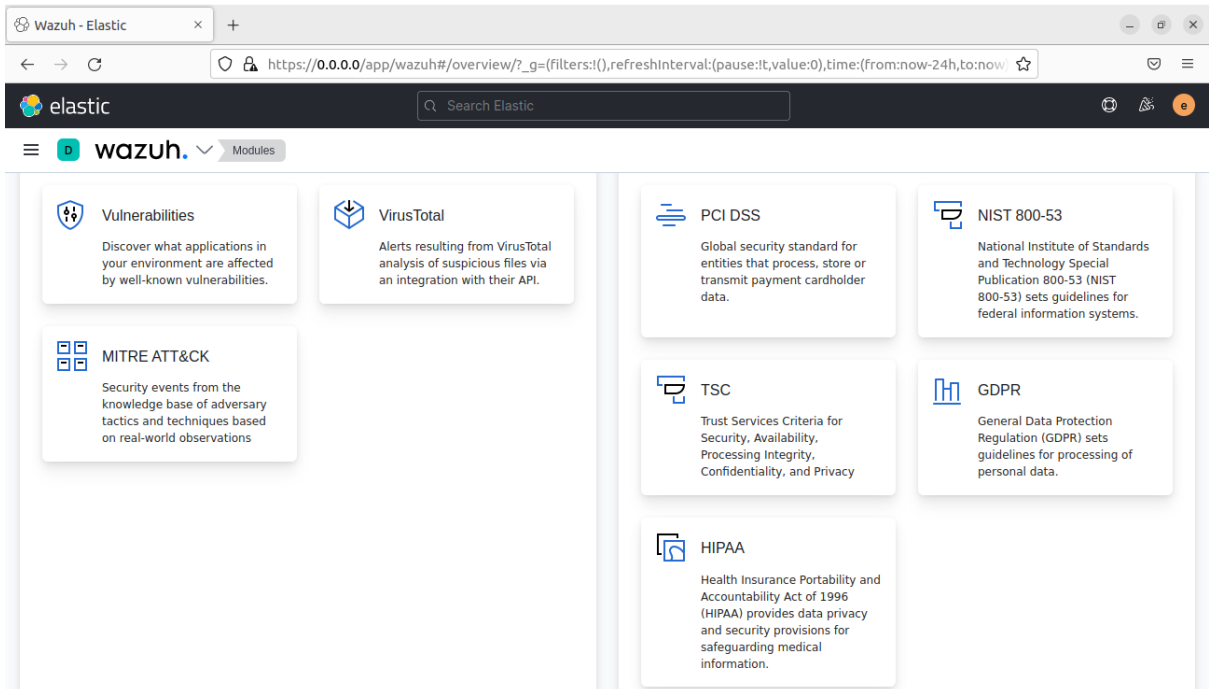


Figure 54: Wazuh home page 2

The figure above represents Wazuh's home page that provides shortcuts to the application modules (Threat detection and response, information management security, audit and policy monitoring, etc.).

III.6.2 Agents page

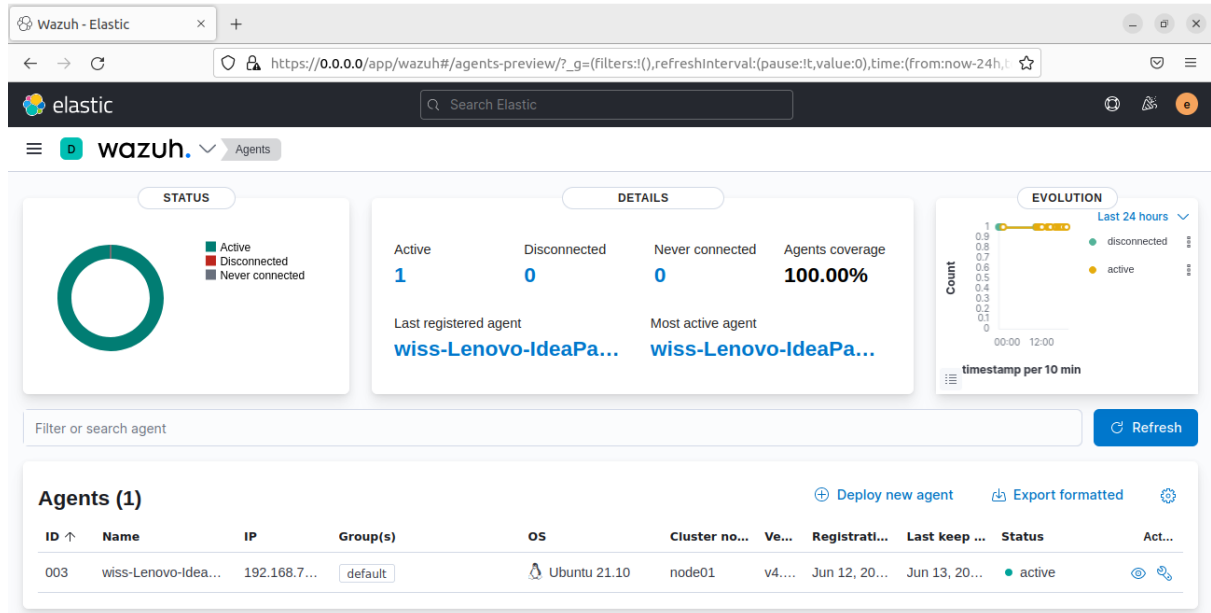


Figure 55: Wazuh agent page

The agent page gives the user all information about agents, for example, the agent status (Active, Disconnected, Never connected), its data, it can also deploy a new agent, etc.

To deploy a new agent, the user selects “Deploy new agent,” then a list of information needs fields, such as the type of operating system, Wazuh server address...

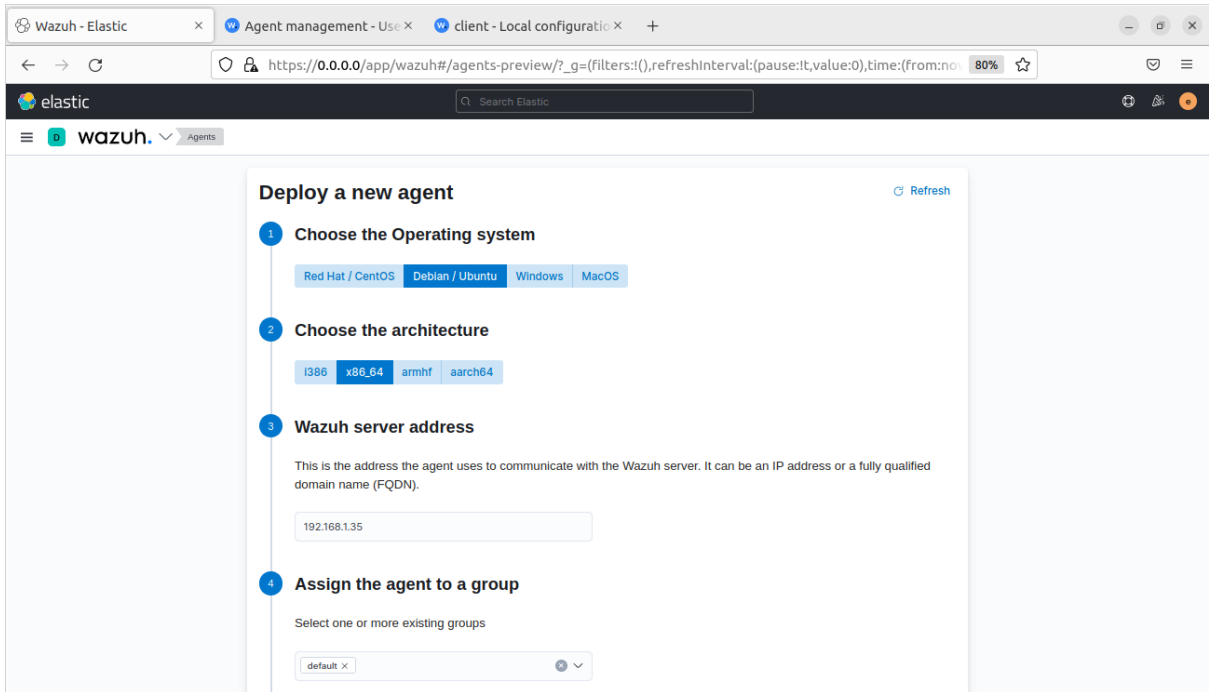


Figure 56: Deploying agent 1

According to the information filled, a downloading command is generated.

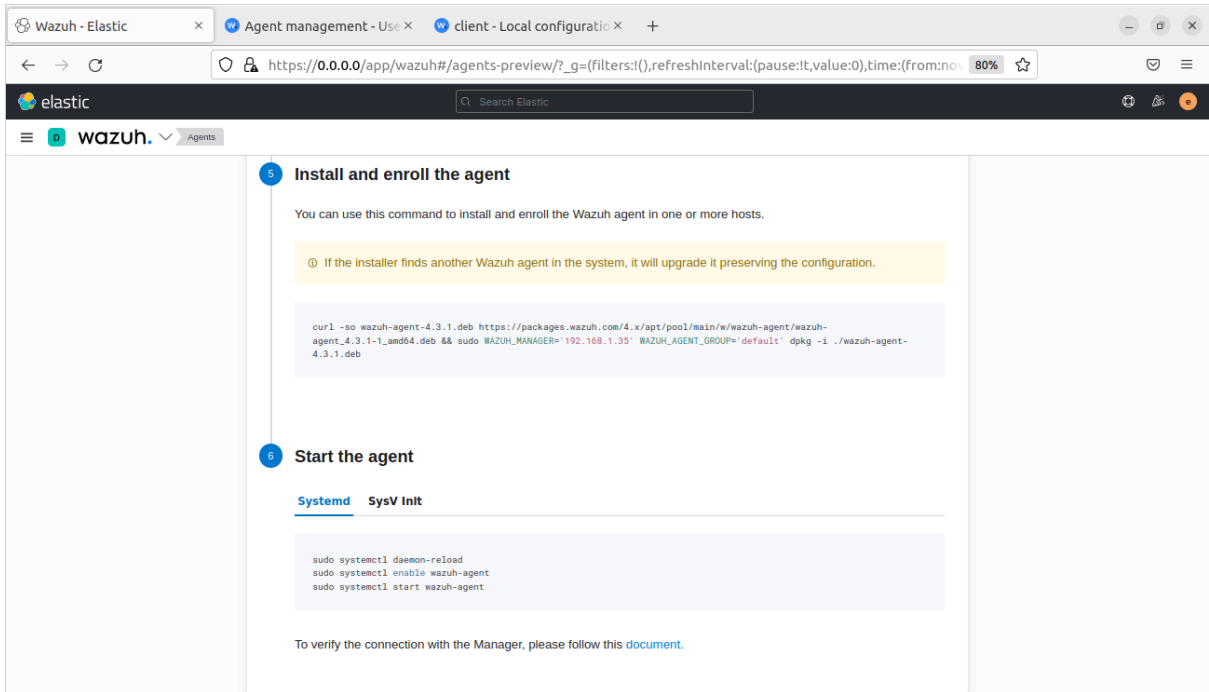
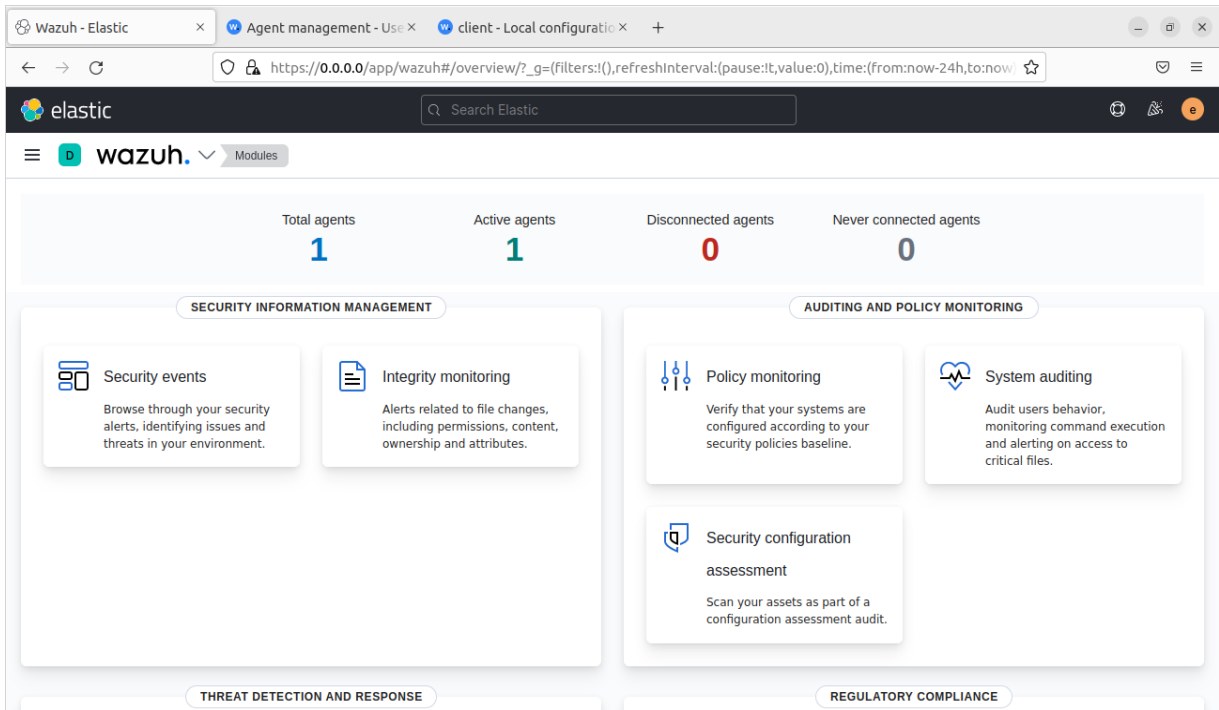


Figure 57: Deploying agent 2

We execute the command in the machine that we need to deploy this agent.

```
root@wiss-Lenovo-IdeaPad-S410p: /home/wiss# curl -so wazuh-agent-4.3.1.deb https://packages.wazuh.com/4.x/apt/pool/main/w/wazuh-agent/wazuh-agent-4.3.1-1_amd64.deb && sudo WAZUH_MANAGER='192.168.1.35' WAZUH_AGENT_GROUP='default' dpkg -i ./wazuh-agent-4.3.1.deb
Selecting previously unselected package wazuh-agent.
(Reading database ... 214178 files and directories currently installed.)
Preparing to unpack ./wazuh-agent-4.3.1.deb ...
Unpacking wazuh-agent (4.3.1-1) ...
Setting up wazuh-agent (4.3.1-1) ...
root@wiss-Lenovo-IdeaPad-S410p: /home/wiss# systemctl daemon-reload
root@wiss-Lenovo-IdeaPad-S410p: /home/wiss# systemctl enable wazuh-agent
Synchronizing state of wazuh-agent.service with SysV service script with /lib/systemd/systemd-sysv-install.
Executing: /lib/systemd/systemd-sysv-install enable wazuh-agent
Created symlink /etc/systemd/system/multi-user.target.wants/wazuh-agent.service -> /lib/systemd/system/wazuh-agent.service.
root@wiss-Lenovo-IdeaPad-S410p: /home/wiss# systemctl start wazuh-agent
root@wiss-Lenovo-IdeaPad-S410p: /home/wiss#
```

After restarting wazuh manager, the agent activates.



- Agent Dashboard

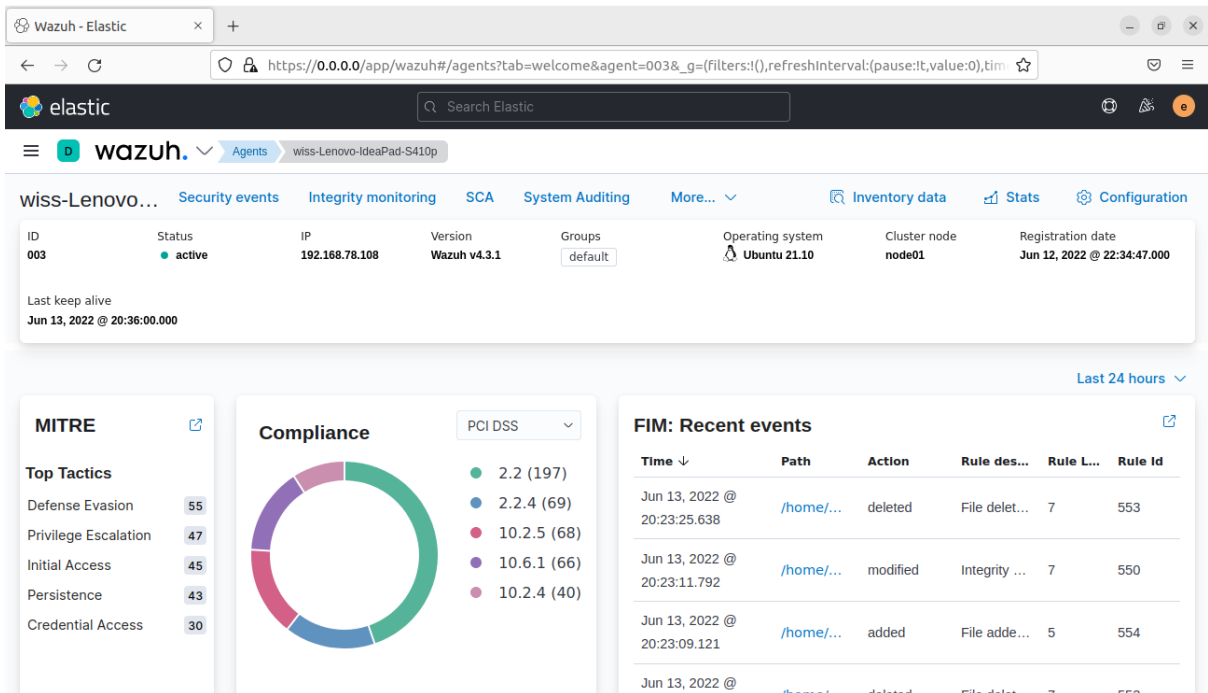


Figure 58: Agent dashboard 1

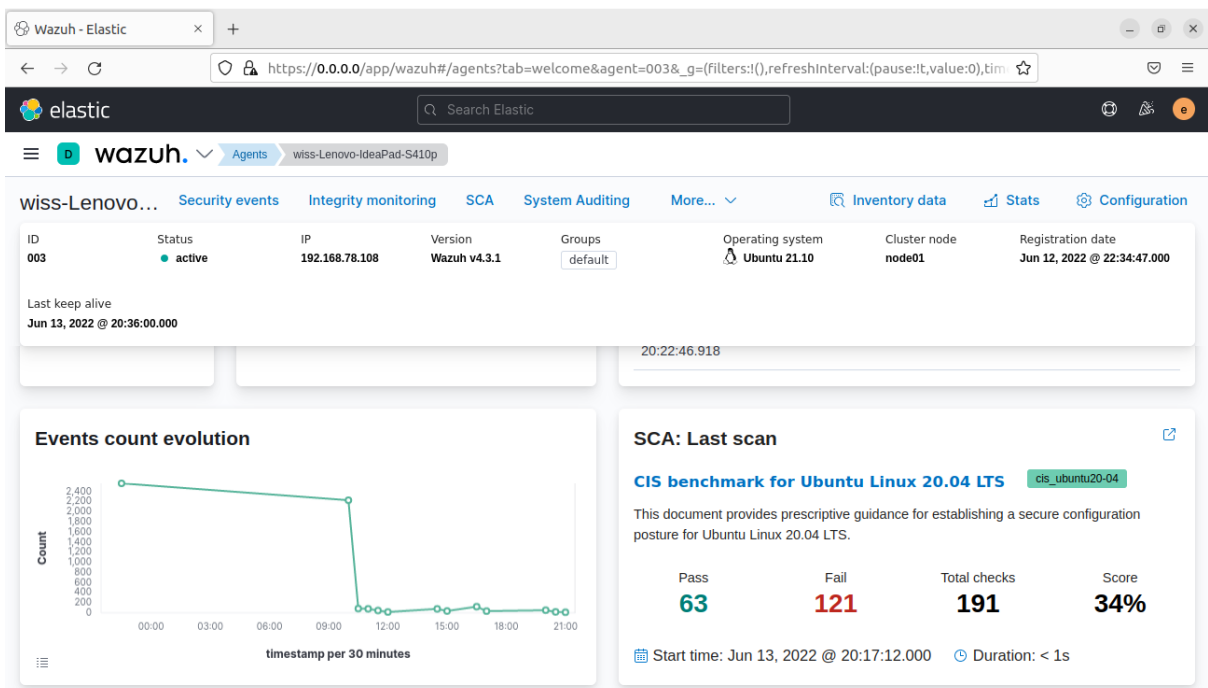


Figure 59: Agent dashboard 2

The agent dashboard provides detailed information about the selected agent, the status, the agent’s IP address, and the operating system... As well as visualizations for the application functionalities.

