Improved Log Monitoring UsingHost-based Intrusion Detection System

Mr. James Sani

M.Sc. in Electronics and Communications, Project Officer, BCN NIG LTD, Nigeria

Abstract

There has been an issue ever since computers were invented, security. Evolving in technology is evolving cybercriminals. Hackers are becoming knowledgeable day by day making it hard for us to handle the security of our machines. Then, the computer-generated records were generated and they are being analyzed to find any anomalies in the computers. These records generated by devices, applications, and networks are then time-sequenced into logs. A log is a record of an event that has occurred in a machine. These logs are not fully documented or consistently formed across applications or devices which can be understood by professionals working with logs, which further builds the case for log analytics. Log analytics helps a normal user get insights into system performance and can indicate possible problems such as hardware failure or any security breaches. To do this, a powerful technique to monitor the computers is used, namely an intrusion detection system. A sub-technique called host-intrusion detection system is used to understand more in detail about a machine, server, or computer. These logs are collected, analyzed, and put out as stats in the form of UI in web applications. This helps the user understand what is happening and to overcome and prevent security issues in an organization.

Keywords: wazuh, network, security, attack, monitoring

INTRODUCTION

It has been zero minutes since the last security breach. There has been an evident rise in the number of cybercrimes happening day by day. Once patched a new loophole will be found, thus making it hard for the security analysts and security engineers to recognize and stop the breaches from happening. Organizations are facing an increased number of cyber threats ranging from financially motivated and opportunistic malware to more advanced targeted attacks.

This can be taken care of by using an intrusion detection system. In specific, the sub-type of IDS known as Host Intrusion Detection System is used.

It can help identify all the events happening on a computer, laptop, or server. Thus, helping in getting an insight into what all are happening in a machine. HIDS can help collect the log information of what is happening in a machine. It is hard to analyze the logs in their original raw format. In order to help the user in getting a clear understanding of what logs are and what information they carry a Host intrusion detection system (HIDS) is integrated with a web app.

OVERVIEW OF WAZUH AND EXPERIMENT SETUP

Wazuh is an open-source security solution that provides threat detection, visibility, and compliance management. It is designed to help organizations of all sizes detect, respond to, and recover from security incidents.

Wazuh collects and analyzes security-related data from multiple sources such as logs, configuration files, registry keys, and system events in real time. It correlates the events to identify potential security threats and generates alerts. The solution also provides automated responses to mitigate and contain security incidents.

Wazuh's architecture includes agents that are installed on servers, endpoints, and cloud instances to collect security data. A central manager receives and analyzes the data, while an API allows users to access and manage the security data programmatically. Wazuh uses the Elastic Stack, which includes Elasticsearch, Logstash, and Kibana, to store, process, and visualize the security data.

A. Wazuh Architecture:

Wazuh architecture consists of several components that work together to achieve its objectives:

Agents:Wazuh agents are installed on servers, endpoints, and cloud instances to collect security-related data such as logs, configuration files, registry keys, and system events. The agents normalize and encrypt the data before sending it to the central manager.

Manager: The central manager receives and analyzes the data from agents. It correlates the events to identify potential security threats and generates alerts. The manager can be deployed on-premises, in the cloud, or as a hybrid solution.

API: Wazuh provides an API allowing users to programmatically access and manage the security data. The API can be used to integrate Wazuh with other security tools or custom applications.

Elastic Stack: Wazuh uses the Elastic Stack, which includes Elasticsearch, Logstash, and Kibana, to store, process, and visualize the security data. Elasticsearch provides a scalable and distributed search and analytics engine, Logstash is used for data processing and transformation, and Kibana is a web-based user interface for data visualization and dashboarding.

Ruleset: Wazuh ruleset is a collection of security rules that define the behavior of the system and the events to be monitored. The ruleset can be customized to meet the specific security needs of an organization.

Integrations: Wazuh integrates with various security tools and platforms such as threat intelligence feeds, vulnerability scanners, and security information and event management (SIEM) systems. This enables organizations to have a holistic view of their security posture and respond to threats more effectively.

