CYBERSECURITY ANALYTICS

*Note: Sub-titles are not captured in Xplore and should not be used

Suranchana Komenmontri
School of Applied Digital Technology
Mae Fah Luang University
6531501129@lamduan.mfu.ac.th

Nabeen Mohammad
School of Applied Digital Technology
Mae Fah Luang University
6531501154@lamduan.mfu.ac.th

Ratchanon Teakmoon
School of Applied Digital Technology
Mae Fah Luang University
6531501165@lamduan.mfu.ac.th

Saharat Yaikamuk
School of Applied Digital Technology
Mae Fah Luang University
6531501173@lamduan.mfu.ac.th

Abstract

At present, cyber threats are continuously increasing, and the ability to identify and mitigate malicious IP addresses plays an important role in creating network security. Traditional security systems, especially rule-based or signature-based detection mechanisms, are often insufficient to cope with Zero-day threats, Advanced Persistent Threats (APT), or large-scale Distributed Denial-of-Service (DDoS) attacks. This results in significant challenges in terms of inspection, accuracy, and the capability of existing defense systems. This project therefore aims to address these problems by developing a cybersecurity analytics model using artificial intelligence (AI), which leverages log data from the FortiSIEM system, a comprehensive Security Information and Event Management (SIEM) platform. The collected data will be pre-processed, structured, and converted into a machine-readable format to enable model training. Using machine learning techniques, including supervised learning and natural language processing (NLP), the model is trained to automatically classify IP addresses as either benign or malicious. This system improves the accuracy of detecting abnormal behaviors. The experimental results show that AI-driven methods are effective, indicating the transformative potential of cybersecurity analytics in organizational environments. It helps enable intelligent decision-making and increases resilience in responding to evolving cyberattacks. Keywords— cybersecurity, analytics, malicious IP detection, SIEM, FortiSIEM, machine learning, natural language processing, AI-driven security

Introduction

In recent years, digital technology has rapidly transformed the way organizations work, bringing new opportunities to education, business, and even government. However, at the same time, it has also opened up more space for cyber risks. As networks grow larger and connect a large number of devices together, they are often attacked by threats such as malware, phishing emails, brute-force login attempts, or IP scans. At Mae Fah Luang University, the IT team uses a Security Information and Event Management (SIEM) system to collect and monitor logs from firewalls, IDS/IPS systems, endpoint protection, and servers. This system helps show what is happening. As a result, problems such as false detection, delayed detection, or even missing serious threats can occur. This project aims to address these issues by adding Artificial Intelligence (AI) into SIEM analysis. Logs from FortiSIEM are transformed into datasets that can be used with machine learning. In this way, the system can detect unusual or harmful activities more quickly. Instead of checking everything manually, algorithms are used to classify IP addresses and help reduce repetitive workloads for staff. The project aims to achieve two main

objectives Build an AI system that can identify malicious IP addresses and automatically disconnect them from the access point. • Use the results to strengthen the university's security policies and improve incident management. By achieving these goals, the project will help reduce the workload of IT staff, increase accuracy in detection, and make the university's network more prepared to handle cyberattacks.

BACKGROUND AND RELATED WORK

Cybersecurity analytics is a field that combines theory, technology, and practical techniques to detect and prevent digital threats. Fundamentally, this analysis relies on three key foundations: Machine Learning Theory Machine learning enables systems to recognize patterns in data and predict outcomes without explicit programming. Various algorithms demonstrate strong capability in detecting abnormal behaviors. What makes these algorithms effective is their ability to generalize from past data, which allows them to identify new and previously unseen types of threats. Natural Language Processing (NLP) Security logs from SIEM systems are often unstructured or semi-structured text, which makes them difficult to analyze directly. NLP techniques such as tokenization, Term Frequency-Inverse Document Frequency (TF-IDF), and word embeddings such as Word2Vec or BERT help transform these raw logs into structured vectors. Doing so enables the AI system to better understand the context and relationships between different events, which greatly improves anomaly detection. Network Security Concepts A deep understanding of how attacks manifest within a network is extremely important in designing features for detection. Events such as brute-force login attempts, IP scanning, or DDoS attacks usually leave distinct patterns in network logs. Using this knowledge allows developers to identify relevant indicators, such as unusually high connection rates or abnormal traffic volumes, and feed these into machine learning models.