III.6.3 Security information management

III.6.3.1 Security events

This module allows the safety browsing through the security alerts and identifying issues and threats in the environment.

- **Dashboard**

From the dashboard of security events, we can see the total of the events in the last seven days (wazuh can display all the events that happen in real-time to a year later), also the number of authentication failures, authentications success, as well as the special event that have a 12 level or higher.

The events are also demonstrated in graphical types such as curves and graphic circles.

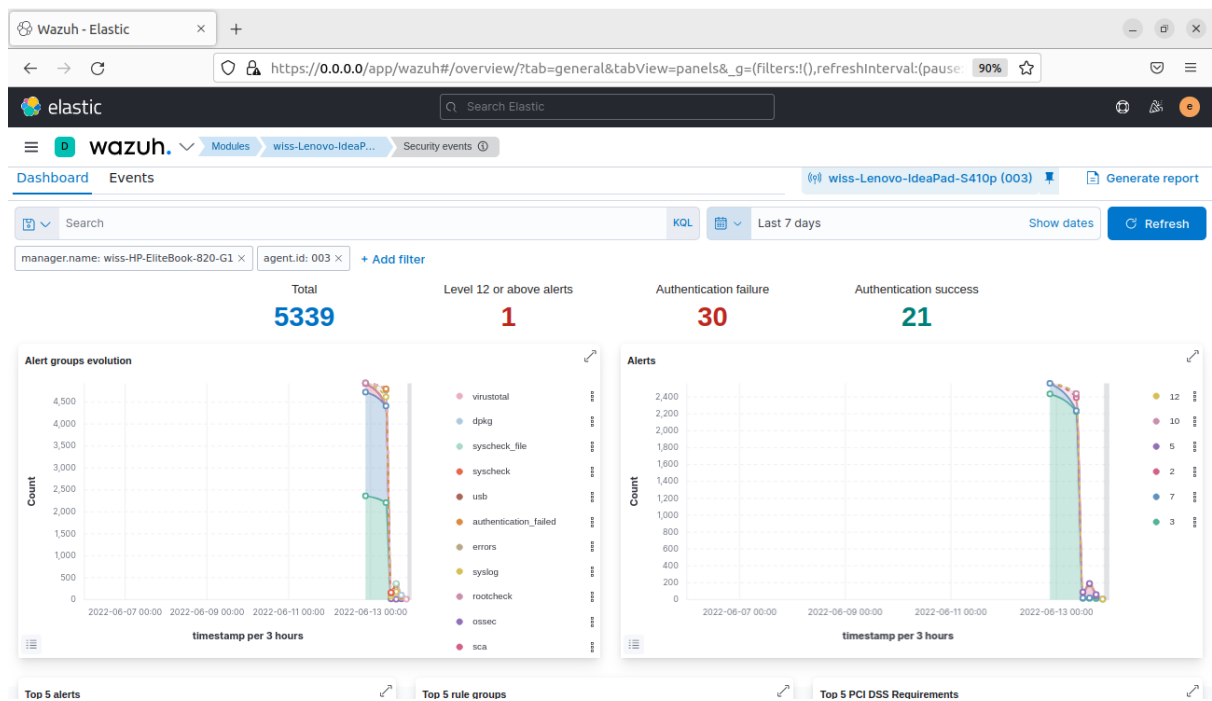
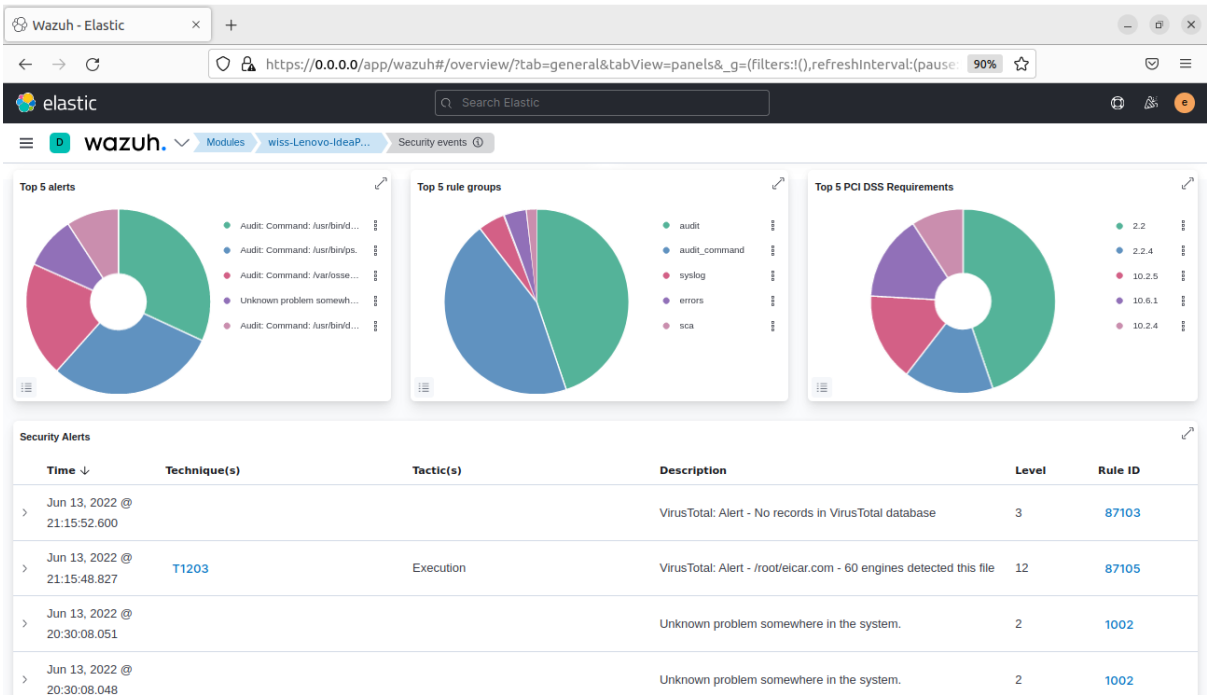


Figure 60: Security events dashboard



- **Events**

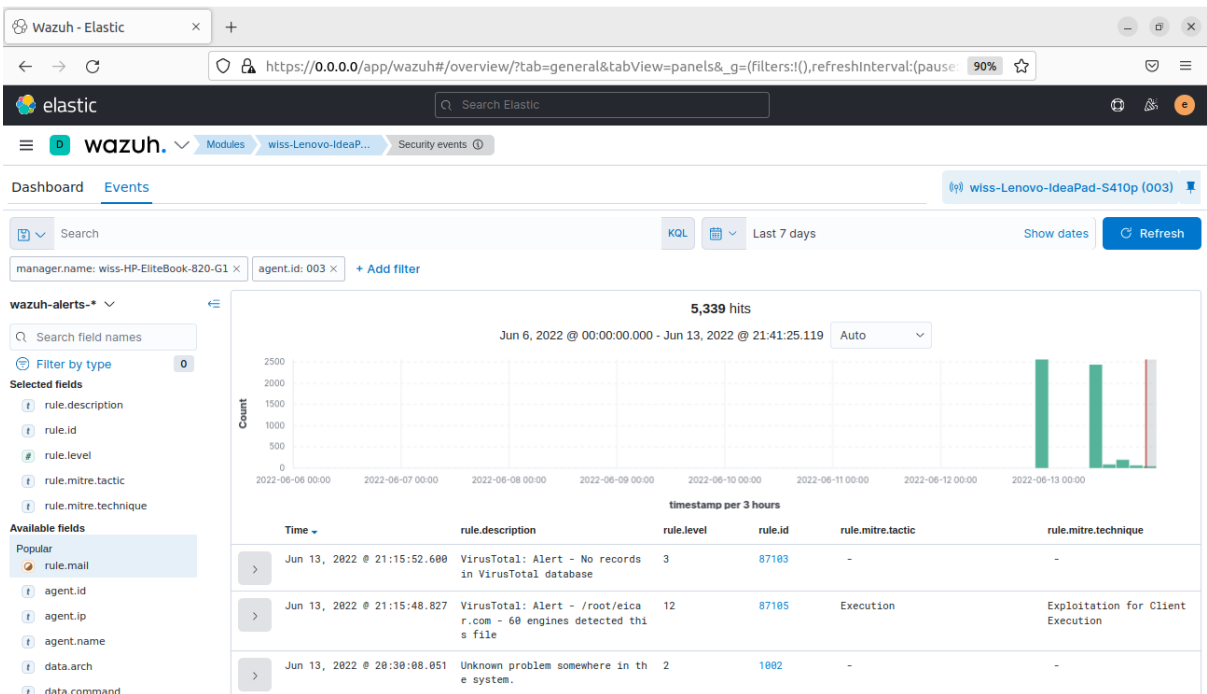


Figure 61: Security events, events

The figure above represents the list of security events “index: wazuh-alerts-*” the user can filter by type to display the necessary information.

III.6.3.2 Integrity monitoring

The Wazuh integrity monitoring module displays the alerts related to file changes, including permission, content, ownership, and attributes.

- **Dashboard**

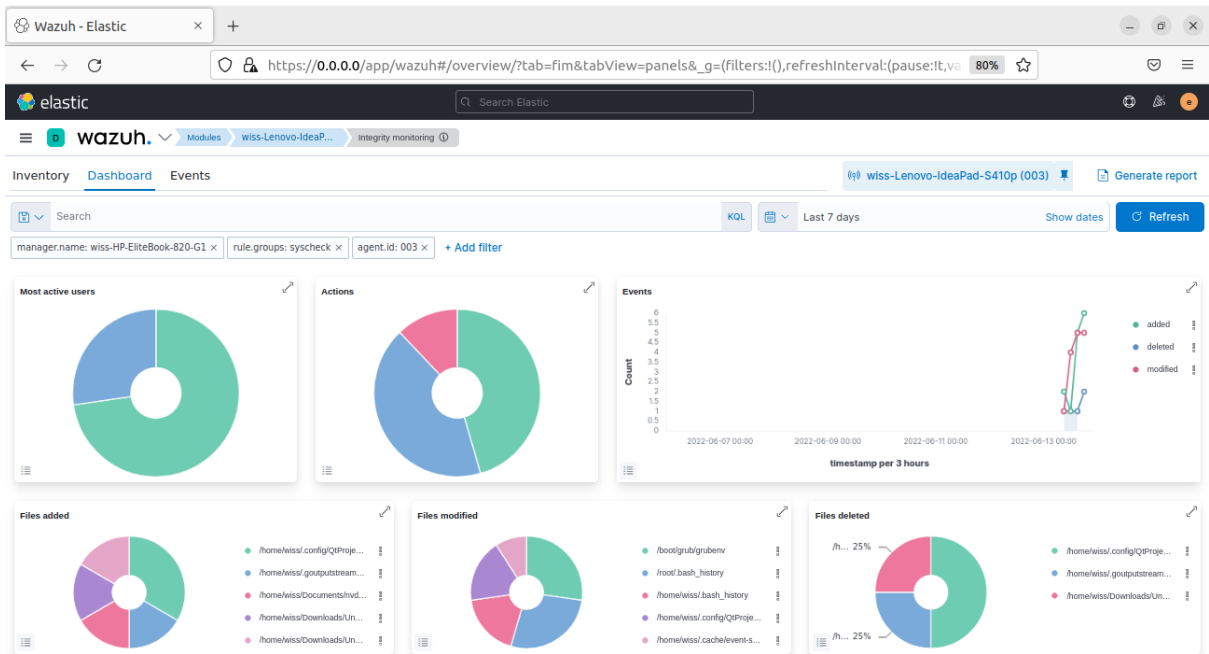


Figure 62: Integrity monitoring dashboard

- **Events**

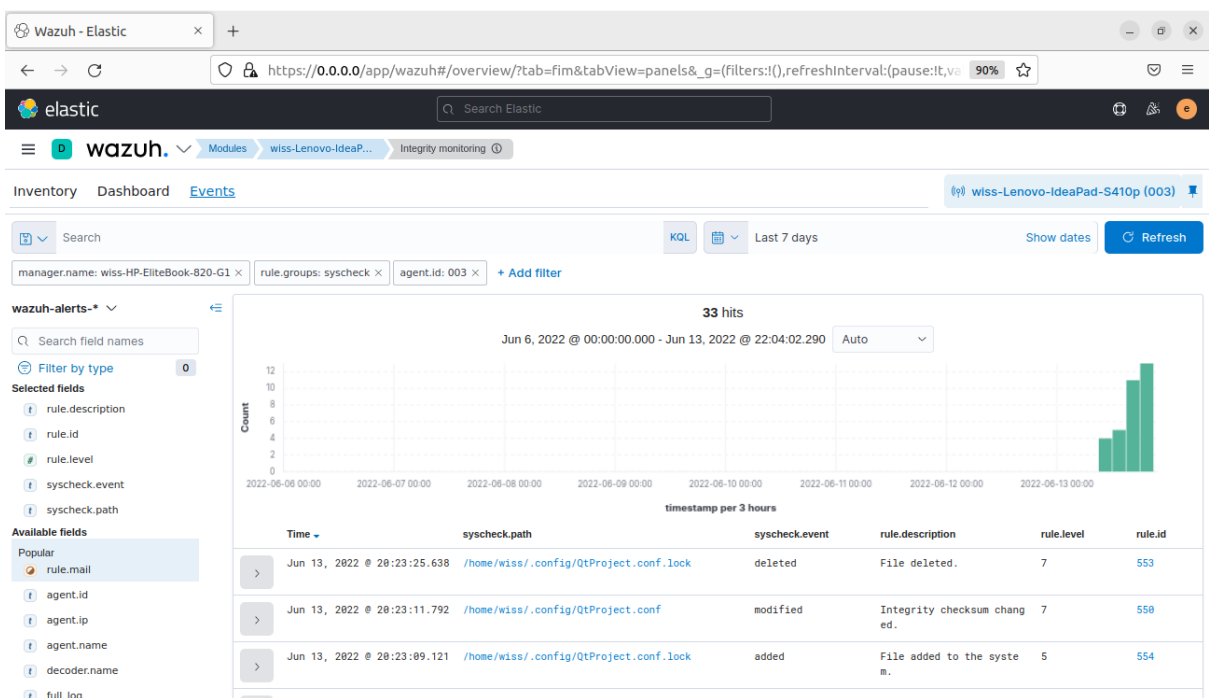


Figure 63: Integrity monitoring events

III.6.4 Auditing and Policy Monitoring

Auditing and Policy monitoring verifies that all the agent systems conform to predefined rules regarding configuration settings and approved application usage.

III.6.4.1 Policy monitoring:

The Policy monitoring module verifies that the systems are configured according to the security policies baseline.