Overall, Wazuh architecture provides a comprehensive and scalable approach to security by leveraging agents, a central manager, APIs, Elastic Stack, rulesets, and integrations.

B. Wazuh provides the following capabilities:

Security analytics: Security analytics refers to the process of analyzing data from various security sources, such as network logs, system logs, and security devices, to detect and respond to security threats. The goal of security analytics is to identify patterns and anomalies in data that may indicate a security breach or attack.

Intrusion Detection: Intrusion detection is the process of monitoring computer networks or systems to detect unauthorized access, misuse, or other malicious activity. The goal of intrusion detection is to identify security threats and take appropriate action to prevent or mitigate the impact of an attack.

Log Data Analysis: Log data analysis is the process of examining and analyzing data generated by computer systems and applications, typically stored in log files. Log data contains information about system events, user activity, and other important data that can be used to monitor and troubleshoot systems, detect security threats and gain insights into system performance.

Vulnerability Detector: A vulnerability detector is a type of software tool that is used to scan computer systems and applications for known vulnerabilities that could be exploited by attackers. Vulnerability detectors typically use a database of known vulnerabilities, such as the Common Vulnerabilities and Exposures (CVE) database, to identify vulnerabilities that are present in the system being scanned.

Configuration Assessment: Configuration assessment is the process of evaluating the security configuration of computer systems and applications to ensure that they are properly configured and secured. Configuration assess ment involves reviewing system settings, policies, and procedures to identify potential vulnerabilities and areas of weakness.

C. Components of Wazuh:

Wazuh agent: These are the endpoints in the wazuh architecture. These are installed in the machines to be monitored for intrusions. They help in securing the machine by providing threat prevention, detection, and response capabilities.
Wazuh server: A server is used to manage the agents that are installed on the machines to be monitored. They analyze the data received from the wazuh agents. It then processes it through various decoders and rules. This is done by using threat intelligence to know if any compromise has happened to the network. One server can analyze data from a huge number of agents. These can be scaled horizontally when set up as a cluster. A server is of two different types a master and a worker. A master can be used to monitor the workers and balance the load. A worker is used to monitor the various agents and gather information for monitoring.

Elastic Stack:It is a group of various products that are used to gather the logs from the manager or workers and forward them for indexing using the filebeat tool. The forwarded logs from the filebeat are collected at the elastic search where the logs are indexed and stored. Using these logs that are indexed are shown in the form of UI with the help of Kibana.

D. Other Components:

Auditd: It is a daemon in Linux and is used for logging events related to file integrity monitoring, killing a process, or creating a network connection. It is a Linux kernel feature that logs system calls such as opening a file, killing a process, or creating a network connection. It must be configured to decrease the wastage of memory due to unnecessary logging of less-priority data. This can be done by configuring the audit rules. Auditd can be configured to meet specific auditing requirements, such as tracking specific system calls or logging user activity on specific files or directories. It is also flexible in terms of data storage and can be configured to log data to a local file system, remote server, or database. One of the key benefits of using auditd is that it provides detailed information about system activity, making it easier to identify potential security breaches or unauthorized activity. This information can also be used for system troubleshooting and performance tuning.

FileBeat: Filebeat is an open-source log data shipper for the Elastic Stack. It is a lightweight agent that is installed on servers, shipping logs, and other data to Elasticsearch for centralization and analysis. Filebeat provides real-time data collection and forwarding capabilities and is designed to handle large volumes of logs from multiple sources. Filebeat works by reading files and sending data to the output, which can be Elasticsearch, Logstash, or another output. It is also able to tail files and send new lines as they are written, making it suitable for real-time monitoring. Filebeat also has a powerful configuration file, which allows administrators to define specific files to monitor, configure how data is sent, and control the behavior of the agent. This makes it easy to customize Filebeat for different use cases, such as processing logs from different sources and sending data to different outputs. In conclusion, Filebeat is an essential tool for centralized log management and analysis, allowing administrators to gather, process and analyze log data from multiple sources in real time.