A. Related Research A number of studies have demonstrated how effective the integration of AI into cybersecurity can be: • Zhang et al. (2020): Applied supervised learning (SVM and Random Forest) to firewall logs and achieved more than 90% accuracy in detecting malicious IPs. • Chen et al. (2019): Used NLP techniques to analyze SIEM logs, which helped improve real threat detection and reduce false positives. • Siddiqui et al. (2021): Leveraged deep learning to study log behaviors and successfully identified botnets and command-and-control traffic. These studies show that AI-driven log analysis can significantly improve both the efficiency and accuracy of modern cyber defense. B. Related Tools and Technologies Building an effective cybersecurity analytics framework requires the integration of tools and platforms, each serving a specific purpose: • Machine Learning Frameworks: Libraries such as scikit-learn, TensorFlow, and PyTorch are widely used in developing and evaluating detection models. • NLP Libraries: spaCy, NLTK, and Hugging Face Transformers support text preprocessing and embedding generation. • Data Processing Tools: pandas, NumPy, and jsonlines assist in organizing, cleaning, and structuring raw SIEM log data. • SIEM Platforms: FortiSIEM is the primary log collection and incident management system used in this project, providing the necessary datasets for analysis. • Visualization Tools: matplotlib, seaborn, and Plotly help present results in clear visual formats, making it easier for administrators to interpret system outputs.

SYSTEM MODEL

The system's operation can be divided into two main parts: (1) Log Analysis and (2) IP Address Blocking.

Log Analysis

In this step, the system receives log data from a CSV file uploaded by the user via the web application. This file is stored in the database. When the user clicks Start Analysis, the system converts the log data into vectors for the artificial intelligence (AI) model to process and analyze any suspicious or malicious IP addresses. Upon completion of the analysis, the system displays the results of the detected IP addresses via the web application and forwards this information to the IP Blocking module.

IP Blocking

The system receives the IP addresses detected through the analysis process and automatically generates an Access Control List (ACL) on the Cisco Catalyst 9800-CL Wireless Controller access point used in the experiment via a CLI command using the Netmiko Library. This prevents the IP address from connecting to the wireless (Wi-Fi) network via the access point. If the system detects an IP address that has already been blocked, the system will not re-create the ACL. Once the blocking process is complete, the blocked IP addresses are saved to the database for later display via the web application.

Result Comparison

Experiments with the developed analytics system revealed that Ollama's AI model can correctly detect suspicious IP addresses from uploaded FortiSIEM log data (CSV file format), particularly in logs containing keywords such as drop or deny in conjunction with WLAN/WIRELESS from the firewall. When the analytics system detects a suspicious or risky IP, it automatically creates an Access Control List (ACL) on the Cisco Catalyst 9800-CL Wireless Controller to block the connection of that IP via a CLI command using the Netmiko Library. This is to prevent unsafe IP addresses from accessing the university's wireless (Wi-Fi) network and to enhance the security of the internal network.

Implementation

Cybersecurity Analytics systems have a basic working overview of storing log data into a database for artificial intelligence (AI) models to analyze and create Access Control Lists (ACLs) to block IP addresses that are detected as risky or malicious. The implementation can be divided into four main parts: (i) Web Application, (ii) Database, (iii) AI Processing and (iv) Network.

(i) Web Application
The system can be operated via a Web UI for user convenience. Users can upload log files from FortiSIEM in CSV format via the Web UI, track the number of transactions processed, and view a real-time overview of detected at-risk IP addresses. A screen is also provided to display the blocking and unblocking history.

(ii) Database
Database systems are divided into two types: MongoDB for storing raw logs and block history; and ChromaDB for storing log text embeddings converted from CSV files into vector format, enabling AI models to efficiently extract data for analysis.

(iii) AI Processing
Log analysis and detection of risky or malicious IP addresses are performed using the Ollama Large Language Model (LLM) in conjunction with LangChain. It consists of two main components:

Ollama Embeddings: Converts log messages into high-dimensional vectors and stores them in ChromaDB.

Ollama LLM: Extracts, analyzes, and interprets unusual behavior, such as drop or deny events, in wireless networks.

The process works as follows: When a new log is uploaded via the web UI, the system saves the data to MongoDB and creates a vector stored in ChromaDB. A similarity search is then performed to find similar data and sends it to the LLM for processing to identify any suspicious or malicious IP addresses before sending the results to the blocking step.

(iv) Network
For IP addresses detected as potentially at risk, the system automatically blocks them using Netmiko, connecting to the Cisco Wireless Controller via SSH, and running CLI commands to create rules in the Access Control List (ACL), preventing malicious IP addresses from connecting to the wireless network (Wi-Fi) through the access point. Each block and unblock attempt is logged back to MongoDB for historical data storage and can later be displayed in web applications.

The performance is shown in the image below.