- **Dashboard**

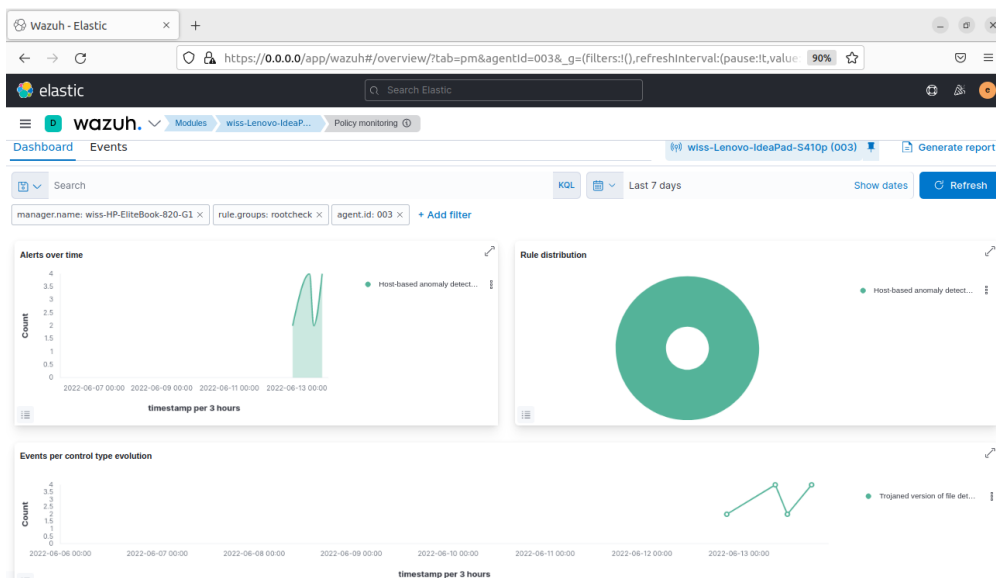


Figure 64: Policy monitoring dashboard

- **Events**

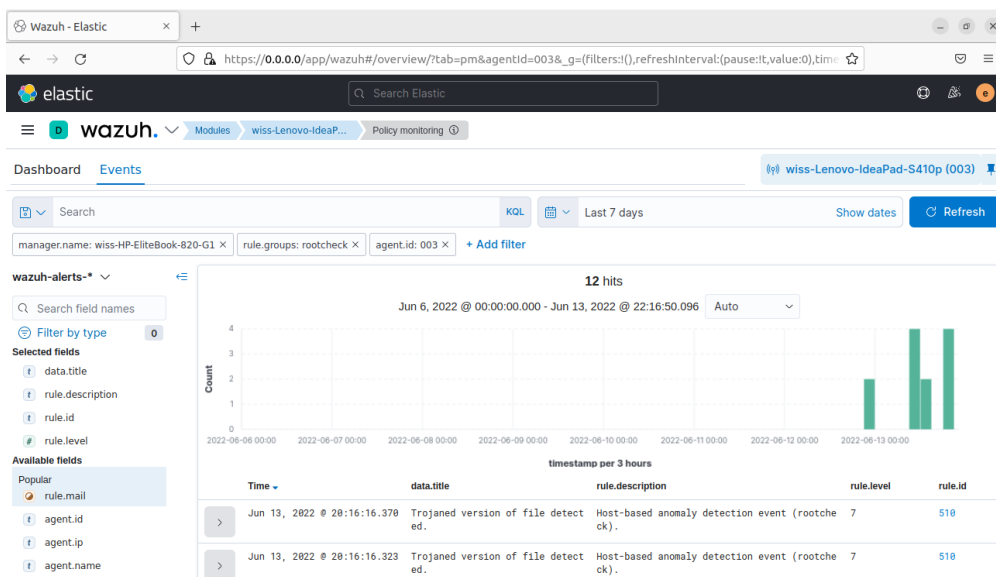


Figure 65: Policy monitoring events

III.6.4.2 System auditing

The System auditing module provides the audit users' behavior, monitoring the executing command and alerting on access to critical files.

- **Dashboard**

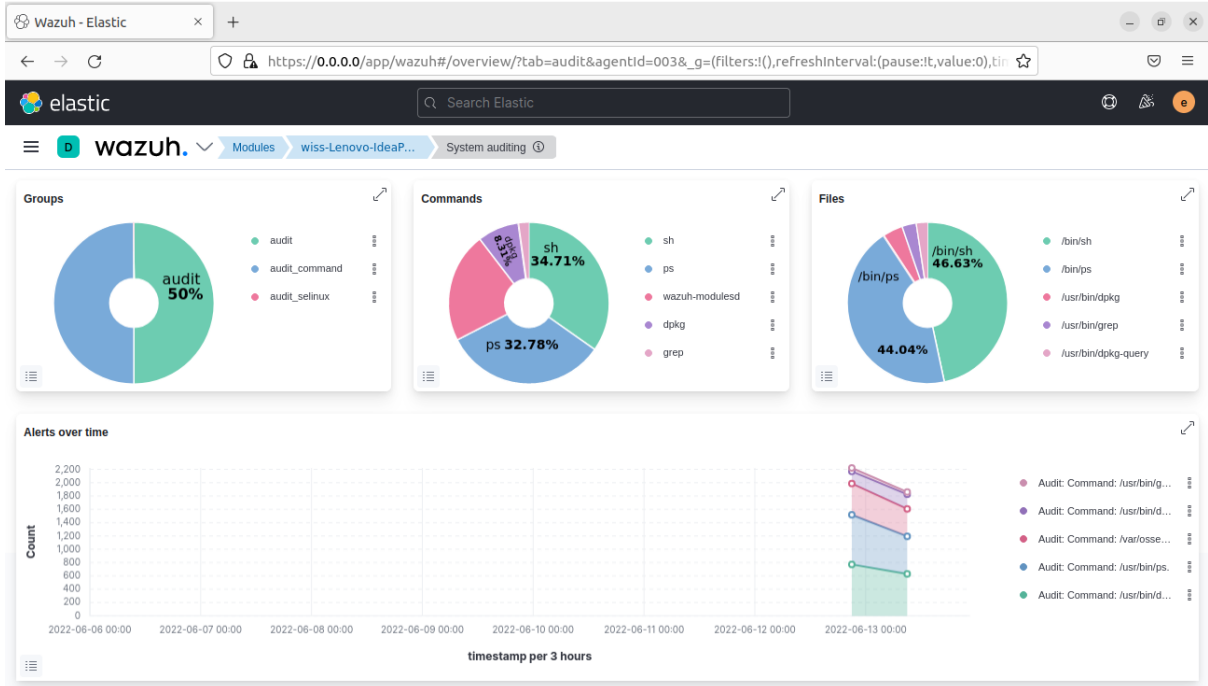


Figure 66: System auditing dashboard

- **Events**

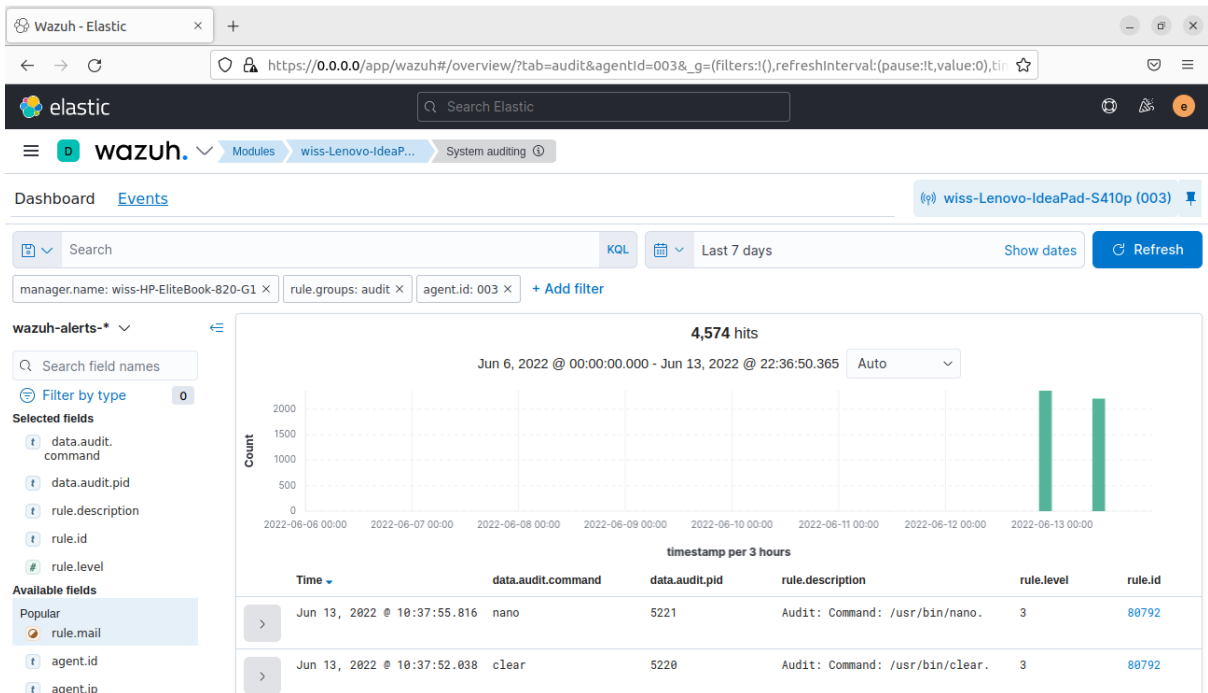


Figure 67: System auditing events

III.6.4.3 Security configuration assessment

The Security configuration assessment is a group of configuration checks; each check uses a rule or a set of rules to verify the state of a system. These rules can run custom commands, scan configuration files and review the running processes or Windows registry keys. (Security Configuration Assessment (SCA) · Wazuh · The Open Source Security Platform). This module loads the default policies specific to the agent operating system.

- **Inventory**

In this case, the OS of the agent is Ubuntu, so the module automatically loads the CIS benchmark for Ubuntu. Center for Internet Security is an entity whose mission is to identify, develop, validate, promote and support cyber defense best practices. CIS benchmarks are configuration references for security configuration systems. The security policies are kept by the Wazuh Manager, which distributes them to all agents.

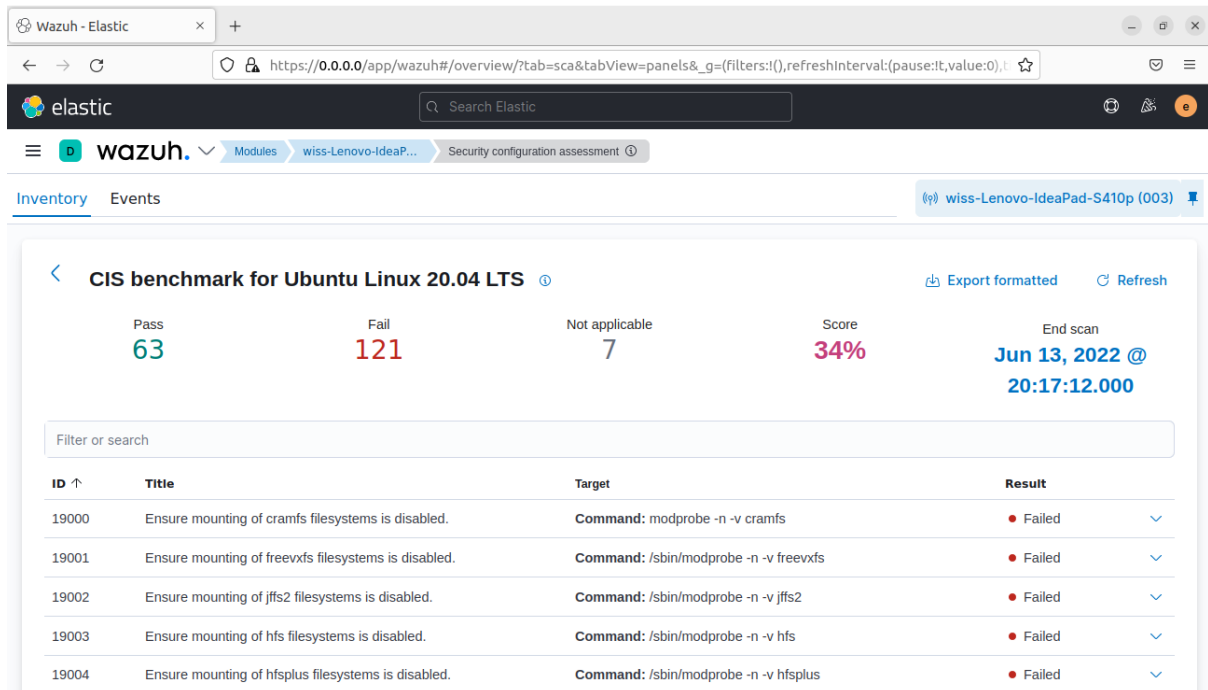


Figure 68: SCA inventory

- **Events**

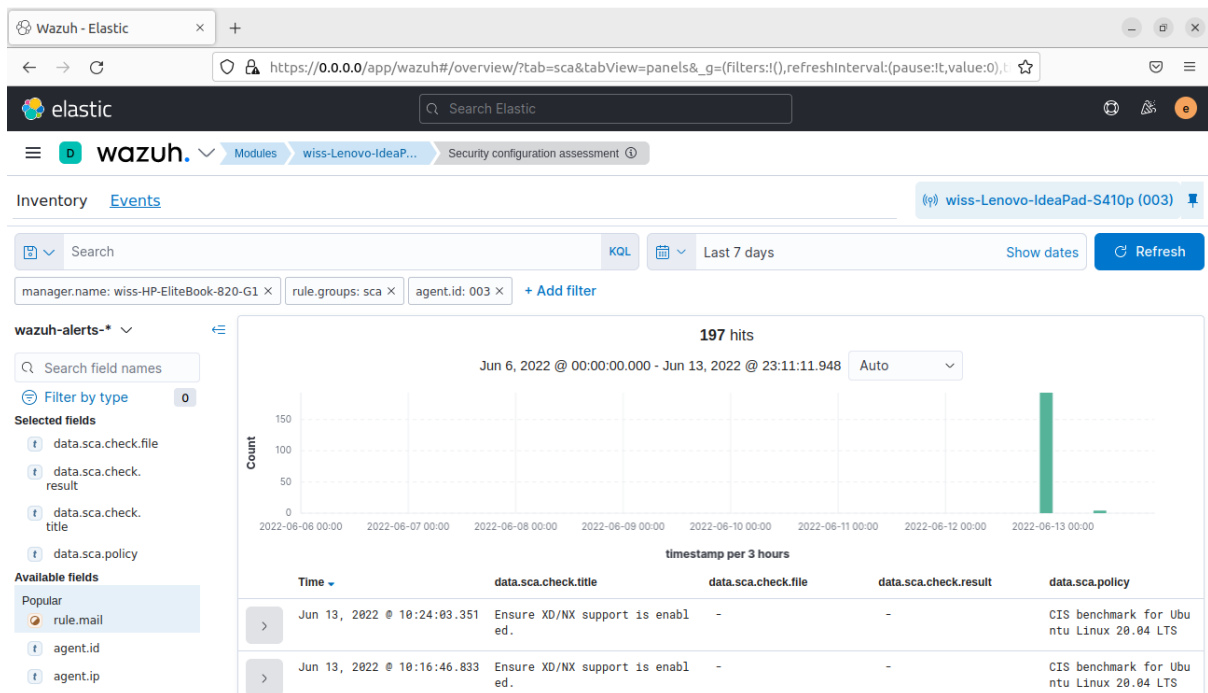


Figure 69: SCA events

The user can create custom security rules and policies. For example, the rules can check the existence of files, directories, keys, and registry values of processes running and can recursively test the presence of files in directories.

These rules have a specific syntax, and it starts with a type of location) which will be the test's target, followed by the actual specification. These tests fall into two categories: existence and content checks. There are five main types of rules described below:

Table 16: Rules type

Type	Character
File	f:
Directory	d:
Process	p:
Commands	c:
Registry (Windows only)	r:

III.6.5 Threat detection and response

III.6.5.1 VirusTotal

VirusTotal is a service that analyzes suspicious files and facilitates the real-time detection of viruses, worms, Trojan horses, and all kinds of malicious software detected by antivirus engines.

- **Dashboard**

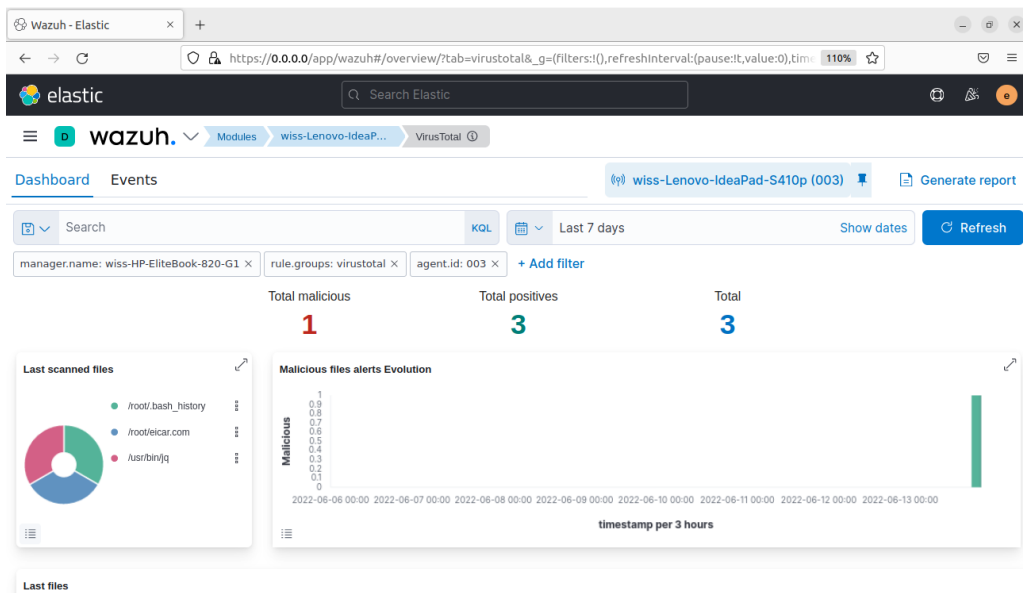


Figure 70: VirusTotal dashboard

- **Events**

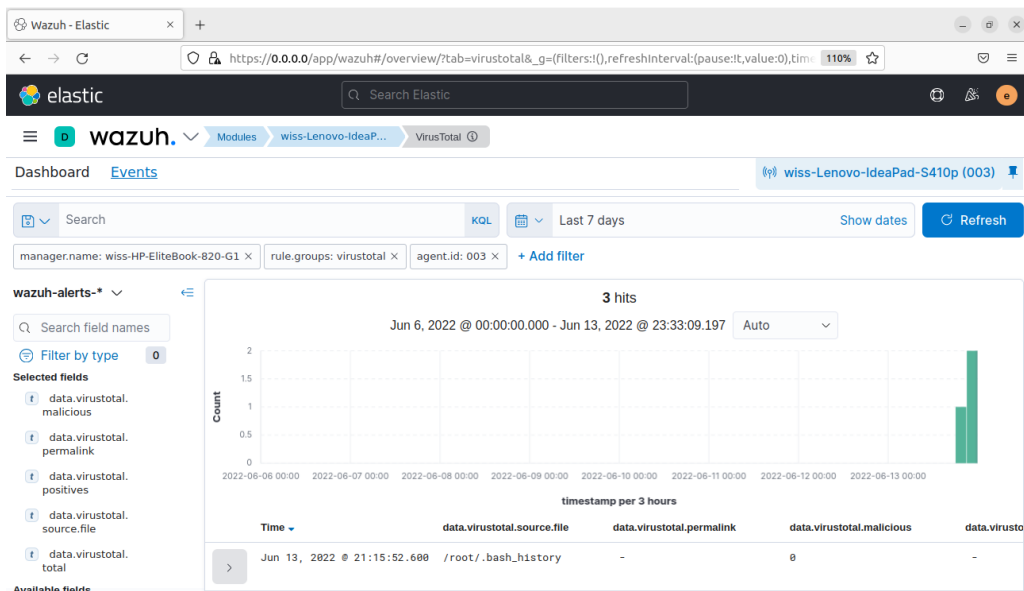


Figure 71: VirusTotal events

III.6.5.2 MITRE ATT&CK

MITRE ATT&CK Matrix is a database of tactics and techniques based on known cyber-attacks. This open and accessible structure is open to continuous improvement. It significantly contributes to the systematic performance of various tests for red and blue teams. MITRE ATT&CK techniques and procedures provide behavioral observability to detect attacks by analyzing the network and end systems. (Al-Shaer et al., 2020).