Elastic Search: Elasticsearch is a versatile and scalable search engine that can be used to handle a wide range of use cases. Its combination of fast search, powerful analytics, and easy integration makes it an ideal choice for organizations with large and complex data requirements. With its ability to scale and handle huge amounts of data, it is a popular choice for organizations with large datasets and complex search requirements. Elasticsearch provides near real-time search and analytics capabilities, allowing users to search and analyze massive amounts of data quickly and efficiently. It supports a wide range of query languages and APIs, making it easy to integrate into existing systems. The engine's powerful query language, combined with its ability to index and analyze structured and unstructured data, makes it ideal for applications such as log analytics, business intelligence, and content search. Additionally, Elasticsearch can be easily scaled horizontally, allowing for the addition of more nodes to the cluster as the size of the data set grows.

Kibana: Kibana is an open-source data visualization and exploration platform used for analyzing and visualizing data stored in Elasticsearch. It provides a web-based user interface for creating and sharing dynamic dashboards, histograms, pie charts, heat maps, and other visualizations of data. Kibana is commonly used for log analysis, security analysis, and business intelligence. With Kibana, you can create custom dashboards to monitor metrics and trends, perform ad-hoc analysis, and identify correlations in your data. It also provides powerful search capabilities, enabling you to easily find the information you need. Kibana is used across a wide range of industries and applications, including log analysis, security analytics, performance metrics, and business intelligence. It is highly customizable, so you can tailor it to your specific needs and workflows. And, because it is open source, it has a large and active community of users who contribute to its development and provide support and resources.

A. MITRE ATT&CK

MITRE ATT&CK is a framework for describing the tactics and techniques used by attackers during a cyber-attack. ATT&CK stands for Adversarial Tactics, Techniques, and Common Knowledge, and it was created by the MITRE Corporation. The framework is used to organize and describe the steps that an attacker may take during a cyber-attack, and it is designed to help defenders understand and defend against these attacks.

MITRE ATT&CK also includes a matrix that maps the tactics and techniques to specific stages of the cyber kill chain, which is a model that describes the different stages of a cyber-attack, from the initial reconnaissance to the final exfiltration of data.

The MITRE ATT&CK framework is widely used in the cybersecurity industry, and it is an important tool for defenders to understand the tactics and techniques that attackers may use, so they can develop effective defense strategies.

B. ELK Architecture:

The ELK stack is a popular open-source software suite used for log management and analytics. It consists of three main components:

Elasticsearch - a distributed search engine that stores and indexes data

Logstash - a data processing pipeline that ingests, processes, and transforms log data

Kibana - a web-based user interface that provides visualization and analysis capabilities for data stored in Elasticsearch.

In a typical ELK architecture, log sources send log data to the Logstash server(s) via a variety of input plugins. Logstash then processes and transforms the data using a variety of filters and outputs the data to Elasticsearch for indexing and storage. Finally, Kibana provides a web-based user interface for searching and visualizing the log data stored in Elasticsearch.

The ELK stack provides a scalable, flexible, and cost-effective solution for log management and analysis that can be customized to meet the needs of different organizations.

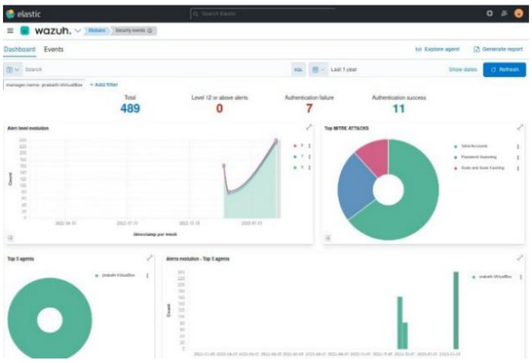
EXPERIMENT, RESULTS, AND SYSTEM PERFORMANCE

In most cases, the attacks that have occurred will be saved as log messages in that system, device, or application. With the help of wazuh, the user can manage and analyze those logs and hence can perform a much better threat detection. The machine which is being monitored will have a wazuh agent installed and it is responsible for reading log messages inthe operating system of the monitored machine. The agent then forwards all the required information to the Wazuh server, where the analysis is performed. With the help of decoders, wazuh can decode all the information from the log and identify the source where it came from. Then with all the information that is collected analysis is performed.

