International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 12 Issue: 01 | Jan 2025 www.irjet.net p-ISSN: 2395-0072

Enhancing Security Information and Event Management Systems in Cybersecurity Using Artificial Intelligence

Saba Gour[1]

[1] *Software Engineer, Seattle, USA*

Abstract – *Cybersecurity has always been important to an organization to ensure its smooth and healthy functioning. However, in recent years, cybersecurity has become increasing important due to an exponential increase in security threats. Today, we hear of critical systems in hospitals, banks, retail being hijacked by malicious actors for financial gains and enterprises are left helpless with no real choice but to concede to the demands of the attackers. With the advent of artificial intelligence (AI), attackers now use sophisticated techniques and tactics to breach systems and are successful in causing a lot of harm. Thus, it has become very important to enhance cybersecurity strategies and use artificial intelligence to improve the prevention, detection, mitigation and resolution of security threats. Applying AI correctly to cybersecurity gives us the much-needed edge to combat security attacks. Security Operations Center (SOC) is the team within an organization whose sole function is to respond to security incidents. The SOC team uses different tools; however, a Security information and event management (SIEM) system is a critical tool used for cybersecurity. We examine an SIEM system, which is a central system that collects logs from various sources within the network, correlates and analyzes the logs and uses it for efficient monitoring and alerting. The SOC team relies heavily on the SIEM system to detect Indications of Compromise (IoC) and if a compromise is detected then the Incident Response Team (IRT) is engaged. Through a comprehensive analysis, this paper not only provides insights into SIEM systems, their different sources of data, their functions and challenges but also anticipates future trends and developments in the field.*

Key Words: cybersecurity, security information and event management (SIEM), artificial intelligence (AI), machine learning (ML), false positive alerts, correlation rules, security operations team (SOC), information technology (IT)

1.The main functions of SIEM systems

The main purpose of an SIEM system is to collect and correlate logs from different sources in real time, set up monitoring and alerting which are then used for incident response and, compliance and auditing.

1.1 Collecting logs

The logs are collected by the SIEM system from different network devices like switches, routers, firewalls, network segments, load balancers, domain name system (DNS) servers/resolvers, hypervisors, hosts, virtual machines (VMs), intrusion detection systems (IDS), intrusion prevention systems (IPS), endpoint protection (EPP) tools, and data loss prevention (DLP) tools.

Switches, routers: Switch is a network component which ties together different devices like computers, printers, storage servers in a small setting like a home or a small office. Router is a network component which will connect multiple switches together to connect networks in a single location or across multiple locations to form a larger network. Router is an interface through which traffic flows between the network and the internet. A larger network is broken up into smaller subnetworks, it is called network segmentation to ensure smooth network administration and enhanced security which could be on-premises or in the cloud.

Firewalls: Firewall is a network component that will control the incoming and outgoing traffic between your network and the internet. It will protect your network from malicious actors based on certain predefined security rules. Larger networks benefit from having standalone firewalls for enhanced protection, whereas in your home a firewall is integrated into the router. A firewall is the first line of defense for any network and is often intruded by attackers.

Load balancers: Load balancer is a network device that distributes network or application traffic across multiple hosts/servers for balancing capacity, improving response time and combating network latency. Mostly, high volume customer facing systems are load balanced. They help in diverting malicious traffic to specific hosts, keeping it away from the main network, which helps in the case of distributed denial of service (DDoS) attacks.

DNS servers/resolvers: Domain name system (DNS) is the phonebook of the internet. It transforms a domain into an IP address so that it can be understood by the browser and the correct pages can be loaded. A DNS recursive resolver is the first device that receives the request and will recursively look for the IP address until it reaches the authoritative DNS server which is the final stop, and it returns the IP address to the recursive resolver that initially made the request. Many times, DNS servers are targeted by attackers that want to divert the traffic to malicious IP addresses.