MITRE ATT&CK Matrix is improved continuously, and there are three types of MITRE Matrix: Enterprise, Mobile, and ICS. We used the MITRE ATT&CK Enterprise Matrix-type for our study.

Due to the diversity of assets in organization control systems, the ATT&CK Enterprise matrix was created, focusing on the functional levels of the enterprise architecture and asset classes to make the correct classification. While each title in the matrix defines a tactical name, techniques used for the relevant tactic are described under each heading. Tactics correspond to what hackers are trying to accomplish, while individual techniques fit how well they achieve those milestones or goals.

A technique can be classified under more than one tactic according to its intended use. The full MITRE ATT&CK Enterprise Matrix⁴.

Table 17: MITRE ATT&CK Enterprise Matrix

Enterprise Matrix

The full ATT&CK Matrix™ below includes techniques spanning Windows, Mac, and Linux platforms and can be used to navigate through the knowledge base.

Last Modified: 2018-10-17T00:14:20.652Z

Initial Access	Execution	Persistence	Privilege Escalation	Defense Evasion	Credential Access	Discovery	Lateral Movement	Collection	Exfiltration	Command and Control
Drive-by Compromise	AppleScript	.bash_profile and .bashrc	Access Token Manipulation	Access Token Manipulation	Account Manipulation	Account Discovery	AppleScript	Audio Capture	Automated Exfiltration	Command and Control
Exploit Public-Facing Application	CMSTP	Accessibility Features	Accessibility Features	BITS Jobs	Bash History	Application Window Discovery	Application Deployment Software	Automated Collection	Data Compressed	Command and Control Through Removal Media
Hardware Additions	Command-Line Interface	Account Manipulation	AppCert DLLs	Binary Padding	Brute Force	Browser Bookmark Discovery	Distributed Component Object Model	Clipboard Data	Data Encrypted	Command and Control Proxy

⁴ <https://attack.mitre.org/matrices/enterprise/>

Tactics correspond to what hackers are trying to accomplish, while individual techniques compare to how well they achieve those milestones or goals.

- **Techniques**

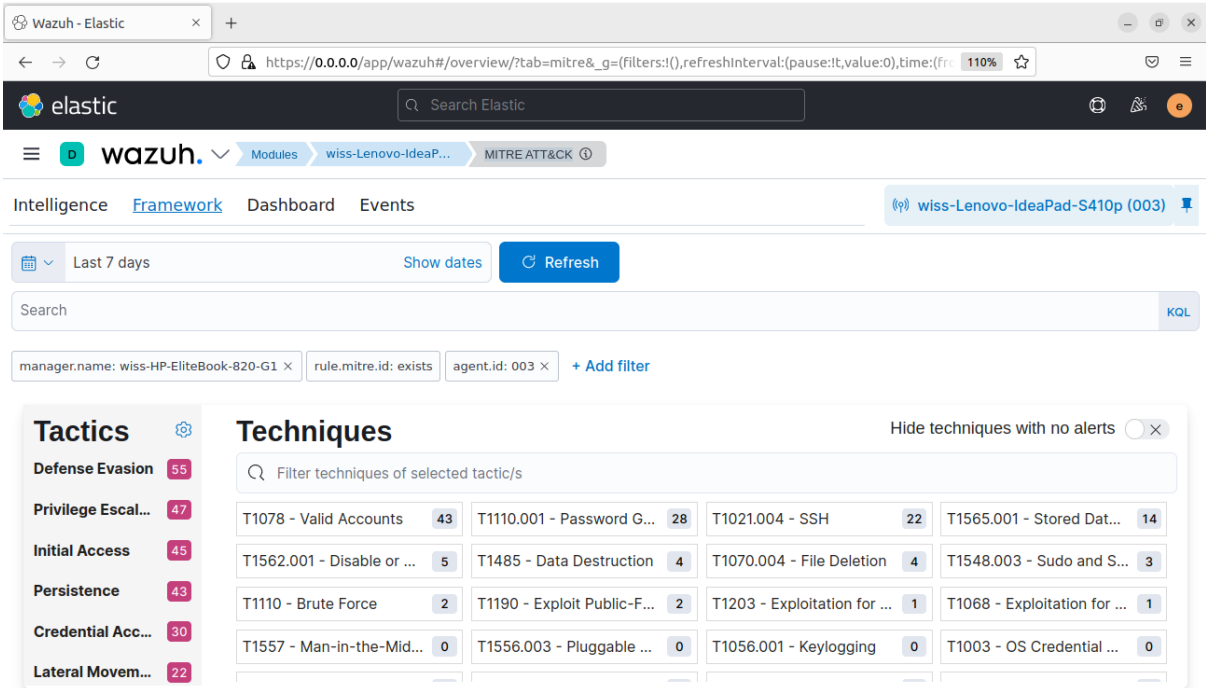


Figure 72: MITRE ATT&CK techniques

An overview of the Valid Accounts technique.

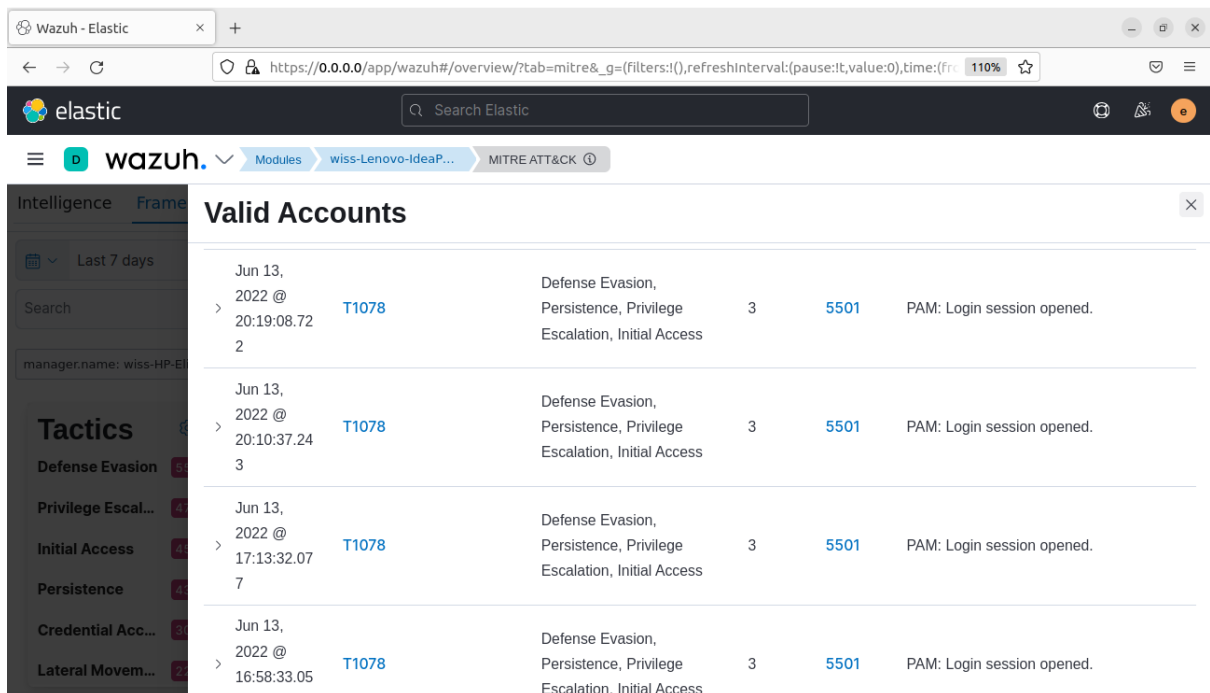


Figure 73: MITRE ATT&CK valid accounts technique

- **Dashboard**

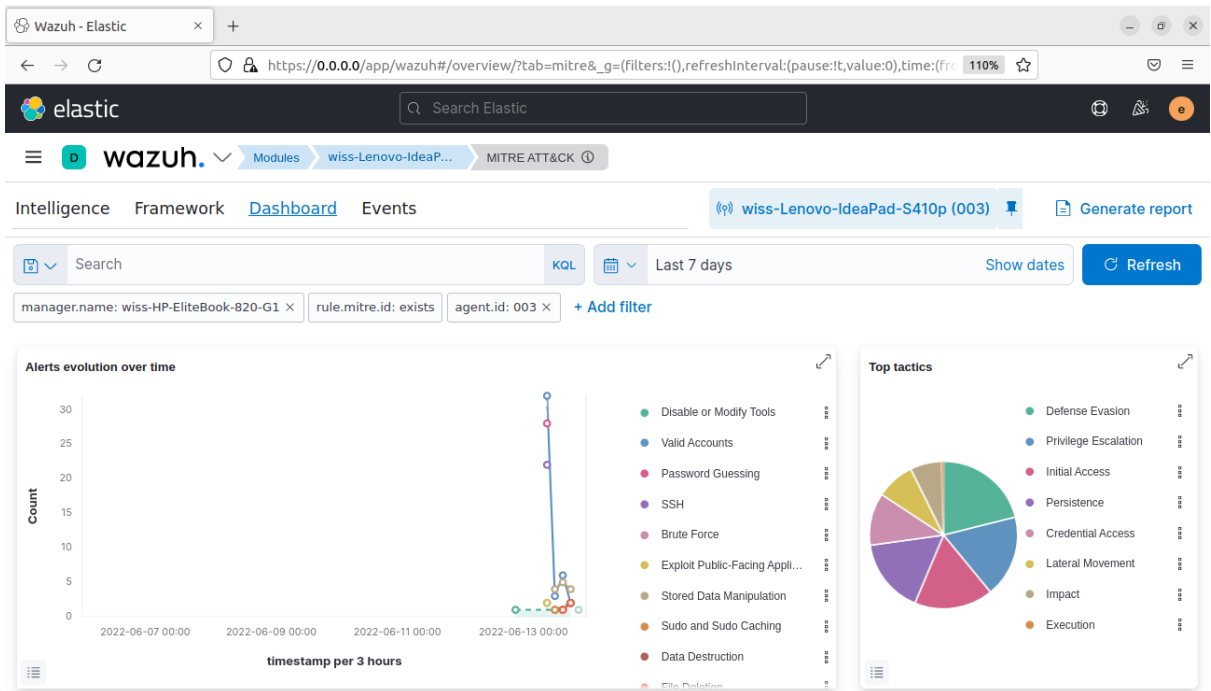
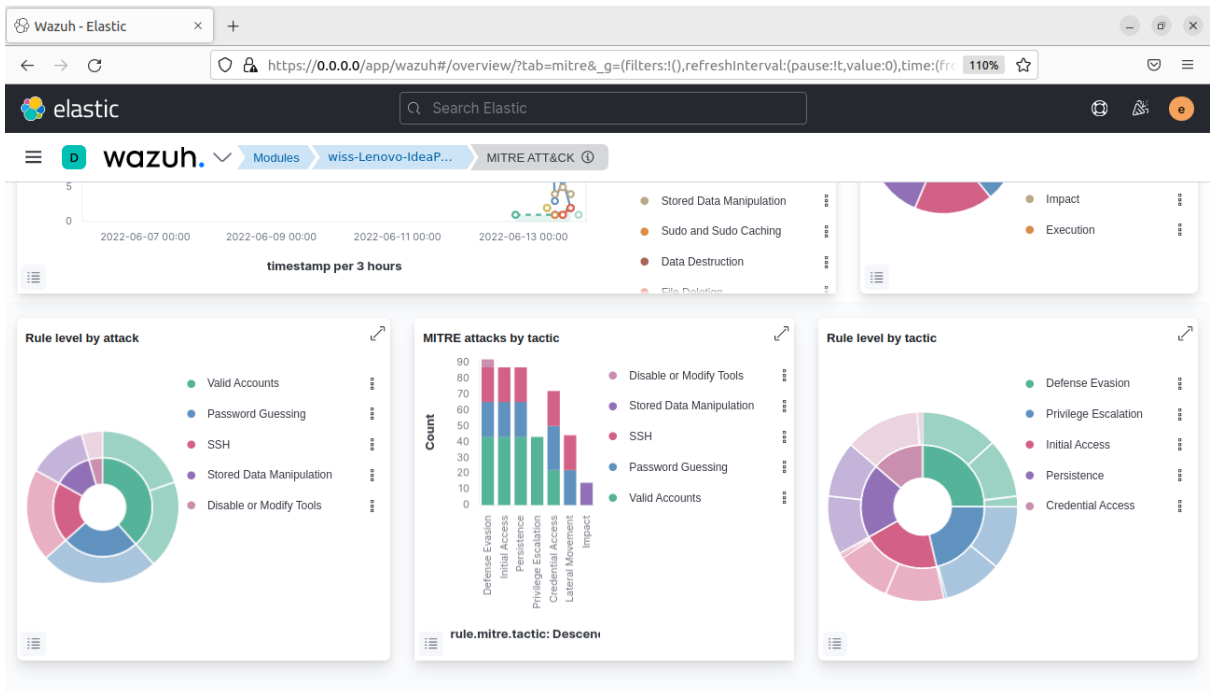


Figure 74: MITRE ATT&CK dashboard



- **Events**

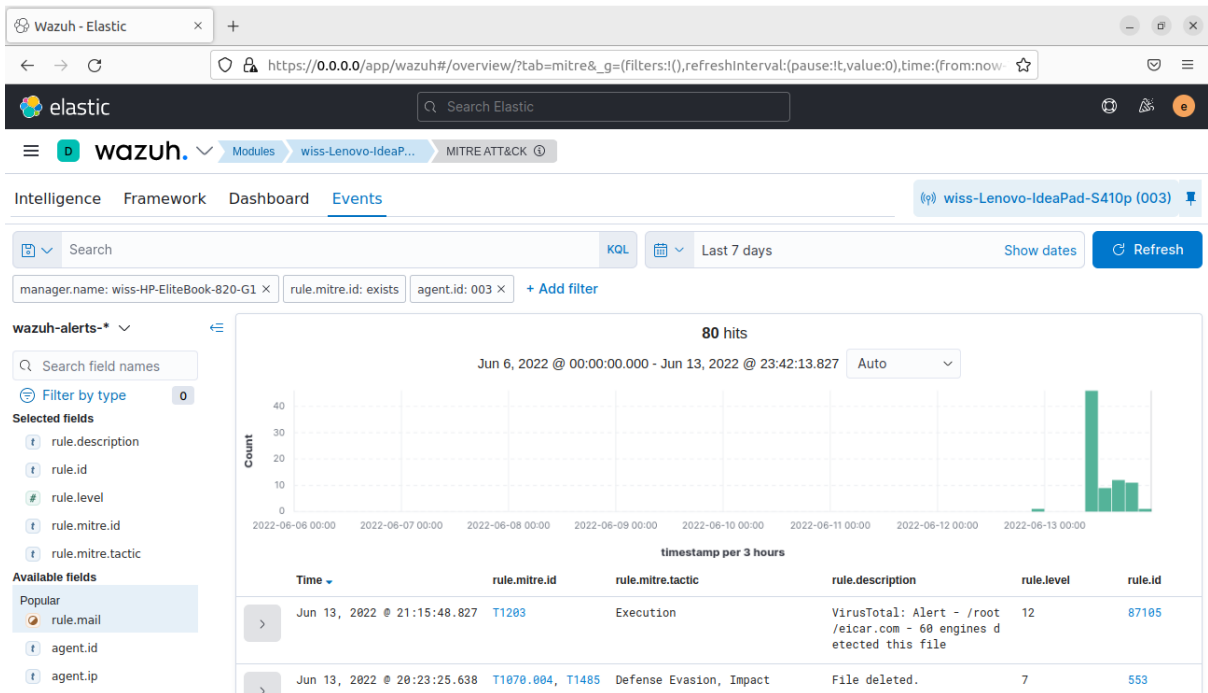


Figure 75: MITRE ATT&CK events

III.6.6 Regulatory compliance

Wazuh rules also include mapping with the following list of regulatory compliance support standards:

- PCI DSS: Payment Card Industry Data Security Standard.
- GDPR: The General Data Protection Regulation (EU).
- HIPAA: Health Insurance Portability and Accountability Act.
- NIST 800-53: NIST Special Publication 800-53.
- GPG13: Good Practice Guide 13.
- TSC SOC2: Trust Services Criteria.