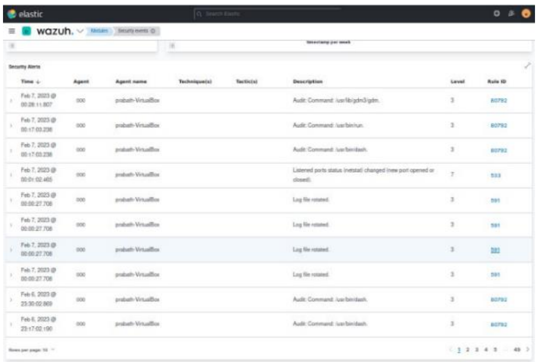
Whenever an event occurs the agent collects it and the manager generates an alert if it matches a rule that has a priority level higher than the predefined threshold. If an alert is generated then the wazuh server sends those alerts to the wazuh indexer component where they are stored and indexed. Then with the help of Wazuh dashboard, we can search, analyze, and visualize the data.

Below is an example screenshot of the interface:

Dashboard:

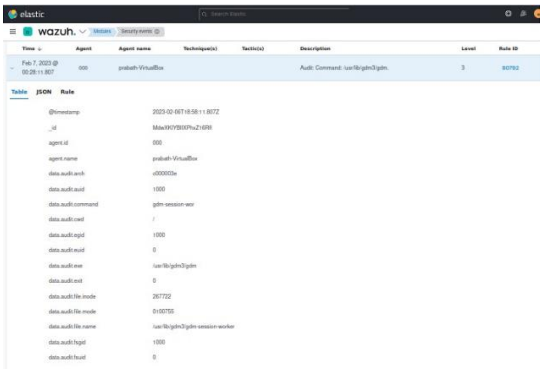


Security Alerts:



Time ↓	Agent	Agent name	Technique(s)	Tactic(s)	Description	Level	Rule ID
Feb 7, 2023 @ 00:28:11.807	000	prabath-VirtualBox			Audit: Command: /usr/lib/gdm3/gdm.	3	80792
Feb 7, 2023 @ 00:17:03.238	000	prabath-VirtualBox			Audit: Command: /usr/bin/run.	3	80792
Feb 7, 2023 @ 00:17:03.238	000	prabath-VirtualBox			Audit: Command: /usr/bin/dash.	3	80792
Feb 7, 2023 @ 00:01:02.465	000	prabath-VirtualBox			Listened ports status (netstat) changed (new port opened or closed).	7	533
Feb 7, 2023 @ 00:00:27.708	000	prabath-VirtualBox			Log file rotated.	3	591
Feb 7, 2023 @ 00:00:27.708	000	prabath-VirtualBox			Log file rotated.	3	591
Feb 7, 2023 @ 00:00:27.708	000	prabath-VirtualBox			Log file rotated.	3	591
Feb 7, 2023 @ 00:00:27.708	000	prabath-VirtualBox			Log file rotated.	3	591
Feb 6, 2023 @ 23:30:02.869	000	prabath-VirtualBox			Audit: Command: /usr/bin/dash.	3	80792
Feb 6, 2023 @ 23:17:02.190	000	prabath-VirtualBox			Audit: Command: /usr/bin/dash.	3	80792

Details about the log:



Time ↓	Agent	Agent name	Technique(s)	Tactic(s)	Description	Level	Rule ID
Feb 7, 2023 @ 00:28:11.807	000	prabath-VirtualBox			Audit: Command: /usr/lib/gdm3/gdm.	3	80792

@timestamp	2023-02-06T18:58:11.807Z
_id	MdwXKIYBIXPhxZ16RR
agent.id	000
agent.name	prabath-VirtualBox
data.audit.arch	c000003e
data.audit.auid	1000
data.audit.command	gdm-session-wor
data.audit.cwd	/
data.audit.egid	1000
data.audit.euid	0
data.audit.exe	/usr/lib/gdm3/gdm
data.audit.exit	0
data.audit.file.inode	267722
data.audit.file.mode	0100755
data.audit.file.name	/usr/lib/gdm3/gdm-session-worker
data.audit.fsgid	1000
data.audit.fsuid	0

A. Result:

Data analysis of the log is performed and all the important information’s shown in a better manner so that when an alert is generated, we can have a better understanding of the issue and fix it as soon as possible.