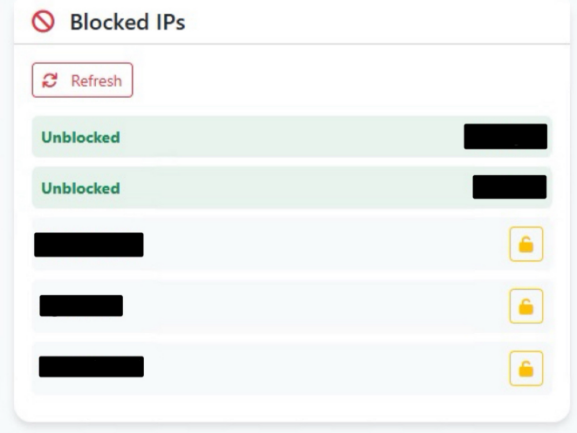

Fig1. Blocking results

Suggestion

This study found that relying solely on log data from a SIEM system may not be sufficient for real-world applications. Linking the data to external threat databases, such as AbuseIPDB or AlienVault OTX, can increase the reliability of detecting IP addresses with suspicious behavior, as in some cases, the actual threat may not be present in internal logs.

Regarding the detection process, using a rule-based approach alone, despite its simplicity, cannot cover all scenarios, especially new attack vectors. Therefore, the researchers recommend combining rule-based approaches with anomaly detection, or supplementing them with basic machine learning. While this approach is simple, it can reduce the false positive rate and increase system resilience. Furthermore, it can be extended to real-time log processing, which receives data from the FortiSIEM firewall as new logs are generated, enabling the system to detect and respond to threats in real time.

Regarding the blocking process, relying on manual, step-by-step operations by administrators may not be able to provide a timely response. Therefore, the system should be designed to automatically send block commands to access points via SSH or API. This approach will significantly shorten incident response times and reduce administrator workload.

Regarding the effectiveness evaluation Clear and concrete criteria should be established, such as detection accuracy rates, false alarm frequency, response time, and the ability to handle increased log volumes without impacting system performance. These metrics better reflect the true value of the system than theoretical explanations alone.

Finally, this research also demonstrates that internal threats are a significant issue, as infected devices can evade firewall inspections. Therefore, blocking them at the access point level is an effective practical measure. In the future, integration with Security Orchestration, Automation, and Response (SOAR) systems may be considered to enhance threat protection capabilities for greater coverage and automation.

Acknowledgement

The researcher would like to express deepest gratitude to Asst. Prof. Dr. Mahamah Sebakor, the thesis advisor, for his valuable guidance, insightful suggestions, and continuous support, which were essential to the success of this research. Special thanks are extended to Mr. Pakornphat Keawrakmook and Ms. Patcharin Manowan, both serving as Network Engineers in the Critical Information Infrastructure Division, Center for Information Technology Services, Mae Fah Luang University (MFU). Their professional advice and technical support regarding SIEM systems were instrumental in the successful completion of this study. The researcher also wishes to thank all individuals who provided encouragement and support throughout the research process, making this study possible.

References

AbuseIPDB. (n.d.). AbuseIPDB: A collaborative project to track abusive IPs. Retrieved August 16, 2025, from https://www.abuseipdb.com/

AlienVault. (n.d.). Open Threat Exchange (OTX). AT&T Cybersecurity. Retrieved August 16, 2025, from https://otx.alienvault.com/

Chiba, Z., & Abou El Kalam, A. (2019). SIEM-based intrusion detection and incident response: A systematic literature review. International Journal of Information Security, 18(6), 623–644. https://doi.org/10.1007/s10207-019-00441-2

Gartner. (2017). Innovation Insight for Security Orchestration, Automation and Response. Stamford, CT: Gartner Research.

Scarfone, K., & Mell, P. (2007). Guide to Intrusion Detection and Prevention Systems (IDPS) (NIST Special Publication 800-94). National Institute of Standards and Technology.

Owen, T. (2020). SIEM solutions perform lookups for IP addresses and integrate with threat intelligence. Lewis University MSIS Project Paper. Retrieved from https://www.cs.lewisu.edu/mathcs/msisprojects/papers/SIEM_TylerOwen.pdf

Riquelme, F., González, J., & Alam, M. (2018). RIPEx: Extracting malicious IP addresses from security forums using cross-forum learning. Proceedings of the International AAAI Conference on Web and Social Media (ICWSM'18). Retrieved from https://arxiv.org/abs/1804.04760

Alsaeedi, A., Khan, S., & Ahmad, R. W. (2019). Mining actionable information from security forums: The case of malicious IP addresses. Proceedings of the ACM Conference on Information and Knowledge Management (CIKM'19). Retrieved from https://arxiv.org/abs/1804.04800

Bartoli, A., Crispo, B., & de Ruiter, J. (2014). Closing the loop of SIEM analysis to secure critical infrastructures. arXiv preprint arXiv:1405.2995. Retrieved from https://arxiv.org/abs/1405.2995

Alhazmi, O. H., Malaiya, Y. K., & Ray, I. (2021). Threat intelligence improving SIEM cybercriminality awareness using information from IP blacklists. ResearchGate. Retrieved from https://www.researchgate.net/publication/353659612