DS-670

CAPSTONE: BIG DATA & DATA SCIENCE

WEEK-2 ASSIGNMENT

SAI HEMANTH KUMAR SANGEPOGU

BHAVANI GODDINDLA

GOPICHAND VEMULA

9th MARCH 2025

**Literature Review:**

Cybernetics, a concept rooted in control systems and communication theory, is crucial in cybersecurity. The principles of feedback loops and adaptive control mechanisms are integral to modern SIEM systems, ensuring continuous improvement in threat detection and response. SIEM solutions leverage these principles to adjust security policies dynamically based on evolving threats.

Anomaly detection, behavioral analytics, and artificial intelligence are also transforming cybersecurity. SIEM systems increasingly incorporate machine learning models to identify deviations from normal network behavior, enabling proactive threat mitigation. Integrating these technologies enhances the effectiveness of SIEM solutions, providing organizations with a comprehensive security posture.

Moreover, cybersecurity frameworks such as MITRE ATT&CK and the Zero Trust Model are being incorporated into SIEM systems to enhance security postures. These frameworks provide structured methodologies for identifying attack techniques and ensuring that only authenticated entities gain network access. Integrating these frameworks into SIEM solutions ensures stronger defenses against adversaries as cyber threats evolve.

MongoDB, a NoSQL database, has emerged as a pivotal technology in modern cybersecurity solutions, particularly in environments that require the real-time processing and analysis of large, unstructured datasets. When integrated with tools like the ELK Stack (Elasticsearch, Logstash, and Kibana), Kafka, and Wireshark, MongoDB plays a key role in storing, processing, and visualizing security logs and network traffic data for enhanced cybersecurity monitoring. This

section explores how MongoDB contributes to SIEM-based cybersecurity monitoring systems, focusing on its scalability, flexibility, and integration capabilities.

## SIEM Systems

System Information and Event Management (SIEM) technology is an integral cybersecurity tool that collects, normalizes, and analyzes logs and security events from various sources within a network. The main function of SIEM is to provide real-time analysis, detect security threats, and ensure compliance with regulatory standards. As López Velásquez et al. (2023) outline in their systematic review, SIEM solutions have evolved from postmortem investigations to real-time event monitoring with advanced analytics capabilities. Importantly, SIEM systems are designed to adapt to the ever-changing cybersecurity landscape, providing a sense of reassurance to organizations.

SIEM technology operates through centralized log collection, enrichment with context data, event correlation, and security alerts. Modern SIEMs integrate with emerging technologies such as blockchain, cloud-based architectures, and containerization to enhance performance and adaptability (López Velásquez et al., 2023). These advancements allow organizations to shift towards a security-as-a-service model, reducing maintenance costs while improving detection accuracy.

## The ELK Stack: A Versatile SIEM Component

The ELK stack, composed of Elasticsearch, Logstash, and Kibana, has become a robust open-source toolset for SIEM functionalities. Each component plays a distinct role in handling large-scale security event data:

- Elasticsearch: A distributed search and analytics engine that stores, indexes, and retrieves vast amounts of log data in real time. It enhances security monitoring by enabling quick search and retrieval of logs, allowing organizations to detect threats promptly (Poat et al., 2023).
- Logstash: A data processing pipeline that collects, processes, and forwards log data to Elasticsearch. It supports various input sources, making it adaptable for SIEM integration. Logstash helps normalize security event data, ensuring log consistency (Poat et al., 2023).
- Kibana is a visualization tool that provides interactive dashboards, allowing cybersecurity analysts to monitor threats effectively. Its graphical representations make detecting anomalies and analyzing security trends easier (Poat et al., 2023).

Poat et al. (2023) highlight how organizations integrate the ELK stack with intrusion prevention systems (IPS) to improve real-time security monitoring. This approach enhances visibility into network anomalies and optimizes response times. The ELK stack enables organizations to centralize logs from multiple sources, making detecting patterns and conducting forensic analysis during security incidents more manageable.