Hypervisors, hosts, virtual machines (VMs): Hypervisor is the virtualization software or firmware that creates virtual machines on a physical machine. It is the underlying component that manages memory, CPU usage,

storage and isolation of virtual machines. A host can be a physical or virtual machine. Attackers target the hypervisor to compromise the entire machine. By having control of the hypervisor, they can inflate the memory and CPU resources used by a single virtual machine, causing the other virtual machines to crash because resource management across the VMs is also done by the hypervisor. An attacker getting control of the hypervisor can be dangerous.

Intrusion detection systems (IDS), intrusion prevention systems (IPS): IDS and IPS are most used together and are the same set of tools. They constantly monitor the network for intrusive traffic and threats, they are used for reporting and alerting the security operations center (SOC) team. Thus, the logs of the IDS and IPS are crucial.

Endpoint protection (EPP) tools: It is a security solution that protects endpoint devices like phones, laptops, computers within an organization. This set of tools will alert in case of abnormalities on devices like the existence of malware and multiple incorrect password attempts causing the device to get locked.

Data loss prevention (DLP) tools: It is a set of security tools aimed at protecting the sensitive data in an organization from unauthorized access, unintentional sharing and information leaks. A DLP system can have a database of hashes for files that contain trade secrets or crucial information regarding internal applications. A hash is a value derived by applying certain cryptographic formulae and it uniquely identifies a file or resource. Any manipulation done to the file, will change the original hash value and is widely used to detect tampering. When a data transfer is detected, the DLP can check against this database to verify the sensitivity of the information being shared and if needed it can stop the transfer and alert the security team immediately. A DLP solution will work best when sensitivity of data within an organization is clearly determined.

1.2 Correlating logs

The logs are correlated in real time using correlation rules established within the SIEM system by the organization. This is an important operation to derive meaningful metrics and trace the path and actions of an attacker across multiple networks and devices in the organization in relation to time. This gives the organization a chance to define security thresholds and priorities according to the criticality of their business operations. For example, if some systems are more critical than others then correlation rules can be defined to protect more networks and hosts that support those systems and applications. Applications that support financial transactions are a good example and rules can be defined to closely track financial activity within the organization and with external parties.

1.3 Monitoring and alerting

Using the correlated logs of the SIEM system, monitoring and alerting is set up. This is crucial for the security operations centre (SOC) team to look for anomalies and threats and respond to them as soon as possible. Alerting is based on certain thresholds and commonly seen errors or warnings in the logs in case of various risk situations which are then investigated and remediated by the security analysts and the incident response team.

1.4 Compliance and security auditing

The SIEM system helps in maintaining compliance with rules and regulations within an organization and the country in which the organization is operating. For example, medical records are protected by Health Insurance Portability and Accountability Act (HIPAA) in America which means that patient records can only be accessed by authorized individuals and any information breach needs to be reported to the concerned authorities. Similarly, there is General Data Protection Regulation (GDPR) law to protect the personal data of European Union (EU) citizens. If a data breach happens to occur, then the security team would need to check and report the breach within 72 hours to concerned authorities. Security audits are essential in assessing the health of the IT infrastructure of an organization, SIEM logs are very useful in carrying out security audits because by studying the logs, one would understand the measures put in place to make the systems resilient to threats.

2. Challenges of SIEM systems

SIEM systems face many challenges that have been prevalent for a while like receiving high volumes of data, integration complexity, scalability, difficulty in optimizing correlation rules, blind spots in threat detection, and an important one is false positives.

Large amounts of data: An SIEM system receives large volumes of data from different sources which leads to many alerts being generated and clogging the inbox of the security analysts. The security analysts are then sifting through these emails causing them to get fatigued, this is called alert fatigue. It negatively impacts the productivity and functioning of the security team as many of these alerts could be irrelevant and redundant.

Complexity in integrating different sources of data: There are different components within a network and within the IT infrastructure of an organization. Each of these components needs to be integrated with an SIEM system to ensure security, but it is difficult to integrate different sources of data with an SIEM system. Some systems remain unintegrated, and the security team needs to monitor and set up alert mechanisms independently for these components. The security analysts need to check these components separately thus increasing the scope of their work and the time needed to troubleshoot issues. This can have a

significant impact when the organization is under a cyber-attack and mitigating the risk quickly is of high importance.