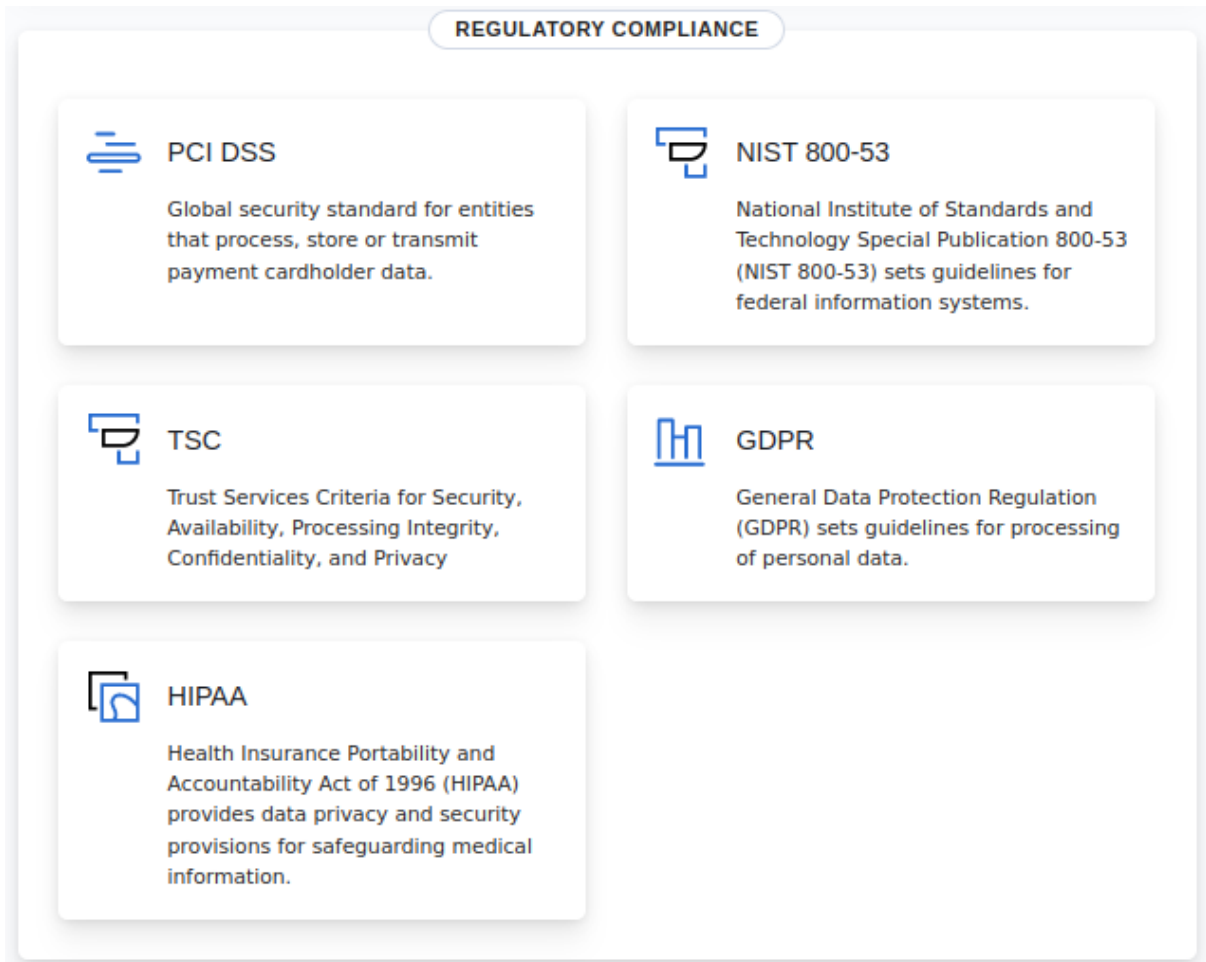


Figure 76: Wazuh regulatory compliance

III.6.7 Wazuh roles

Wazuh manager includes role-based access control. This feature manages endpoints (agents) access using policies and user roles. It also allows the user to assign roles and procedures to users who will perform different functions in the environment. The manager custom the roles according to the organization's policies; in this case, the rules are by default.

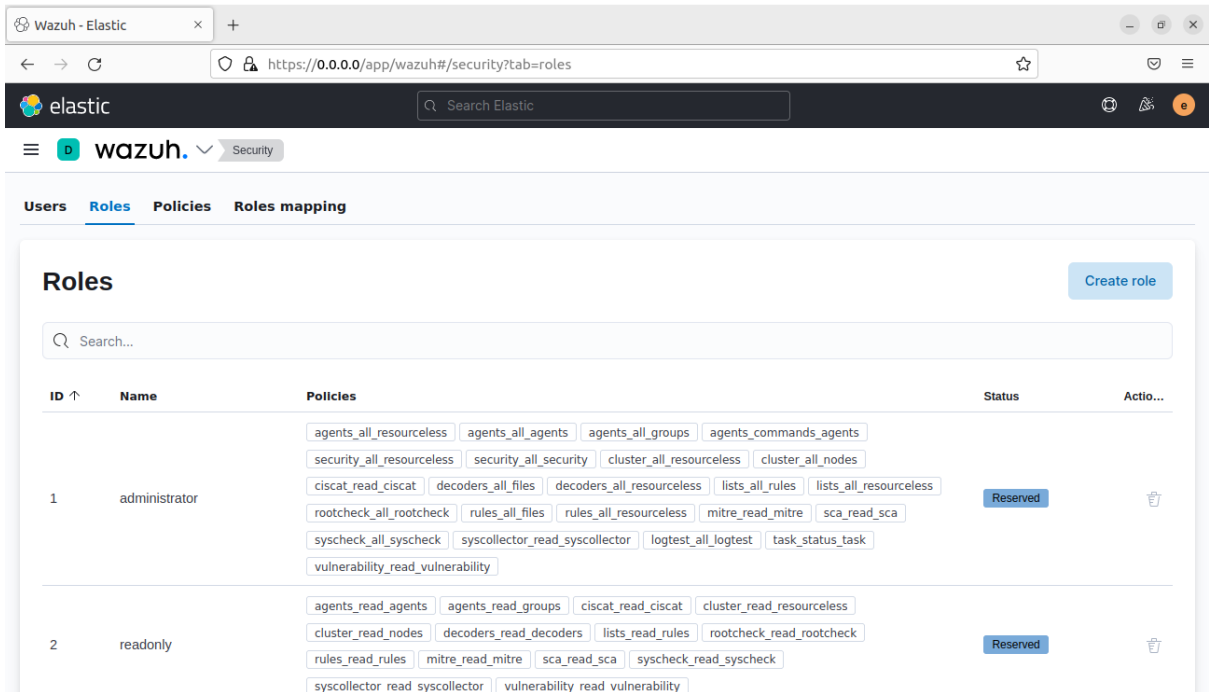


Figure 77: Wazuh roles

III.7 Elastic user interface

III.7.1 Integrations

Elastic provides 249 integrations for collecting and analyzing data from different sources; the user can choose a log source from the list of log sources displayed to connect and explore it with the elastic user interface.

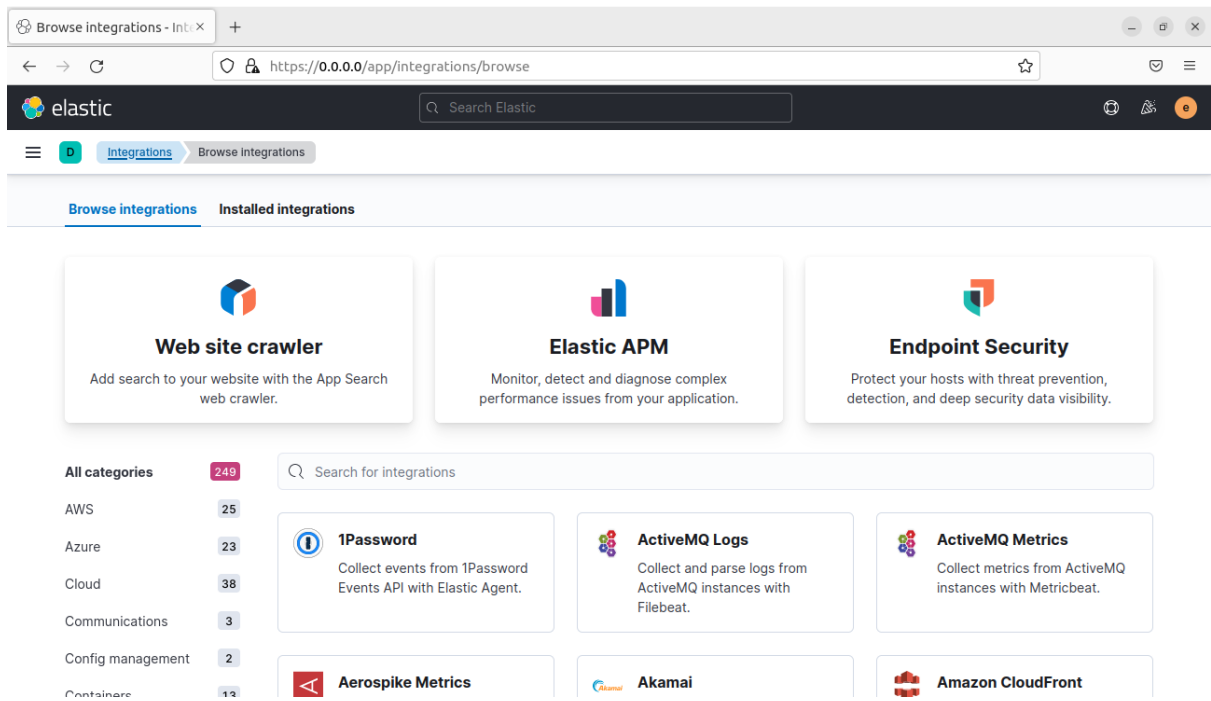


Figure 78: Elastic integrations

III.8 Alerts notifications

To receive real-time alert notifications, we suggest that the company use the communication platforms to facilitate the monitoring of events; in our case, we propose slack.



Figure 79: Stack platform icon

Slack is a business messaging app connecting people to the information they need. Slack transforms organizational communication by connecting people to collaborate as one unified team.(Johnson, 2018).

We created a workspace named “ Alerts Supervision “ *alertssupervision.slack.com*. It contains multiple channels; we create a #Supervision channel to receive the alerts notifications, and the workspace manager can add the employees to monitor.

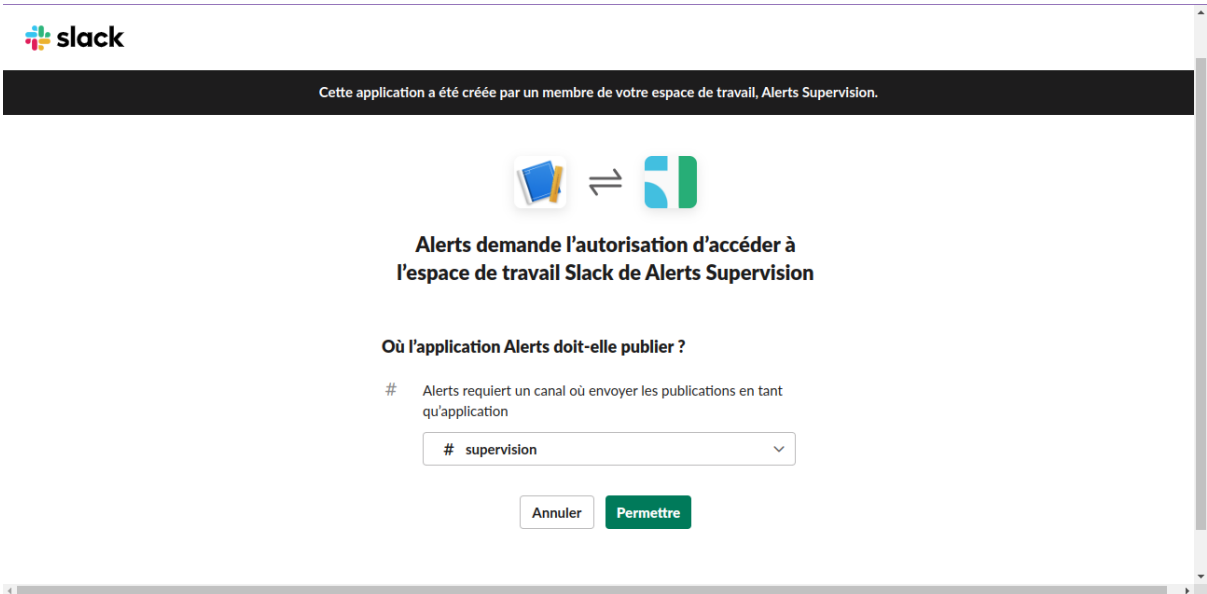


Figure 80: Integrating Slack with Wazuh

The Slack integration must be added in the Wazuh manager configuration file.

```
<integration>
<name>slack</name>
<hook_url>https://hooks.slack.com/services/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N </hook_url> <!-- Replace with your Slack Webhook -->
<level>5</level>
<alert_format>json</alert_format>
</integration>
```

All the previous alerts notification loaded to the #supervision channel in slack.

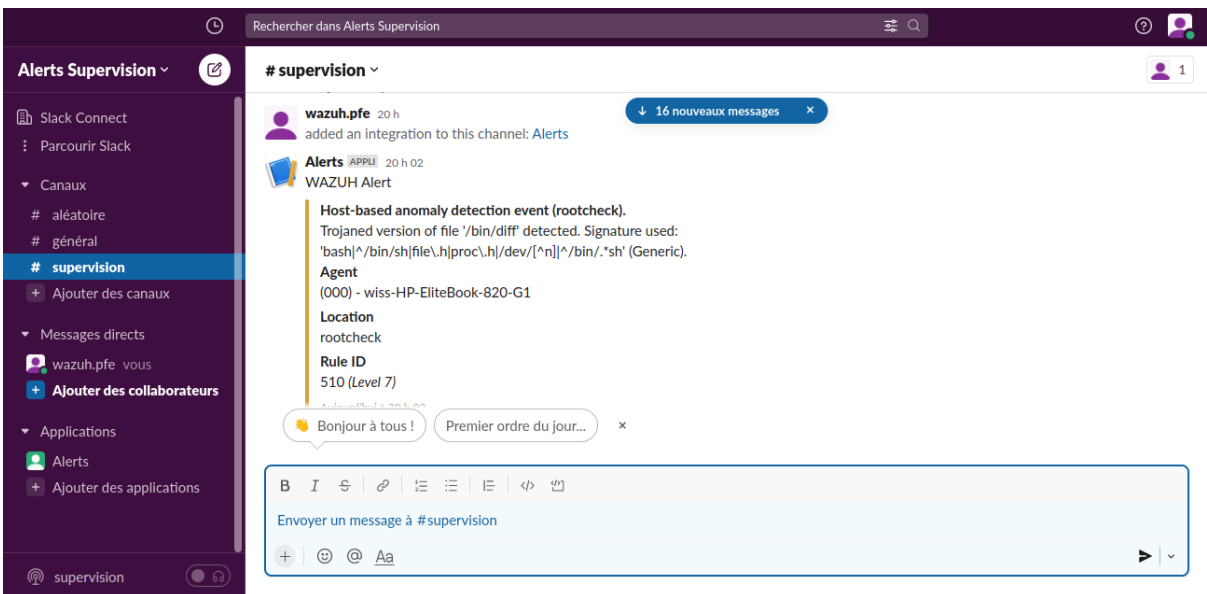


Figure 81: Alerts notification in Slack

Conclusion

In the previous chapter, we understood that the solution we arrived at is a combination of ELK Stack and Wazuh open sources tools; each tool has a robust architecture that provides the missing features of the other.

Above, we describe our solution's conceptual study and the detailed architecture of the tools used. Then, we presented the realization of the application by exposing the working environment and illustrating the application's graphical interfaces and functionalities. Finally, we presented the interfaces of the graph visualization and illustrations on incident detection.

Chapter IV



Tests and results

Introduction

Multiple methods of IS cyber attacks are applied from one system to another, depending on our victim.

Kaspersky ICS CERT, a global project run by Kaspersky to coordinate the efforts of industrial automation system vendors and industrial facility owners and operators, reports that the malicious objects that are blocked by Kaspersky products on ICS computers during 2021 fall into many categories. (Threat Landscape for Industrial Automation Systems. Statistics for H2 2021 | Kaspersky ICS CERT).

The following figure revealed estimated percentages of ICS computers on which the activity of malicious objects from different categories had been prevented:

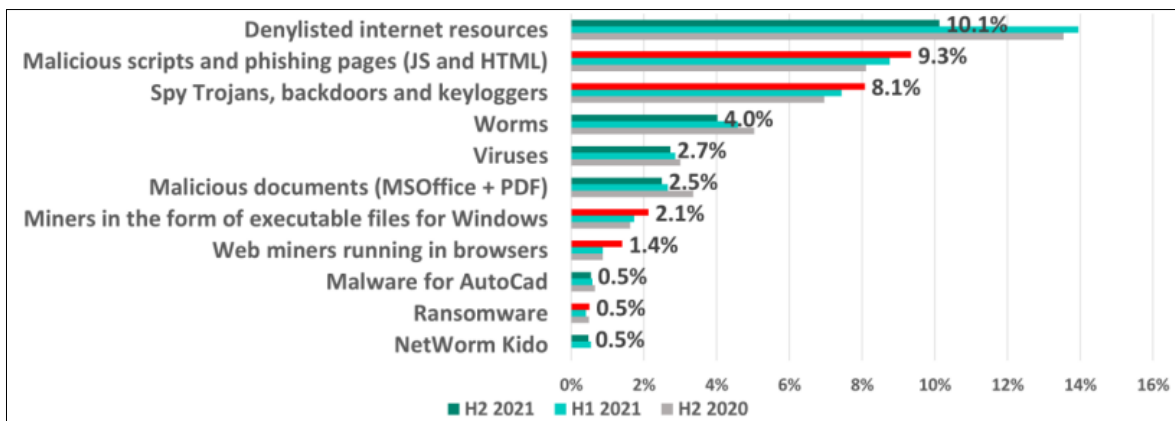


Figure 82: Percentage of ICS computers on which malicious objects from various categories were blocked

According to our prerequisites in the fundamental notions in cyber security, we did simple tests while creating different scenarios of attacks to verify the effectiveness of our project using our lab environment.

IV.1 Detecting a Brute-force attack:

IV.1.1 Overview

This test performs a brute-force attack on a machine with a Wazuh Agent installed. It can detect and identify in the Wazuh manager the multiple succession authentication attempts by a user as a brute-force attack. Furthermore, it aims to highlight the capabilities of a monitoring and logging tool to identify any cyber attack, in this test, as a brute-force

attack. We use Hydra, an open-source parallelized login cracker, to perform the brute-force attack. (Kali Linux, 2021).