Apache Kafka and Wireshark in Cybersecurity

Apache Kafka and Wireshark play significant roles in cybersecurity, particularly in data streaming and network traffic analysis:

- Apache Kafka: A distributed event streaming platform for real-time data ingestion and processing. In cybersecurity, Kafka supports log aggregation and enhances SIEM's ability to process large-scale event data efficiently. By enabling real-time event streaming, Kafka

helps organizations detect and respond to security threats faster. Kafka also provides scalability, allowing security teams to handle an increasing volume of security events (Shameem et al., 2024).

- Wireshark is a network protocol analyzer widely used for inspecting live network traffic and identifying anomalies such as Distributed Denial-of-Service (DDoS) attacks and malware transmissions. It allows deep packet inspection, enabling cybersecurity professionals to analyze communication between network nodes (Jain & Anubha, 2021).

Shameem et al. (2024) emphasize Wireshark's role in detecting suspicious HTTP traffic and utilizing geolocation data to track real-time cyber attacks. Jain and Anubha (2021) further explore how SNORT and Wireshark enhance network intrusion detection by capturing and analyzing malicious packets. Wireshark provides deep packet inspection capabilities, making it a valuable tool for forensic analysis and cybersecurity investigations.

The reviewed literature demonstrates how SIEM technologies, coupled with open-source tools like the ELK stack, Apache Kafka, and Wireshark, significantly enhance cybersecurity. Hybrid machine learning approaches provide additional layers of security by improving intrusion detection accuracy. The role of cybernetics in SIEM adds a dynamic element, allowing organizations to adapt to emerging threats continuously. As SIEM evolves, integration with real-time data processing, geospatial analytics, and AI-driven threat intelligence will shape the future of cybersecurity solutions. By leveraging traditional security methods and innovative technologies, organizations can create a robust cybersecurity framework capable of effectively mitigating modern cyber threats ("Security information and event management," (2023))

**MongoDB:**

In a Security Information and Event Management (SIEM) system, various data streams, including logs, network traffic, and event data, need to be collected, processed, and analyzed in real-time to detect security incidents and vulnerabilities. MongoDB offers a scalable and flexible database solution that efficiently handles these high-volume, unstructured data types, which are typically generated by security tools like Wireshark and network monitoring systems.

MongoDB's document-oriented model is particularly suitable for handling log files and security events, which often vary in structure and format. The flexibility to store semi-structured data, such as JSON-like documents, allows MongoDB to accommodate the diverse types of security event data that are typically generated by network devices, servers, and security appliances (Sadalage & Fowler, 2012). This is essential in cybersecurity monitoring, where logs from different sources may have different formats, making MongoDB an ideal storage solution.

One of the primary advantages of MongoDB in a cybersecurity monitoring context is its ability to handle vast amounts of unstructured or semi-structured data efficiently. Security logs, network traffic, and security events are often varied and dynamic, making traditional relational databases less suitable for such data. MongoDB’s flexible schema design ensures that the data can evolve as security events change over time, making it an ideal solution for cybersecurity monitoring applications that require continuous updates and scalability (Sadalage & Fowler, 2012).

Moreover, MongoDB's support for horizontal scaling ensures that the system can accommodate increasing data loads over time. As organizations grow and generate more security event data, MongoDB can scale to handle the increased volume of logs, network traffic data, and event information without compromising performance. The ability to store and index large datasets quickly allows security analysts to respond to threats faster and more effectively (Pell, 2017).

**References**

Ahmed, A., Asim, M., Ullah, I., Zainulabidin, & Ateya, A. A. (2024). An optimized ensemble model with advanced feature selection for network intrusion detection. *PeerJ Computer Science*, *10*, e2472. https://doi.org/10.7717/peerj-cs.2472

Ariffin, M. A. M., Darus, M. Y., Haron, H., Kurniawan, A., Muliono, Y., & Pardomuan, C. R. (2022). Deployment of Honeypot and SIEM Tools for Cyber Security Education Model in UITM. *International Journal of Emerging Technologies in Learning*, *17*(20), 149–172. https://doi.org/10.3991/ijet.v17i20.32901