Blind spots in threat detection: Not being able to collect data from different sources, can lead to blind spots in threat detection and leave certain components and devices vulnerable to attacks. Every organization would like to reduce the attack surfaces of their networks however due to blind spots it comes difficult.

Optimizing correlation rules: The data that is received from different sources is not in the same format. Many a times, this data is malformed and unstructured. Applying correlation rules to such data is very challenging. Modifying these rules depending upon the situation cannot be easily achieved.

False positive alerts: Poorly defined correlation rules and loosely defined thresholds can lead to many false positives. Security analysts being paged in the middle of the night to triage a false positive threat will burn out and will eventually get desensitized to alerts coming from real threats. This is one of the major issues that a security operation center (SOC) team faces within an organization.

Difficulty in scaling: It is difficult to scale an SIEM system due to huge volumes of unstructured data it receives. If an organization has an SIEM solution that is on-premises, then scaling it would require a major overhaul of their infrastructure.

3. Applying artificial intelligence to solve the challenges of SIEM systems

An artificial intelligence (AI) system driven by machine learning (ML) is powerful at pattern recognition. Machine learning is the core technology behind pattern recognition where large amounts of data are analyzed to identify complex patterns which would not be possible for humans to do. For example, natural language processing (NLP) is a result of machine learning algorithms and linguistics. NLP helps a computer to understand human language. Thus, AI that is powered by machine learning can perform text analysis on different kinds of data including human speech which makes it incredible at exposing vulnerabilities and, detecting abnormalities, threats, frauds and incidences of compromise (IoC) in cybersecurity. An AI system placed in front of an SIEM solution can take away many of the challenges faced by SIEM systems. We can have each network component send all its logs to the AI system, the AI will then collect, correlate and analyze the logs of each component based on the pattern recognition rules that we feed to it. The AI can be trained to recognize abnormalities and inconsistencies in the logs based upon real threats that have been detected over the years. And over time, AI can intelligently learn to not only recognize threats but also recommend remedies to fix the problems.

Handling large amounts of data: An AI system has increased capacity, so it can handle logs from various components across the IT infrastructure of an organization. The organization could have network segments across different geographical areas and numerous corporate devices like laptops, computers, mobiles but all this data can be processed by AI because it is designed to be trained on large datasets.

Easily analyzing different data formats: An AI system is proficient in analyzing different data formats including unstructured and malformed data. For example, encrypted data often looks like gibberish and is not human readable. Many a times, network components block legitimate encrypted traffic assuming that it is a server-side injection attack. Using AI, we can recognize legitimate encrypted traffic and allow it to flow into an organization by setting up automation and triggering an email to the web application firewall (WAF) management team.

Aid in setting up correlation rules: An AI system can analyze the logs of each network component individually looking for intrusion patterns and indications of compromise. Even if one device in the network perimeter shows an abnormality, an alarm can be raised. This can give the much-needed leverage to the incident response team to quickly respond to a threat. For example, a honey pot is a virtual machine that is set up to look like it carries sensitive information to lure the attackers and study their methods. Such valuable data from the honeypot will help in training the AI system in combating different cyber-attacks. Thus, an AI system will aid in setting up pattern recognition and data correlation rules because they are adept at analyzing large amounts of data and fetching fine-tuned metrics from it. AI will automatically begin to recognize patterns that it sees on a regular basis and learn from it. This is very time consuming for humans to do. Modifying correlation rules will be easier because by merely forwarding the logs of different devices, the AI will get trained on different scenarios that may arise.

Ease of integrating different components: An AI system makes it easy to integrate with the different components of an IT infrastructure whether physical or virtual. It would merely mean transferring logs via an application programming interface (API) call, or a file transfer solution through scheduled batch jobs. Once the logs are on the AI system, it will take care of the remaining steps to process, understand and analyze the log files.