IV.1.2 Prerequisites

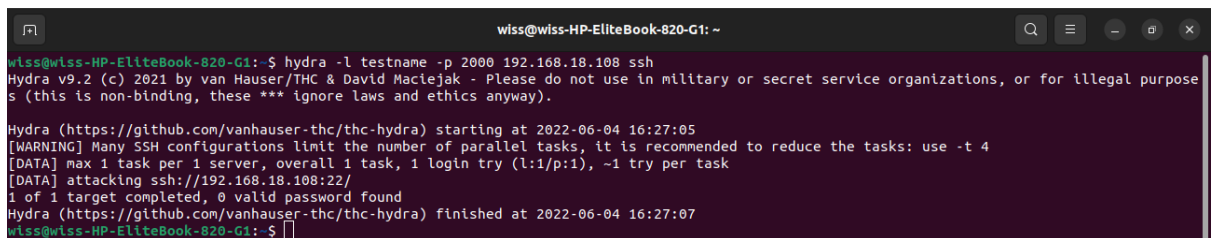
- Attacker machine: Linux system with Hydra and SSH installed.
- Victim machine: Wazuh agent installed.

IV.1.3 Test Steps

Run the following command to use Hydra to perform multiple login attempts on a machine with a Wazuh Agent.

We first run the following command one time to perform a log-in attempt on a victim machine.

hydra -l [username] -p [password] [host_ip] [ssh]



```
wiss@wiss-HP-EliteBook-820-G1: ~  
wls@wls-HP-EliteBook-820-G1:~$ hydra -l testname -p 2000 192.168.18.108 ssh  
Hydra v9.2 (c) 2021 by van Hauser/THC & David Maciejak - Please do not use in military or secret service organizations, or for illegal purposes (this is non-binding, these *** ignore laws and ethics anyway).  
  
Hydra (https://github.com/vanhauser-thc/thc-hydra) starting at 2022-06-04 16:27:05  
[WARNING] Many SSH configurations limit the number of parallel tasks, it is recommended to reduce the tasks: use -t 4  
[DATA] max 1 task per 1 server, overall 1 task, 1 login try (l:i/p:1), ~1 try per task  
[DATA] attacking ssh://192.168.18.108:22/  
1 of 1 target completed, 0 valid password found  
Hydra (https://github.com/vanhauser-thc/thc-hydra) finished at 2022-06-04 16:27:07  
wls@wls-HP-EliteBook-820-G1:~$
```

- Test name: A guessing user name of our victim machine.
- 2000: A guessing password of our victim machine.
- 192.168.18.108: Victims' IP address .
- Ssh: Type of the communication protocol.

In this case, an Sshd: Attempt to log in using a non-existent user is performed on the attacker's machine. This attack tries to gain access to the “victim” ’s account.

IV.1.4 Test Results

After executing the previous command once, we access the Wazuh Alerts in the Security Events module of the Wazuh Kibana plugin. We find that in response to the experiment, Wazuh generated an alert with rule ID 5710.



Figure 83: Graphic of the alerts generated by Wazuh in response to the ssh attempt

The details of Wazuh alerts were developed in response to the ssh attempt.

Table 18: Wazuh alerts information developed in response to the ssh attempt

Field Name	Alert Id: 5710
rule.description	Sshd: Attempt to login using a non-existent user
rule.firedtimes	4
rule.level	5
rule.mitre.tactic	Credential Access, Lateral Movement, Defense Evasion, Persistence, Privilege Escalation, Initial Access
rule.mitre.technique	Password Guessing, SSH, Valid Account
full.log	Jun 4 16:27:06 wiss-Lenovo-IdeaPad-S410p sshd [17932]: Failed password for invalid user testname from 192.168.18.93 port 45304 ssh2
location	var/log/auth.log

Another time, we executed the previous command, but this time, around 4 tries, we accessed the Wazuh Alerts in the Security Events module of the Wazuh Kibana plugin. Again, a brute-force attack is performed. We find that in response to the experiment, Wazuh has generated alerts with rule IDs 5503 and 5712.

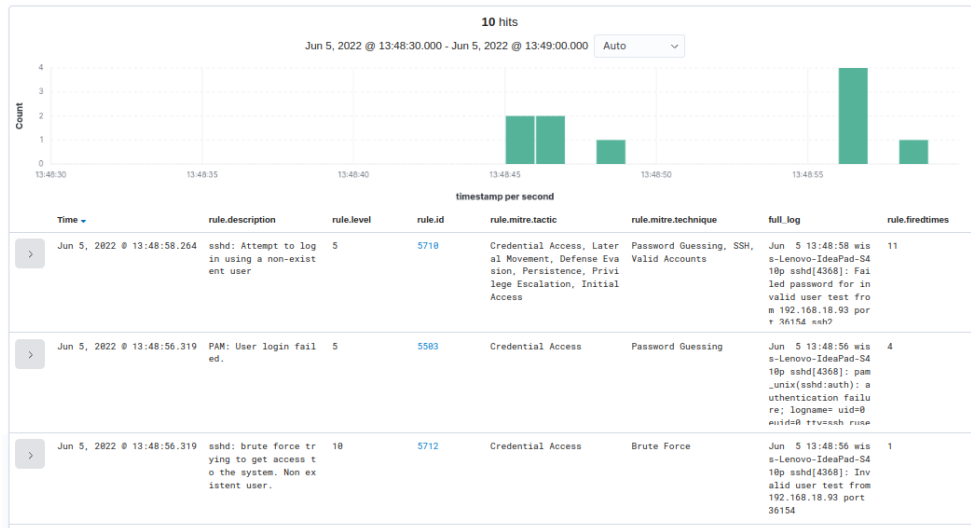


Figure 84: Graphic of the alerts generated by Wazuh in response to the brute force attack

The details of Wazuh alerts were generated in response to the brute-force attack.

Table 19: Wazuh alerts information generated in response to the brute-force attack

Field Name	Alert Id 5503	Alert Id 5712
rule.description	PAM: User login failed.	sshd: brute force trying to get access to the system. Non existent user.
rule.firedtimes	4	1
Rule.level	5	10
rule.groups	pam, syslog, authentication_failed	syslog, sshd, authentication_failures
rule.mitre.tactic	Credential Access	Credential Access
rule.mitre.technique	Password Guessing	Brute Force
full.log	Jun 5 13:48:56 wiss-Lenovo-IdeaPad-S410p sshd[4368]: pam_unix(sshd:auth): authentication failure; logname= uid=0 euid=0 tty=ssh ruser= rhost=192.168.18.93	Jun 5 13:48:56 wiss-Lenovo-IdeaPad-S410p sshd[4368]: Invalid user test from 192.168.18.93 port 36154
location	/var/log/auth.log	/var/log/auth.log

IV.2 File integrity monitoring

IV.2.1 Overview

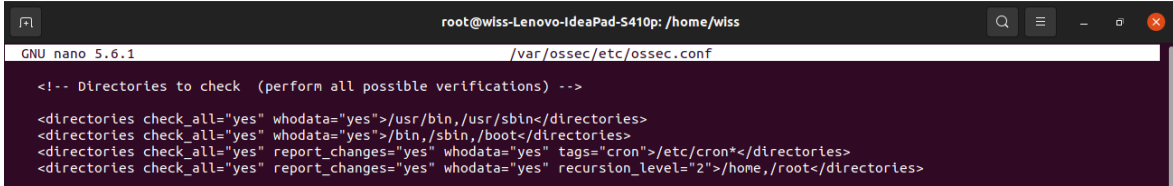
This test can demonstrate if Wazuh can detect the details of file changes within the system by consulting the logged information. These logs include information on when and by whom these changes happen.

IV.2.2 Prerequisites

A Machine where the wazuh agent is installed must contain an auditing subsystem for recording the event in the monitored directories. The installation is already presented in the previous point.

IV.2.3 Test Steps

Adding the monitored directories to the Wazuh agent configuration file `/var/ossec/etc/ossec.conf`.



```
root@wiss-Lenovo-IdeaPad-S410p: /home/wiss
GNU nano 5.6.1 /var/ossec/etc/ossec.conf
<!-- Directories to check (perform all possible verifications) -->
<directories check_all="yes" whodata="yes">/usr/bin,/usr/sbin</directories>
<directories check_all="yes" whodata="yes">/bin,/sbin,/boot</directories>
<directories check_all="yes" report_changes="yes" whodata="yes" tags="cron">/etc/cron*</directories>
<directories check_all="yes" report_changes="yes" whodata="yes" recursion_level="2">/home,/root</directories>
```

- `check_all`: allows, allows recording file sizes, permissions, owner, last modification date...
- `whodata`: use the Linux Audit subsystem to record the author of the events in the directory.
- `report_changes`: ensure that the logs contain the changed content of the file.

IV.2.4 Test Results

The following graph and tables display the Wazuh alerts generated in response to folder `/home/wiss/Pictures` changes.

- Adding a file: in our case, adding a screenshot to the folder Pictures.

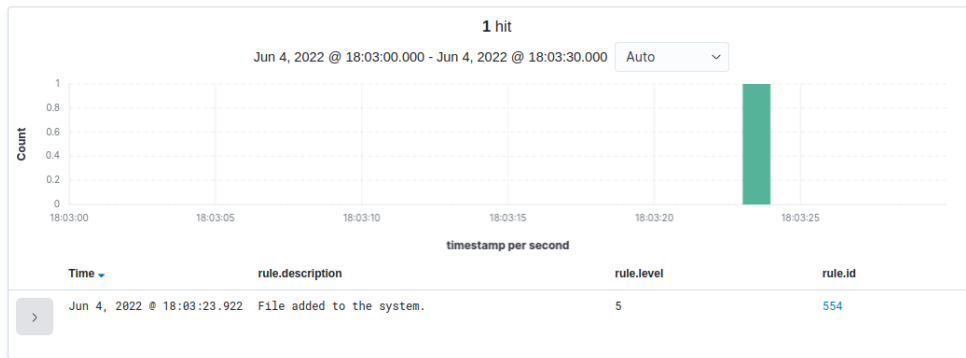


Figure 85:Graphic of the alerts generated by Wazuh in response to adding a file

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

Table 20: Wazuh alerts information generated in response to the creation of a file

Field Name	Alert Id 554
rule.description	File added to the system
rule.firedtimes	1
rule.level	5
Full.log	File' <i>home/wiss/Pictures/Screenshot from 2022-06-04 17-57-16.png</i> ' added Mode : whodata
location	syscheck
rule.groups	ossec, syscheck, syscheck_entry_added, syscheck_file
syscheck.audit.login_user.name	<i>wiss</i>
syscheck.audit.process.name	<i>usr/lib/systemd/systemd</i>
syscheck.event	<i>added</i>
syscheck.path	<i>/home/wiss/Pictures/Screenshot from 2022-06-04 17-57-16.png</i>

- Deleting a file: in our case, deleting a screenshot to the folder Pictures.

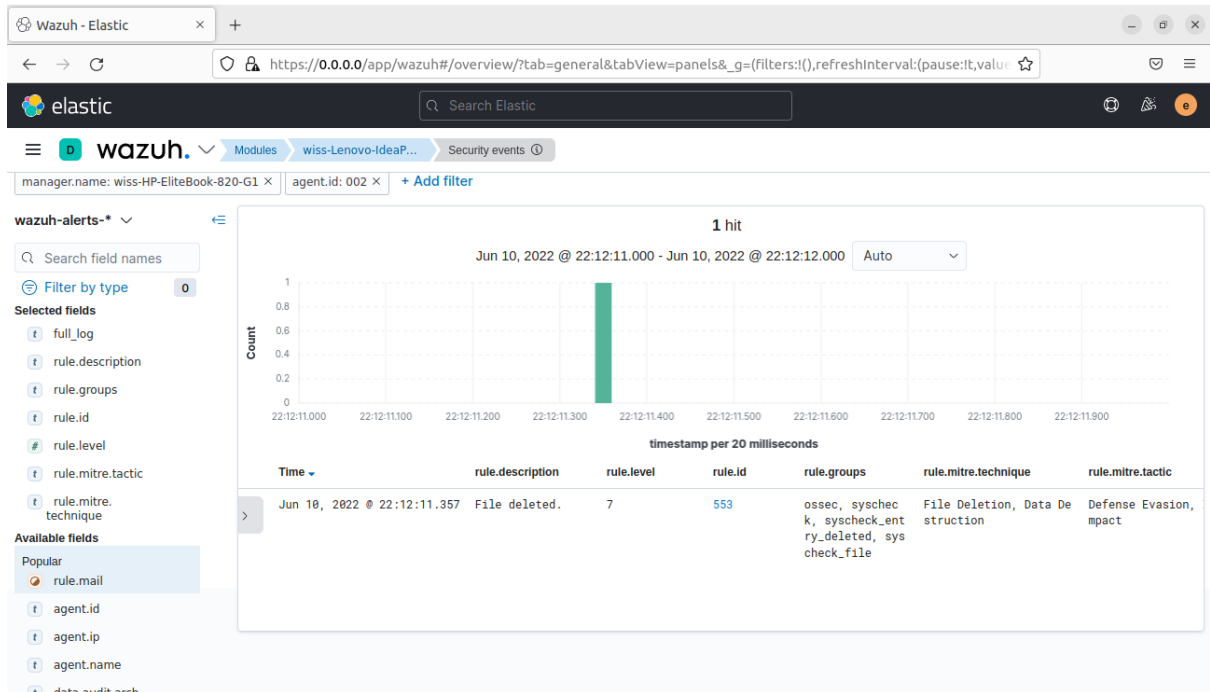


Figure 86: Graphic of the alerts generated by Wazuh in response to deleting a file

Alert generated by Wazuh in response to the deleting of a file.

Table 21: Wazuh alerts information generated in response to deleting a file

Field Name	Alert Id 553
rule.description	File deleted
rule.firedtimes	1
rule.level	7
Full.log	File '/home/wiss/Pictures/Screenshot from 2022-05-30 21-47-58.png' deleted Mode: whodata
location	syscheck
rule.mitre.tactic	Defense Evasion, Impact
rule.mitre.technique	File Deletion, Data Destruction
rule.groups	ossec, syscheck, syscheck_entry_deleted, syscheck_file
syscheck.audit.login_user.name	wiss
syscheck.audit.process.name	usr/lib/systemd/systemd
syscheck.event	deleted
syscheck.path	/home/wiss/Pictures/Screenshot from 2022-05-30 21-47-58.png

- Creating a file named : Test

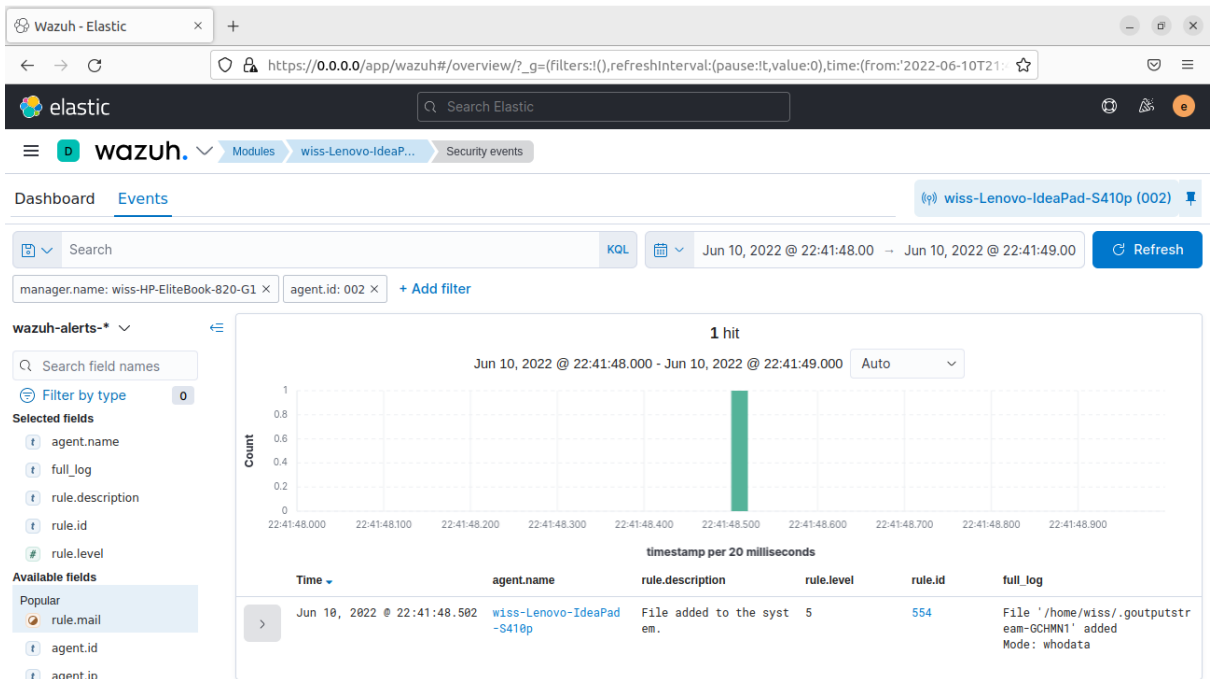


Figure 87: Graphic of the alerts generated by Wazuh in response to adding a file 2

- Changing the content of the previous file

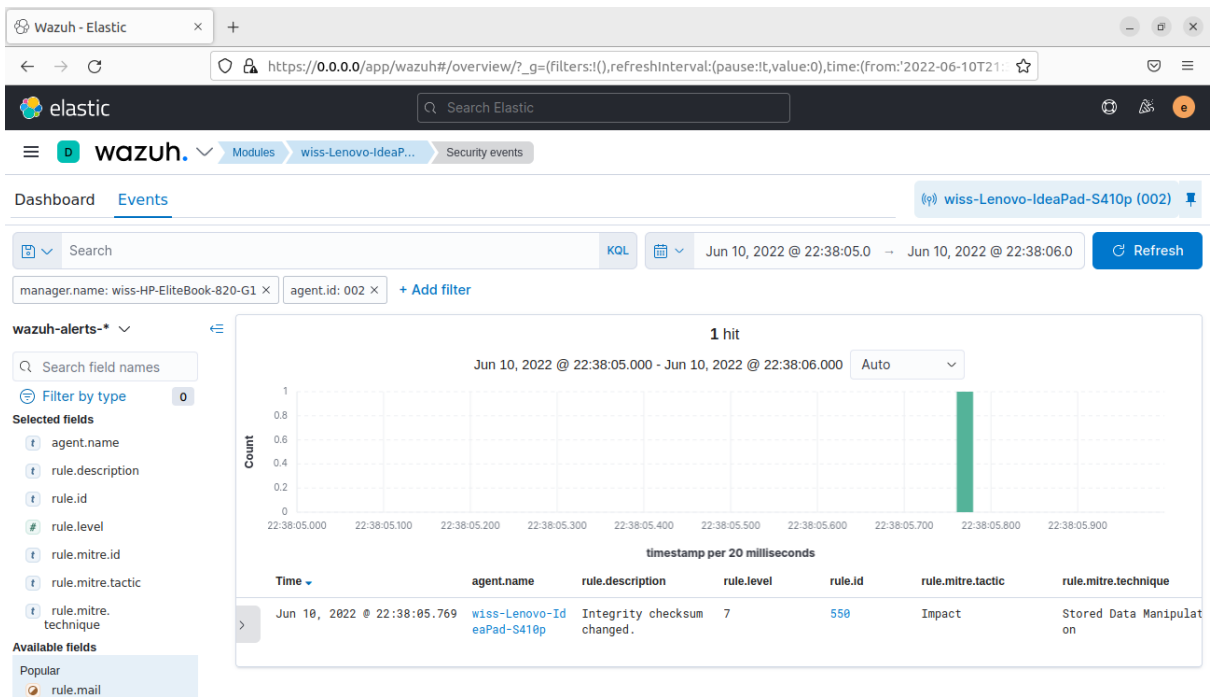


Figure 88: Graphic of the alerts generated by Wazuh in response to modifying a file

Alert generated by Wazuh in response to changing the content of a file.