Calderon, G., del Campo, G., Saavedra, E., & Santamaría, A. (2023). Monitoring Framework for the Performance Evaluation of an IoT Platform with Elasticsearch and Apache Kafka. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-023-10409-2

Coppolino, L., Sgaglione, L., D'antonio, S., Magliulo, M., Romano, L., & Pacelli, R. (2022). Risk Assessment Driven Use of Advanced SIEM Technology for Cyber Protection of Critical e-Health Processes. *SN Computer Science*, *3*(1). https://doi.org/10.1007/s42979-021-00858-4

Jain, G., & Anubha. (2021). Application of SNORT and Wireshark in Network Traffic Analysis. *IOP Conference Series: Materials Science and Engineering*, *1119*(1), 012007. https://doi.org/10.1088/1757-899x/1119/1/012007

Khan, J., Elfakharany, R., Saleem, H., Pathan, M., Shahzad, E., Dhou, S., & Aloul, F. (2025). Can Machine Learning Enhance Intrusion Detection to Safeguard Smart City Networks from Multi-Step Cyberattacks? *Smart Cities*, *8*(1). https://doi.org/10.3390/smartcities8010013

Lakkad, A. K., Bhadaniya, R. D., Shah, V. N., & Lavanya, K. (2021). Complex events processing on live news events using apache kafka and clustering techniques. *International Journal of Intelligent Information Technologies*, *17*(1), 39–52. https://doi.org/10.4018/IJIIT.2021010103

Liu, J. C., Yang, C. T., Chan, Y. W., Kristiani, E., & Jiang, W. J. (2021). Cyberattack detection model using deep learning in a network log system with data visualization. *Journal of Supercomputing*, *77*(10), 10984–11003. https://doi.org/10.1007/s11227-021-03715-6

López Velásquez, J. M., Martínez Monterrubio, S. M., Sánchez Crespo, L. E., & Garcia Rosado, D. (2023). Systematic review of SIEM technology: SIEM-SC birth. *International Journal of Information Security*, *22*(3), 691–711. https://doi.org/10.1007/s10207-022-00657-9

Patil, N. V., Krishna, C. R., & Kumar, K. (2022). KS-DDoS: Kafka streams-based classification approach for DDoS attacks. *Journal of Supercomputing*, *78*(6), 8946–8976. https://doi.org/10.1007/s11227-021-04241-1

Poat, M. D., Lauret, J., & Fedele, D. (2023). Flexible visualization of a 3rd party Intrusion Prevention (Security) tool: A use case with the ELK stack. *Journal of Physics: Conference Series*, *2438*(1). https://doi.org/10.1088/1742-6596/2438/1/012040

Santos, V. F., Albuquerque, C., Passos, D., Quincozes, S. E., & Mossé, D. (2023). Assessing Machine Learning Techniques for Intrusion Detection in Cyber-Physical Systems. *Energies*, *16*(16). https://doi.org/10.3390/en16166058

Shameem, S., Venkatesh, K., Shaik, L., T N D, M., Harsha, S., & Lopes, B. R. (2024). Estimating Malware Impact on Network Traffic Analysis by Using Wireshark. In *J. Electrical Systems* (Vol. 20, Issue 7).

Sharma, A., Rani, S., & Driss, M. (2024). Hybrid evolutionary machine learning model for advanced intrusion detection architecture for cyber threat identification. *PLoS ONE*, *19*(9 September). https://doi.org/10.1371/journal.pone.0308206

Blum, D. (2020). Rational Cybersecurity for Business: The Security Leaders' Guide to Business Alignment. In *Rational Cybersecurity for Business: The Security Leaders' Guide to Business Alignment*. Apress Media LLC. https://doi.org/10.1007/978-1-4842-5952-8

Diao, Q. (2014). MongoDB and its application in modern databases. *International Journal of Computer Applications*, 97(10), 22-29. https://doi.org/10.5120/16762-4073

Rodrigues, L., Almeida, P., & Souza, A. (2020). The role of MongoDB in the Internet of Things. *Journal of Computer Science and Technology*, 35(3), 689-698. https://doi.org/10.1007/s11390-020-0319-2