Removing blind spots in threat detection: Since an AI system can be easily integrated with different network components, it removes blind spots in detecting threats for the security analysts; they do not need to worry about monitoring certain devices separately that were not integrated. For example, AI can scan the emails of the employees of an organization which is not possible by an

SIEM system. A sophisticated phishing email sent to an executive of a company can result in a devastating financial fraud or a phishing email can cause an employee to reveal their credentials which will cause the organization to be compromised by malicious actors. Such emails can be detected and archived by automation and the security team can be notified immediately, even before the email has been read.

Reducing false positive alerts drastically: An SIEM system creates a substantial number of false positive alerts. However, an AI system can be fed with pattern recognition rules, data classification rules, identity and access management (IAM) rules and a bunch of other valuable information like a list of valid locations from which corporate systems can be accessed so that real threats can be distinguished from legitimate traffic. For example, an admin logging into a host from an unauthorized location will be correctly flagged and blocked. Sensitive data that should never leave an organization; if it leaves the network perimeter even by an employee will be detected and reported. However, encrypted sensitive data going to an authorized recipient will not be blocked and will not be reported as a security warning. An AI system can monitor traffic to different interfaces that an organization exposes to the internet and would be familiar with the traffic patterns throughout the day. Only if the traffic spikes suddenly and crosses a certain threshold at an unexpected hour, AI will alert the security team, and traffic can be blocked from certain IP addresses immediately which is indicative of a distributed denial of service (DDoS) attack. A network firewall could be blocking encrypted traffic coming from a trusted source, AI can detect this situation, a notification can be sent, and the traffic can be unblocked by automation as AI has a list of valid IP addresses that the organization can expect traffic from, thus reducing the burden on the security analysts. Hence, drastically decreasing false positives will enhance the efficiency of the incident response team and allow them to focus on real threats and improve security strategies.

Enhanced scanning capabilities: An AI system can scan code repositories for any vulnerabilities. For example, an engineer mistakenly can expose a secret in the application code and forgets to mask it, AI will correctly detect it as it is looking for patterns and the respective team would be alerted. AI will scan the web pages of an organization's customer-facing application for any malicious scripts being executed which will greatly help in combating injection attacks. Many times, organizations realize months later after an information leak has already happened. It is mostly small and medium sized businesses that remain clueless as they do not have a well-developed security posture. AI can help such businesses promptly discover malware across their hosts by persistently scanning the hosts and hence, promptly stop an information leak. An advanced persistent threat (APT) is a sophisticated malware that goes undetected for months as it traverses the network laterally from one host to the other sending information back to the attackers. By continuously scanning the hosts, AI can detect abnormal traffic patterns to entities outside the organization and raise alarms to the security team. A ransomware is an attack where the critical information/files that are needed to operate an organization's systems are encrypted by attackers, thus making them unusable until the ransom is paid. However, AI can keep scanning the critical files/resources of an organization and detect any unwanted attempts at encrypting data thus preventing a ransomware attack.

Keeping up to date with regulations: An AI system can help maintain compliance with laws such as HIPAA and GDPR and other regulations and, report any information leaks immediately to the security team. AI can alert an organization about any vulnerabilities in their infrastructure thus helping them resolve issues much in advance of their security audits.

Natural language processing (NLP) capabilities: AI has the capability to understand human language and identify sensitive information in human speech. For example, if it detects a conversation between employees where credit card numbers and government identification numbers are exchanged. Then by applying automation, the personally identifiable information (PII) can be masked in transit so that if an attacker were to be listening to this conversation, they would not get hold of the sensitive information. Masking is a process of replacing text with symbols to hide the actual content.

Hence, an AI system driven by machine learning that is placed before an SIEM system will help overcome many of the limitations of an SIEM solution. AI would normalize, filter and forward data only regarding real threats to the SIEM system, thus significantly reducing the numerous alerts caused due to data overload. This would result in precise monitoring and alerting and largely removing false positive alerts. It would make it much easier to apply automation. This would free up the capacity of the security team to focus on real threats and work on optimizing security strategies and processes. The security operations center at an organization can perform more effective incident response to prevent, detect, investigate, mitigate and resolve cybersecurity risks. Thus, improving the overall security posture of an organization.