Table 22: Wazuh alerts information generated in response to changing the content of a file

Field Name	Alert Id 550
rule.description	Integrity checksum changed.
rule.firedtimes	1
rule.level	7
Full.log	file '/home/wiss/Test' modified Mode: whodata Changed attributes: size,mtime,md5,sha1,sha256 Size changed from '0' to '4' Old modification time was: '1654897082', now it is '1654897083' Old md5sum was: 'd41d8cd98f00b204e9800998ecf8427e' New md5sum is : '5841af25c604855409fc52cabe606f26'
syscheck.changed_attributes	size, mtime, md5, sha1, sha256
location	syscheck
rule.groups	ossec, syscheck, syscheck_entry_modified, syscheck_file
rule.mitre.tactic	impact
rule.mitre.technique	Stored Data Manipulation
syscheck.audit.login_user.name	wiss
syscheck.audit.process.name	usr/lib/systemd/systemd
syscheck.event	modified
syscheck.path	/home/wiss/Test

IV.3 Detecting Unauthorised Processes

IV.3.1 Overview

This test allows seeing if Wazuh can detect the execution of black-listed processes. Monitoring running processes on a system alert when crucial processes or functions suddenly stop and identifies unknown starting within the system. This test demonstrates the ability of monitoring and logging tools.

IV.3.2 Prerequisites

A machine contains a Wazuh agent configured to obtain a list of currently running processes periodically.

```

<!-- Unauthorized process -->
<localfile>
  <log_format>full_command</log_format>
  <alias>process_list</alias>
  <command>ps -e -o pid,uname,command</command>
  <frequency>30</frequency>
</localfile>

```

And nmap-ncat installed on it

- ps -e -o pid,uname,command: to obtain the running process list, every 30 seconds.

Configuring the Wazuh manager to receive the running process list, by adding a rule.

var/ossec/etc/ rules/local_rules.xml.

```

GNU nano 6.2 /var/ossec/etc/rules/local_rules.xml
</group>
<group name="ossec,">
  <rule id="100050" level="0">
    <if_sid>530</if_sid>
    <match>^ossec: output: 'process list'</match>
    <description>List of running processes.</description>
    <group>process_monitor,</group>
  </rule>
  <rule id="100051" level="7" ignore="900">
    <if_sid>100050</if_sid>
    <match>nc -l</match>
    <description>Netcat listening for incoming connections.</description>
    <group>process_monitor,</group>
  </rule>
</group>
<group name="attack,">
  <rule id="100100" level="10">
    <if_group>web|attack|attacks</if_group>
    <list field="srcip" lookup="address_match_key">etc/lists/blacklist-alienvault</list>
    <description>IP address found in AlienVault reputation database.</description>
  </rule>
</group>

```

- Level = '0': The collection of running processes will not trigger an alert.
- Netcat, a standard computer networking utility that allows port scanning and listening, added a rule to black-list a process.

IV.3.3 Test Steps

Running ncat in the agent to listen to the TCP port

```

w1ss@w1ss-Lenovo-IdeaPad-S410p: ~
w1ss@w1ss-Lenovo-IdeaPad-S410p:~$ sudo nc -l 8000
[sudo] password for w1ss:
^C
w1ss@w1ss-Lenovo-IdeaPad-S410p:~$

```

IV.3.4 Test Results

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

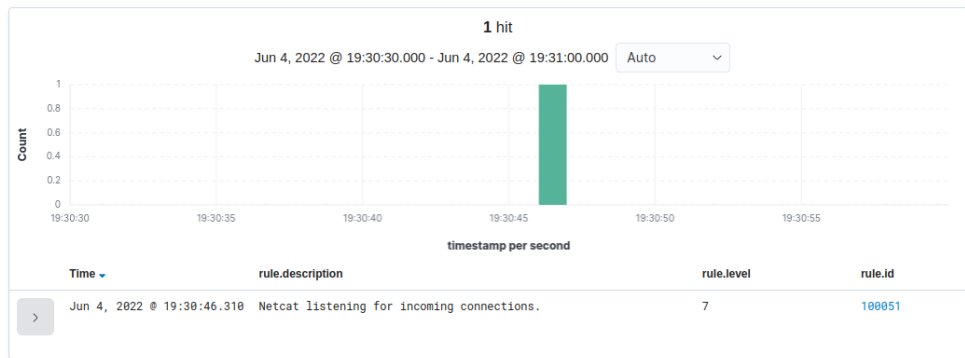


Figure 89: Graphic of the alerts generated by Wazuh in response to the execution of the black-listed Netcat command

The details of the Wazuh alert generated in response to the black-listed Netcat Command.

Table 23: Wazuh information generated in response to the black-listed Netcat Command

Field Name	Alert Id 100051
rule.description	Netcat listening for incoming connections
rule.level	7
Rule.groups	Ossec, process_monitor
Full.log	Ossec: output : ' process list' PID USER COMMAND 1 root /sbin/init splash
location	Process list

IV.4 Detecting an SQL Injection attack

IV.4.1 Overview

This test performs a SQL injection attack on a machine with a Wazuh Agent. In addition, this test can detect and identify the SQL patterns present in the attack.

IV.4.2 Prerequisites

Victim agent machine: - Having an Apache server.

- Agent configured file to capture the event from the Apache server.

```
<!-- Apache -->
<localfile>
  <log_format>apache</log_format>
  <location>/var/log/httpd/access_log</location>
</localfile>
```

IV.4.3 Test Steps

From an external machine, we send the following command to the Apache server:

```
curl -XGET "http://<apache_server_ip>/?id=SELECT+*+FROM+users"
```



IV.4.4 Test Results

After executing the command, we access the Wazuh Alerts in the Security Events module of the Wazuh Kibana plugin. We can see the alert created in response to the previous server request identified as a SQL injection attack.

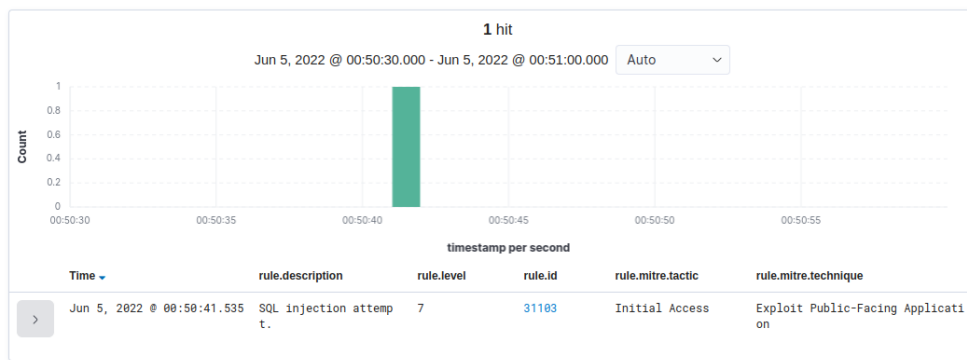


Figure 90: Graphic of the alerts generated by Wazuh in response to SQL injection

The details of the Wazuh alert generated in response to the SQL Injection attack.

Table 24: Wazuh information generated in response to the SQL Injection attack

Field Name	Alert Id 31103
rule.description	SQL injection attempt
rule.firedtimes	1
Rule.level	7
rule.mitre.tactic	Initial Access
rule.mitre.technique	Exploit Public-Facing Application
Rule.groups	Web, accesslog, attack, sql_injection
Full.log	192.168.18.93 - - [05/Jun/2022 00:50:40 + 0100] "GET

	<code>/?id=SELECT+*+FROM+users HTTP/1.1" 403 440 "-- " curl/7.81.0'</code>
location	<code>varlog/apache2/access.log</code>

IV.5 Detecting a Shellshock attack

IV.5.1 Overview

A shellshock attack is a code injection attack, running shell commands on the victim machine by sending maliciously crafted web requests. This test performs a shellshock attack on a machine with a Wazuh Agent installed. The Wazuh manager detects and identifies the shell commands present in the attack.

IV.5.2 Prerequisites

Same as the previous test (SQL injection).

IV.5.3 Test Steps

From an external machine, we send the following command to the Apache server:

```
$ curl -H "User-Agent: () { ;; }; /bin/cat /etc/passwd" < apache_web_server_address >
```

```
root@w1ss-HP-EliteBook-820-G1:/home/w1ss# curl -H "User-Agent: () { ;; }; /bin/cat /etc/passwd" 192.168.18.108
<!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML 2.0//EN">
<html><head>
<title>403 Forbidden</title>
</head><body>
<h1>Forbidden</h1>
<p>You don't have permission to access this resource.</p>
<hr>
<address>Apache/2.4.48 (Ubuntu) Server at 192.168.18.108 Port 80</address>
</body></html>
root@w1ss-HP-EliteBook-820-G1:/home/w1ss# curl -H "User-Agent: () { ;; }; /bin/cat /etc/passwd" 192.168.18.108
<!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML 2.0//EN">
<html><head>
<title>403 Forbidden</title>
</head><body>
<h1>Forbidden</h1>
<p>You don't have permission to access this resource.</p>
<hr>
<address>Apache/2.4.48 (Ubuntu) Server at 192.168.18.108 Port 80</address>
</body></html>
root@w1ss-HP-EliteBook-820-G1:/home/w1ss# curl -H "User-Agent: () { ;; }; /bin/cat /etc/passwd" 192.168.18.108
<!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML 2.0//EN">
<html><head>
<title>403 Forbidden</title>
</head><body>
<h1>Forbidden</h1>
<p>You don't have permission to access this resource.</p>
<hr>
<address>Apache/2.4.48 (Ubuntu) Server at 192.168.18.108 Port 80</address>
</body></html>
root@w1ss-HP-EliteBook-820-G1:/home/w1ss#
```

IV.5.4 Test Results

After executing the previous command, we access the Wazuh Alerts in the Security Events module of the Wazuh Kibana plugin. As a result, the alert created in response to the previous server request identified a Shellshock attack.

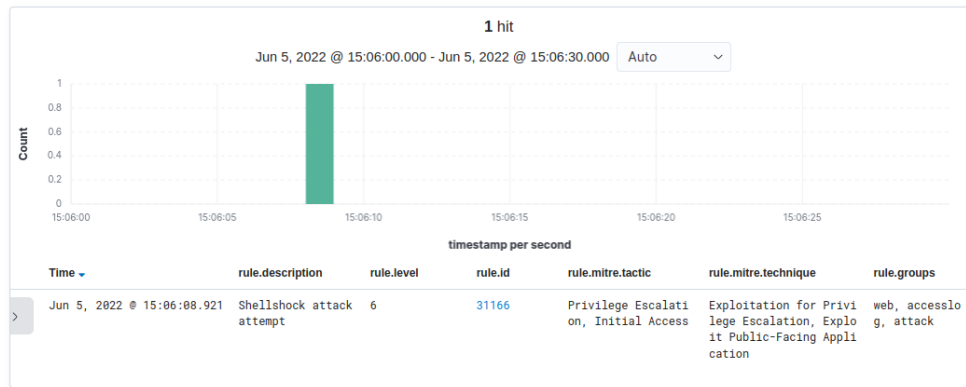


Figure 91: Graphic of the alerts generated by Wazuh in response to the Shellshock attack

The details of the Wazuh alert generated in response to the Shellshock attack.

Table 25: Wazuh information generated in response to the Shellshock attack

Field Name	Alert Id 31166
rule.description	Shellshock attack attempt
rule.firedtimes	1
Rule.level	6
rule.mitre.tactic	Privilege Escalation, Initial Access
rule.mitre.technique	Exploitation for Privilege Escalation, Exploit Public-Facing Application
Rule.groups	web, accesslog, attack
Full.log	192.168.18.93 - - [05/Jun/2022:15:06:08 +0100] "GET / HTTP/1.1" 403 440 "-" "(" { ; } ; /bin/cat /etc/passwd"
location	varlog/apache2/access.log

IV.6 Detecting Suspicious Binaries

IV.6.1 Overview

An original system binary is replaced by a malicious trojan version for this test. This test can demonstrate file monitoring, pattern recognition as well logging tools and for detecting malicious files within a system before they cause damage to it.

IV.6.2 Prerequisites

The `/var/ossec/etc/shared/rootkit_trojans.txt` file in the Wazuh agent comes with a list of common trojan locations and signatures.


```

root@wiss-Lenovo-IdeaPad-S410p: /home/wiss
GNU nano 5.6.1 /var/ossec/etc/shared/rootkit/trojans.txt
# file_name !string_to_search!Description
# Common binaries and public trojan entries
ls !bash|^/bin/sh|dev/[^\u]|\.\tmp|lsfile|duarawkz|/prof|/security|file\.h!
env !bash|^/bin/sh|file\.h|proc\.h|/dev|^/bin/. *sh!
echo !bash|^/bin/sh|file\.h|proc\.h|/dev/[^\u]|^/bin/. *sh!
chown !bash|^/bin/sh|file\.h|proc\.h|/dev/[^\u]|^/bin/. *sh!
chmod !bash|^/bin/sh|file\.h|proc\.h|/dev/[^\u]|^/bin/. *sh!
chgrp !bash|^/bin/sh|file\.h|proc\.h|/dev/[^\u]|^/bin/. *sh!
cat !bash|^/bin/sh|file\.h|proc\.h|/dev/[^\u]|^/bin/. *sh!
bash !proc\.h|/dev/[0-9]|/dev/[hijkz]!
sh !proc\.h|/dev/[0-9]|/dev/[hijkz]!
uname !bash|^/bin/sh|file\.h|proc\.h|^/bin/. *sh!
date !bash|^/bin/sh|file\.h|proc\.h|/dev/[^\u]|^/bin/. *sh!
du !w0rm|/prof|file\.h!
df !bash|^/bin/sh|file\.h|proc\.h|/dev/[^\urdv]|^/bin/. *sh!
login !elite|SuckIT|xlogin|vejeta|porcao|lets_log|sukasuk!
passwd !bash|file\.h|proc\.h|/dev/tty|/dev/[A-Z]|/dev/[b-s,uvxz]!
mngetty !bash|Dimensioni|pacchetto!
chfn !bash|file\.h|proc\.h|/dev/tty|/dev/[A-Z]|/dev/[a-s,uvxz]!
chsh !bash|file\.h|proc\.h|/dev/tty|/dev/[A-Z]|/dev/[a-s,uvxz]!
mail !bash|file\.h|proc\.h|/dev/[^nu]!
su !/dev/[d-s,abuvxz]|/dev/[A-D]|/dev/[F-Z]|/dev/[0-9]|satori|vejeta|conf\.inv!
sudo !satori|vejeta|conf\.inv!
crond !/dev/[^nt]]bash!
gpm !bash|mngetty!
ifconfig !bash|^/bin/sh|/dev/tux|session.null|/dev/[^\cludisopt]!
diff !bash|^/bin/sh|file\.h|proc\.h|/dev/[^n]|^/bin/. *sh!
md5sum !bash|^/bin/sh|file\.h|proc\.h|/dev|^/bin/. *sh!
hdparm !bash|/dev|ida!
ldd !/dev/[^n]|proc\.h|libshow.so|libproc.a!
# Trojan entries for troubleshooting binaries
grep !bash|givemer!
^C Help ^O Write Out ^M Where Is ^K Cut ^T Execute ^C Location M-U Undo M-A Set Mark M-] To Bracket
^X Exit ^R Read File ^I Replace ^P Paste ^J Justify ^_ Go To Line M-B Redo M-C Copy M-_ Where Was

```

IV.6.3 Test Steps

cp -p /usr/bin/w /usr/bin/w.copy : Creating a copy of an original system binary.

Replacing the original system binary with a malicious script:

```

#!/bin/bash echo "'date' this is evil" > /tmp/trojan_created_file
echo 'test for /usr/bin/w trojaned file' >> /tmp/trojan_created_file

#Now running original binary

/usr/bin/w.copy

```

```

root@wiss-Lenovo-IdeaPad-S410p: /home/wiss
root@wiss-Lenovo-IdeaPad-S410p: /home/wiss# cp -p /usr/bin/w /usr/bin/w.copy
root@wiss-Lenovo-IdeaPad-S410p: /home/wiss# echo "'date' this is evil" > /tmp/trojan_created_file
root@wiss-Lenovo-IdeaPad-S410p: /home/wiss# echo 'test for /usr/bin/w trojaned file' >> /tmp/trojan_created_file
root@wiss-Lenovo-IdeaPad-S410p: /home/wiss# /usr/bin/w.copy
 21:25:20 up 9:00, 1 user, load average: 0.91, 0.76, 0.82
USER  TTY  FROM          LOGIN@  IDLE   JCPU   PCPU  WHAT
wiss  tty2  tty2          12:25   8:59m  0.10s  0.09s /usr/libexec/gnome-session-binary --systemd --session=ubuntu
root@wiss-Lenovo-IdeaPad-S410p: /home/wiss# █

```

IV.6.4 Test Results

After executing the previous command, we access the Wazuh Alerts in the Security Events module of the Wazuh Kibana plugin. We can see the alert created in response to the trojan file. Wazuh has successfully detected the suspicious binary.



Figure 92: Graphic of the alerts generated by Wazuh in response to the suspicious binary

The details of the Wazuh alert generated in response to the suspicious binary.