B. System Performance:

The Wazuh manager that is used is installed on an Ubuntu virtual machine which is assigned with these specifications:

8GB of RAM (Random Access Memory),

4 cores with 30GB of storage.

CONCLUSION

This paper aims to perform threat detection by analyzinglogs in a better way. In today's world, all web servers and web applications are connected over the internet and can be exposed if they are not well protected. Hence to protect the host from cyber-attacks and threats we are building a host intrusion detection system in the host which can detect and prevent attacks and also give alerts to the user. This host intrusion detection system is built with the help of Wazuh.

The current tool can monitor all the things happening in the host or the manager machine. It was optimized to reduce the number of logs collected. It could generate raw forms of logs to monitor and analyze.

All the different types of cyber-attacks and threats are presented and the well-knownattacks are being monitored with special attention. Hence, any host with the host-based intrusion detection system installed is well protected from cyber-attacks and threats.

If we integrate the Wazuh tool with antivirus software we will be able to get an even more deep inspection of viruses.

REFERENCES

1. Wu, Yafeng, et al. "Paradise: real-time, generalized, and distributed provenance-based intrusion detection." IEEE Transactions on Dependable and Secure Computing 20.2 (2022): 1624-1640.
2. Kebede, Nibretu, and Gebeyehu Belay Gebremeskel. "In-Depth Analysis of Combine Machine Learning and Open Source Security Tools to Enhance Host-Based Intrusion Detection." (2022).
3. Teixeira, Diogo, Silvestre Malta, and Pedro Pinto. "A Vote-Based Architecture to Generate Classified Datasets and Improve Performance of Intrusion Detection Systems Based on Supervised Learning." Future Internet 14.3 (2022): 72.
4. Wani, Anwaar Ahmad, Juneed Iqbal, and Mudassir Makhdoomi. "Modelling an Intrusion Detection system using ensemble approach based on voting to improve accuracy of base classifiers." Journal of Algebraic Statistics 13.2 (2022): 1844-1865.
5. Ahmet, E. F. E., and İrem Nur ABACI. "Comparison of the host based intrusion detection systems and network based intrusion detection systems." Celal Bayar University Journal of Science 18.1 (2022): 23-32.
6. Gupta, Rajeev Kumar, et al. "Improving collaborative intrusion detection system using blockchain and pluggable authentication modules for sustainable Smart City." Sustainability 15.3 (2023): 2133.
7. Martin Grimmer, Martin Max Röhling, D Kreusel, and Simon Ganz. 2019. A modern and sophisticated host based intrusion detection data set. IT-SicherheitalsVoraussetzung für eineerfolgreicheDigitalisierung (2019), 135—145
8. J. Chandler, "Evaluating Open-Source HIDS with Persistence Tactic of MITREatt&ck", SANS Institute, 2021
9. Mulyadi, L. A. Annam, R. Promya and C. Charnsripinyo, "Implementing Dockerized Elastic Stack for Security Information and Event Management", 5th International Conference on Information Technology, 2020.
10. O. Negotia, M. Carabas "Enhanced Security Using Elastic Search and MachineLearning", Advances in Intelligent Systems and Computing, vol. 1230, July, 2020.

11. Jyothsna V. and Rama Prasad V. V. 2011 A Review of Anomaly based IntrusionDetection Systems International Journal of Computer Applications
12. J. Chandler, "Evaluating Open-Source HIDS with Persistence Tactic of MITRE att&ck", SANS Institute, 2021.
13. Sharma, V. (2016). Getting Started with Kibana. In: Beginning Elastic Stack. Apress, Berkeley, CA.
14. M. Moh, S. Pininti, S. Doddapaneni, T.S. Moh "Detecting Web Attacks Using Multi-Stage Log Analysis", 6th IEEE International Conference on Advanced Computing,2016
15. Mvula, Paul K., et al. "Evaluating Word Embedding Feature Extraction Techniques for Host-Based Intrusion Detection Systems." Discover Data 1.1 (2023): 2.