4. Architecture

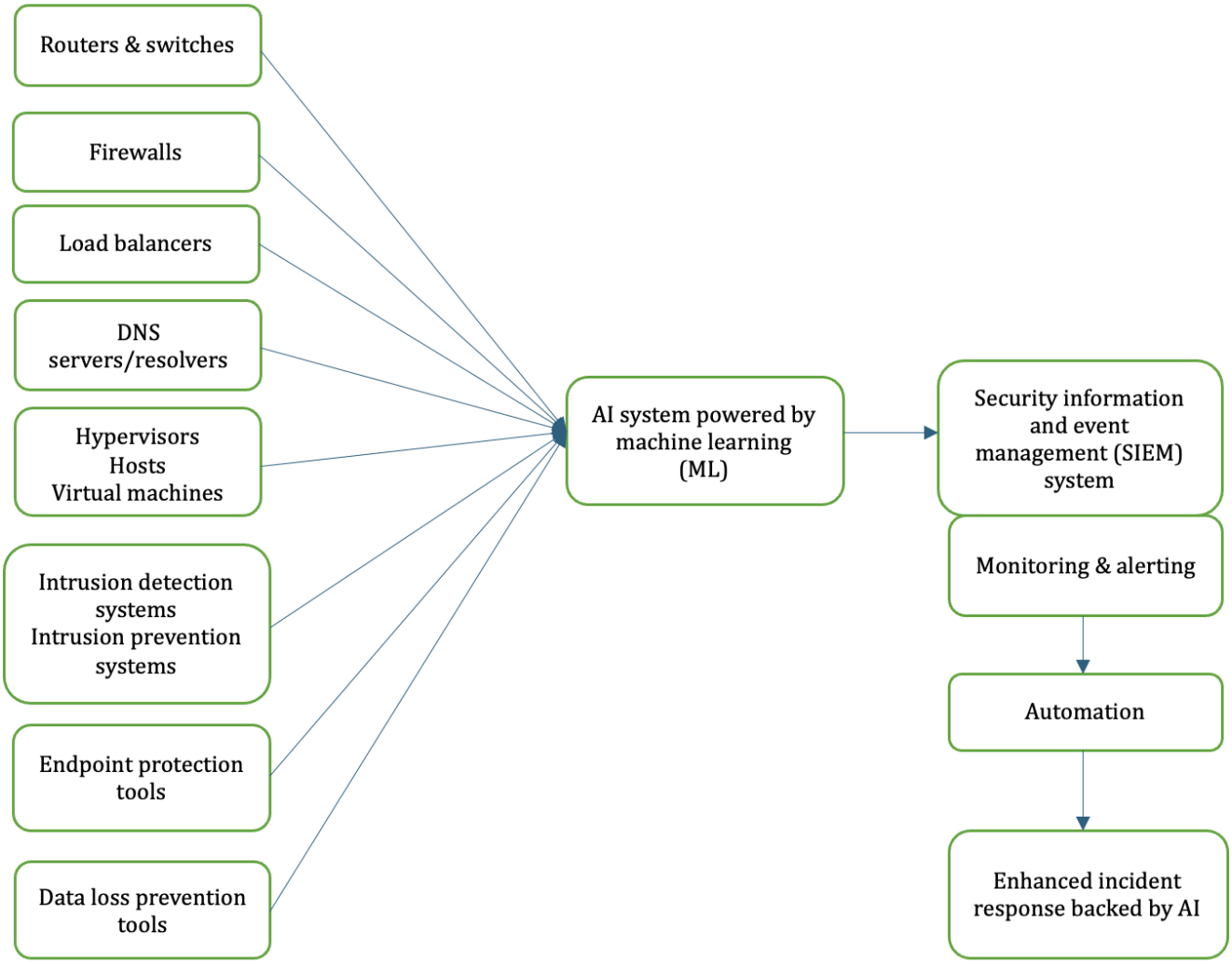

Diagram-1: Cybersecurity architecture of an SIEM system enhanced by an AI system powered by Machine learning