Table 26: Wazuh information generated in response to the suspicious binary

Field Name	Alert Id 510
rule.description	Host-based anomaly detection event (rootcheck)
rule.firedtimes	4
Rule.level	7
rule.groups	Ossec. rootcheck
Full.log	Trojaned version of file '/bin/diff' detected. Signature used: 'bash ^\bin/sh file\.h proc\.h dev [\n]^\bin/.sh' (Generic).
Data.file	usr/bin/diff
Data.title	Trojaned version of file detected

IV.7 Osquery integration

IV.7.1 Overview:

Osquery makes exploring the additional information from the endpoint accessible. Integrating Osquery in wazuh agent can be helpful for threat hinging.

IV.7.2 Prerequisites

An endpoint machine with Osquery installed on it .

IV.7.3 Test steps

No action is required. Wazuh automatically reads the `/var/log/osquery/osqueryd.results.log` file and generates alerts based on these logs.

IV.7.4 Test Results

We access the Wazuh Alerts in the Security Events module of the Wazuh Kibana plugin. We can see the alert created in response that belongs to the Osquery logs.



Figure 93: Graphic of the alerts generated by Wazuh in response to Osquery

The details of the Wazuh alert generated in response to Osquery.

Table 27: Wazuh information generated in response to Osquery

Field Name	Alert Id 24001
Rule.description	Osquery error message
rule.firedtimes	1
Rule.level	5
Rule.groups	osquery
Full.log	E0605 00:03:33.494643 18165 init.cpp:381] osqueryd initialize failed: osqueryd (6679) is already running
location	osquery

IV.8 Detecting malware - Virus Total integration

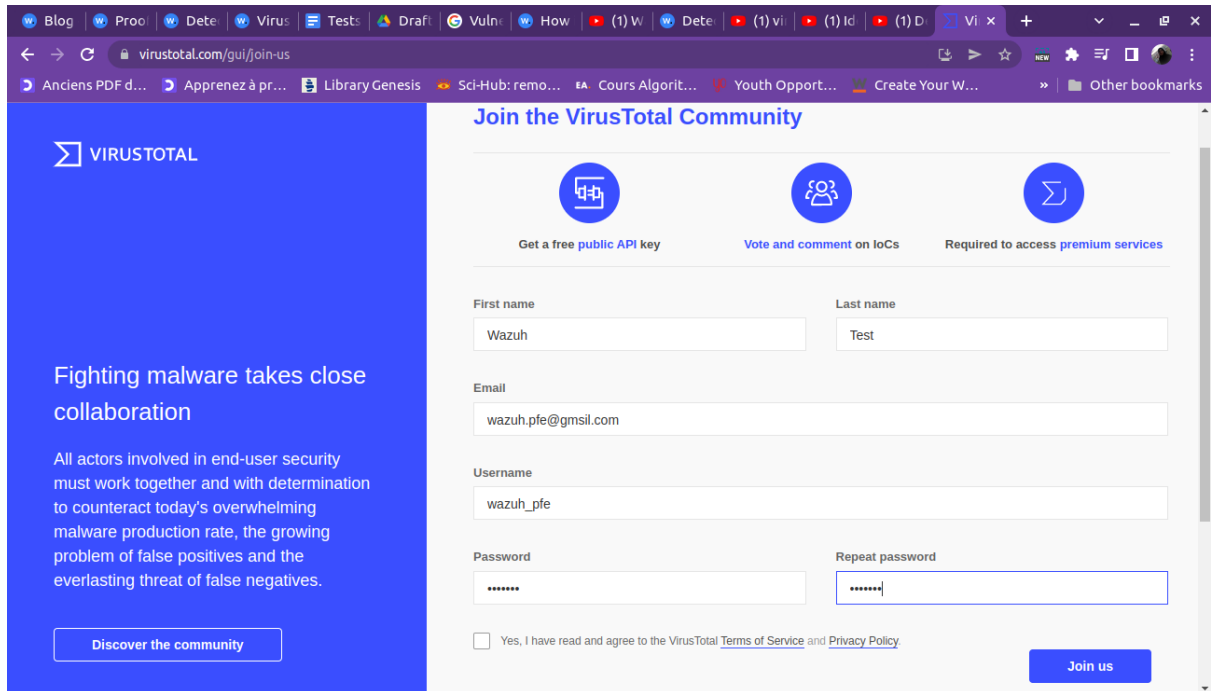
IV.8.1 Overview

The system is vulnerable to different threats, such as viruses and malware...

This test examines the system files and identifies the malicious ones using VirusTotal.

IV.8.2 Prerequisites

- Creating a VirusTotal account to get the API key



* Adding the VirusTotal integration code to the configuration file of the wazuh manager.

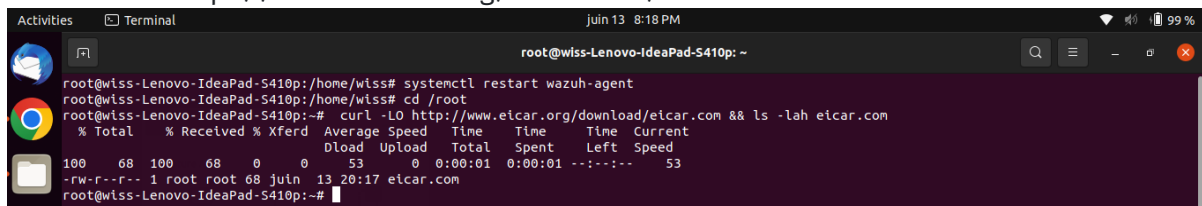
```
<integration>
<name>virustotal</name>
<api_key>6ccb3e9b028329f60e2e65b01488d7c895fad78cc2ed6c3e21d09c6e16b9e63</api_key>
<group>syscheck</group>
<rule_id>100200,100201</rule_id>
<alert_format>json</alert_format>
</integration>
```

IV.8.3 Test steps

In the agent, we download a malicious EICAR test file developed by the European Institute for Computer Antivirus Research (EICAR) and Computer Antivirus Research Organization (CARO) to test the response of computer antivirus programs instead of using real malware.

```
cd /root
```

```
curl -LO http://www.eicar.org/download/eicar.com && ls -lah eicar.com
```



```
root@w1ss-Lenovo-IdeaPad-5410p: /home/w1ss# systemctl restart wazuh-agent
root@w1ss-Lenovo-IdeaPad-5410p: /home/w1ss# cd /root
root@w1ss-Lenovo-IdeaPad-5410p:~# curl -LO http://www.eicar.org/download/eicar.com && ls -lah eicar.com
% Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
           Dload  Upload  Total   Spent    Left   Speed
100  68  100  68  0  0  53  0  0:00:01  0:00:01  --:--:--  53
-rw-r--r-- 1 root root 68 ju1n 13 20:17 eicar.com
root@w1ss-Lenovo-IdeaPad-5410p:~#
```

IV.8.4 Test Results

We access the Wazuh Alerts in the Security Events module of the Wazuh Kibana plugin. We can see the alert created in response that belongs to VirusTotal integration.

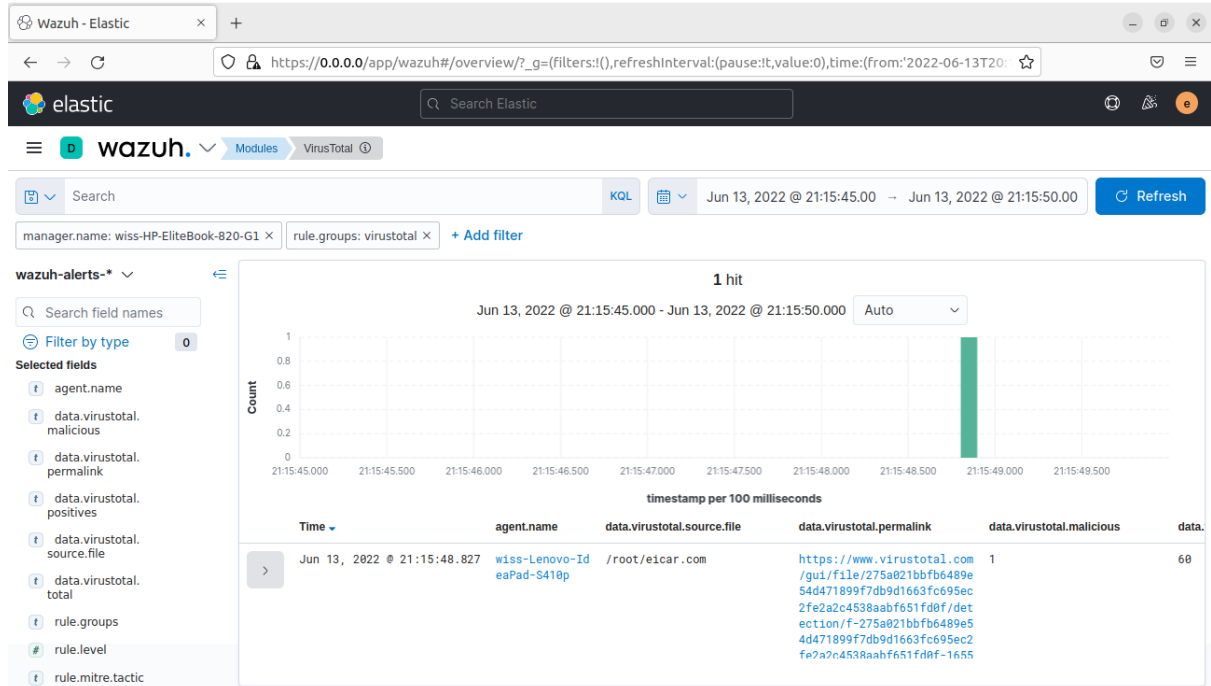


Figure 94: Graphic of the alerts generated by Wazuh in response to malware detection

The details of the Wazuh alert were generated in response to malware detection.

Table 28: Wazuh information generated in response to malware detection

Field Name	Alert Id 87105
rule.description	VirusTotal: Alert - /root/eicar.com - 60 engines detected this file
rule.firedtimes	1
Rule.level	12
rule.mitre.tactic	Execution
rule.mitre.technique	Exploitation for Client Execution
Rule.groups	virustotal
data.virustotal.source.file	/root/eicar.com
data.virustotal.malicious	1
data.virustotal.positives	60
data.virustotal.total	64
location	VirusTotal

Conclusion

Since our SIEM can detect the attack attempts that we have made in our environment and the events that happen in our systems, such as unauthorised processes, SQL injections, and malware.. The project covers the organization's needs and is ready to be implemented in their environment.

Conclusion and future work



This final project focuses on the design and the implementation of a SIEM security information and event management tool for monitoring the audit logs from different sources; in our case, Windows and Ubuntu systems; thus, the trace obtained can be used as evidence against an attacker., BADR Bank can use that for monitoring its networks; in this step, it is clear that the problem raised at the outset has been solved.

This work was an opportunity for us as students in Information System Security to see a part of the professional world and implement the theoretical knowledge acquired during our course. From there, we have identified the network and systems' various critical points and security vulnerabilities.

Our work has been designed with open-source tools ELK and Wazuh, allowing us better flexibility and adaptation to the organization's needs.

Our solution could be enriched by several improvements that can be made:

- The creation of an anomaly detection model using machine learning. We already mentioned it in the theoretical part. Still, for the limited time, we could not implement it because the tools used in our solution contain anomaly detection, but it is a premium feature.
- Using a distributed deployment instead of a single host deployment.
- Improved dashboard by adding graphs that give more information to the user.
- Configuration of security policies to be more adapted to the needs of the BADR.

Bibliography

- Al-Shaer, R., Spring, J. M., & Christou, E. (2020). Learning the Associations of MITRE ATT&CK Adversarial Techniques. *2020 IEEE Conference on Communications and Network Security, CNS 2020*. <https://doi.org/10.48550/arxiv.2005.01654>
- Apache Lucene : tutoriel - IONOS*. (n.d.). Retrieved June 16, 2022, from <https://www.ionos.fr/digitalguide/serveur/configuration/apache-lucene/>
- BADR banque - 318 Mots | Etudier*. (n.d.). Retrieved June 16, 2022, from <https://www.etudier.com/dissertations/Badr-Banque/53162990.html>
- Białecki, A., Muir, R., & Ingersoll, G. (n.d.). *Apache Lucene 4. Case Solution*. (n.d.). Retrieved June 16, 2022, from https://success.trendmicro.com/dcx/s/solution/TP000086250?language=en_US
- Clarence Chio, O'Reilly Media, & David Freema. (2018). *Machine Learning and Security: Protecting Systems with Data and Algorithms*.
- Components - Getting started with Wazuh · Wazuh documentation*. (n.d.). Retrieved June 16, 2022, from <https://documentation.wazuh.com/current/getting-started/components/>
- D. Carr. (2005). *Primer: security information and event management*.
- di Sarno, C., Garofalo, A., Matteucci, I., & Vallini, M. (2016). A novel security information and event management system for enhancing cyber security in a hydroelectric dam. *International Journal of Critical Infrastructure Protection*, 13, 39–51. <https://doi.org/10.1016/j.ijcip.2016.03.002>
- Dorsey, S., de Marchi, S., Feaver, A. P., Balcells, L., & Wibbels, E. (2020). *Machine Learning and Security Studies*.
- DPCASSI-ORGANIGRAMME*. (n.d.).
- Event correlation | Vladimir Potapov Blog*. (n.d.). Retrieved June 16, 2022, from <https://vpotapov.wordpress.com/2016/11/07/correlation-in-siem/>
- González-Granadillo, G., González-Zarzosa, S., & Diaz, R. (2021). Security information and event management (SIEM): Analysis, trends, and usage in critical infrastructures. *Sensors*, 21(14). <https://doi.org/10.3390/s21144759>
- He, S., Zhu, J., He, P., & Lyu, M. R. (2020). *Loghub: A Large Collection of System Log Datasets towards Automated Log Analytics*. <http://arxiv.org/abs/2008.06448>

- Interrogez efficacement vos bases de données avec Elasticsearch.* (n.d.). Retrieved June 16, 2022, from <https://www.data-transitionnumerique.com/search-elasticsearch/>
- INTRODUCTION TO ELK STACK.* (n.d.). Retrieved June 18, 2022, from <https://www.oracle.com/technetwork/java/javase/>
- Ioannis Voulgaris. (2020). *Security in Digital Systems.*
- Johnson, H. A. (2018). Slack. In *Journal of the Medical Library Association* (Vol. 106, Issue 1, pp. 148–151). Medical Library Association.
<https://doi.org/10.5195/jmla.2018.315>
- Liang, W., Li, W., & Feng, L. (n.d.). *SPECIAL SECTION ON RELIABILITY IN SENSOR-CLOUD SYSTEMS AND APPLICATIONS (SCSA) Information Security Monitoring and Management Method Based on Big Data in the Internet of Things Environment.*
<https://doi.org/10.1109/ACCESS.2021.3064350>
- Logstash - Internal Architecture.* (n.d.). Retrieved June 16, 2022, from https://www.tutorialspoint.com/logstash/logstash_internal_architecture.htm
- Mattia Incoronato. (2020). *Wazuh The Open Source Security Platform.*
<https://computerscience.unicam.it/marcantoni/tesi/Wazuh%20-%20The%20Open%20Source%20Security%20Platform.pdf>
- Phillipe Martinet. (2006). *Mise en oeuvre d'un prototype d'architecture OSSIM.*
- Présentation de la banque BADR - CAW JIJEL.* (n.d.). Retrieved June 16, 2022, from <https://www.cawjijel.org/fr/accueil/financement-agricole/87-presentation-de-la-banque-badr>
- Rajendra Kumar Dwivedi, Arun Kumar Rai, & Rakesh Kumar. (2020). *10th International Conference on Cloud Computing, Data Science & Engineering : proceedings of the Confluence 2020 : 29-31 January 2020, Amity University, Uttar Pradesh, India.* (Rajendra Kumar Dwivedi, Arun Kumar Rai, & Rakesh Kumar, Eds.).
- Sagan User Guide Documentation Release 1.2.2 Champ Clark III.* (2022).
- Security Configuration Assessment (SCA) · Wazuh · The Open Source Security Platform.* (n.d.). Retrieved June 16, 2022, from <https://wazuh.com/blog/security-configuration-assessment/>
- Set up X-Pack | Elasticsearch Guide [7.17] | Elastic.* (n.d.). Retrieved June 16, 2022, from <https://www.elastic.co/guide/en/elasticsearch/reference/7.17/setup-xpack.html>
- Stoleriu, R., Puncioiu, A., & Bica, I. (2021). Cyber Attacks Detection Using Open Source ELK Stack. *Proceedings of the 13th International Conference on Electronics,*

- Computers and Artificial Intelligence, ECAI 2021*.
<https://doi.org/10.1109/ECAI52376.2021.9515120>
- Syslog - Definition and Details*. (n.d.). Retrieved June 16, 2022, from
<https://www.paessler.com/it-explained/syslog>
- The SIEM advantage*. (n.d.). Retrieved June 18, 2022, from <http://bit.ly/M2b9Yy>
- Threat landscape for industrial automation systems. Statistics for H2 2021 | Kaspersky ICS CERT*. (n.d.). Retrieved June 19, 2022, from <https://ics-cert.kaspersky.com/publications/reports/2022/03/03/threat-landscape-for-industrial-automation-systems-statistics-for-h2-2021/>
- Tsunoda, H., & Keeni, G. M. (2014). Managing syslog. *APNOMS 2014 - 16th Asia-Pacific Network Operations and Management Symposium*.
<https://doi.org/10.1109/APNOMS.2014.6996575>
- Vazão, A., Santos, L., Piedade, M. B., & Rabadão, C. (2019). SIEM open source solutions: A comparative study. *Iberian Conference on Information Systems and Technologies, CISTI, 2019-June*. <https://doi.org/10.23919/CISTI.2019.8760980>
- Wazuh · The Open Source Security Platform*. (n.d.). Retrieved June 16, 2022, from
<https://wazuh.com/>
- Wazuh agent - Components · Wazuh documentation*. (n.d.). Retrieved June 16, 2022, from
<https://documentation.wazuh.com/current/getting-started/components/wazuh-agent.html>
- What are Beats? | Beats Platform Reference [8.2] | Elastic*. (n.d.). Retrieved June 16, 2022, from <https://www.elastic.co/guide/en/beats/libbeat/current/beats-reference.html>