5. Conclusion

Applying artificial intelligence driven by machine learning in a responsible and intelligent way will help solve the issues that we currently face with SIEM systems. Some of the key features that AI adds to an SIEM system is pattern recognition, natural language processing (NLP) capabilities, the capacity to process large volumes of data and deriving meaningful metrics from it; removing false positives that camouflage real threats; ease of integration with AI leading to the removal of blind spots in detecting threats, and ease of maintaining compliance with laws and regulations and, performing security audits effectively. We live in a scary world where AI is being used by attackers to carry out very sophisticated malware, ransomware and phishing attacks on unsuspecting organizations and individuals. Adopting cybersecurity measures that are backed by AI has become the need of the hour to strategize against enhanced cyber-attacks which can lead to more efficient incident response processes and mechanisms within an organization and increase the security and safety of customer and employee data.

REFERENCES

[1] Geoffrey H. Wold, "Cybersecurity Resilience Planning Handbook", Second Edition, Sept. 2020

[2] Arun E Thomas, "Security Operations Center – Analyst Guide: SIEM Technology, Use Cases and Practices", Sept. 2017

[3] Rob Witcher, John Berti, Josh Lake, "Destination CCSP: The comprehensive Guide", First Edition, Oct. 2024 pp. 131 – 155

[4] Miroslav Kubat, "An Introduction to Machine Learning", Third Edition, Sept. 2021

Factor	Traditional SIEM system	Improved SIEM system powered by AI
Large volumes of data	Unable to handle large volumes of data, leading to data overload and countless alerts causing alert fatigue.	Large volumes of data are easily processed by an AI system, as it is built to be trained on large datasets, thus forwarding only the necessary data to an SIEM system and protecting it from data overload.
Integration	Difficult to integrate different input sources like network devices, hosts, VMs, emails, IDS/IPS, DLP, code repositories etc.	Easy to integrate different input sources like network devices, emails, hosts, code repositories, firewalls etc. which have data in unstructured formats, including human speech. AI is adept at analyzing structured and malformed data.
Scalability	Traditionally, an SIEM system is on-premises within an organization and scaling it has many infrastructure and monetary challenges.	An AI system has enormous capacity to process data and is easily scaled. AI system can reside in cloud infrastructure and can be connected to an on-premises SIEM system thus removing scalability constraints.
Correlation rules and false positives	There are complex patterns that exist within data. Applying correlation rules to detect these patterns is very challenging for humans. Thus, due to inadequate and loose correlation rules many false alerts are generated. This causes security analysts to get desensitized to alerts and gives an attacker the leverage to infiltrate a system.	AI has superior pattern recognition capabilities that can find complex patterns. This drastically improves threat detection and the ability to apply correlation rules. Due to self-learning capabilities of AI, over time it recognizes patterns by itself, thus reducing the complexity of correlation rules. So, only data that is related to real threats gets forwarded to an SIEM system and greatly reduces false alerts.
Automation	An SIEM system, lacks the ability to detect if a firewall is blocking legitimate traffic or if a malicious email resides in an executive's mailbox. Thus, it become difficult to apply automation to combat these situations and the security teams needs to take manual steps.	AI will continuously scan hosts for information leaks, mailboxes of the employees for malicious emails and detect if legitimate traffic is being blocked by the firewall. Thus, automation can be setup on these components effectively so that some burden can be removed from the security team, and they can focus more on critical tasks.
Incident response	With difficulty of integrating different data sources, dealing with blind spots in threat detection and numerous false positive alerts, does not lead to effective incident response in a traditional SIEM system.	With an AI system supporting an SIEM system makes it easy to integrate different data sources, remove blind spots in threat detection, significantly reduce false positives and hence leads to effective incident response to combat threats and any abnormalities in the IT infrastructure.