Marwadi University

Faculty of Engineering and Technology

Department of Information and Communication Technology

Subject: Course: Capstone Project

Academic Year: 2025-26

BEHAVIORAL-ANALYTICS AND USER ACCESS VISUALIZATION (IN SPLUNK)

Name: FAITH JACKSON NKUBA (92200133020)

Abstract

The rising frequency and sophistication of cyberattacks necessitate the development of efficient, scalable, and user-centric monitoring systems. This project proposes the implementation of a Splunk-powered anomaly detection and visualization framework to enhance cybersecurity visibility. The proposed system aims to minimize false positives, detect anomalous activities such as unusual login attempts, and provide security analysts with interactive dashboards for improved incident response. This report outlines the problem statement, objectives, ICT relevance, feasibility analysis, market needs, and novelty of the project, while aligning with IEEE standards for research and academic contributions.

I. Introduction

Information and Communication Technology (ICT) plays a pivotal role in the digital economy, where the security of data and systems remains a pressing concern. Organizations across industries face challenges such as credential misuse, insider threats, and targeted attacks, all of which require effective detection mechanisms. According to IBM's *Cost of a Data Breach Report 2024*, the global average cost of a data breach has risen to **USD 4.88 million**, marking an unprecedented increase [1]. This underscores the urgency of improving anomaly detection and incident visibility.

Traditional intrusion detection systems (IDS) rely heavily on predefined rules, which are often insufficient in detecting evolving and sophisticated threats [2]. Splunk, a widely adopted Security Information and Event Management (SIEM) tool, offers a platform for log analysis, machine learning integration, and real-time visualization. By leveraging Splunk's capabilities, this project aims to address existing limitations by providing **anomaly detection dashboards, scoring mechanisms, and contextualized insights**.

II. Problem Statement

Cybersecurity teams often struggle with **three key challenges**:

1. **High false positive rates** in traditional IDS solutions, leading to analyst fatigue.
2. **Delayed detection** of novel attack techniques due to reliance on static rules.
3. **Limited contextual insights**, making it difficult for analysts to prioritize alerts.

Thus, the specific problem addressed is:
"How can a cost-effective, Splunk-based anomaly detection system be designed to improve the accuracy and timeliness of Cybersecurity threat identification while ensuring user data privacy and scalability?"

III. Objectives

The project sets the following **objectives**:

1. **Develop Splunk dashboards** to detect anomalous login attempts, suspicious user sessions, and targeted attacks within **six months**.
2. **Reduce false positives** in anomaly detection by at least **20%**, compared to baseline rule-based IDS systems.
3. **Implement anomaly scoring mechanisms** using Splunk Machine Learning Toolkit (MLTK) to contextualize alerts.
4. **Deploy and test the solution** on synthetic datasets to ensure scalability across cloud and on-premise environments.
5. **Ensure compliance** with data protection regulations such as GDPR by applying anonymization and user-consent mechanisms.

IV. Relevance to the ICT Domain

The project strongly aligns with multiple ICT fields:

- **Cybersecurity and Network Security:** Focused on intrusion detection and anomaly identification.
- **Artificial Intelligence and Machine Learning:** Leveraging Splunk MLTK for anomaly scoring.
- **Big Data Analytics:** Using Splunk's indexing and querying capabilities for processing large-scale log data.
- **Cloud Computing and DevOps:** Supporting hybrid deployment on local servers or cloud (AWS/Azure).

As Gartner notes, **70% of organizations plan to adopt AI-powered SIEM solutions by 2027**, making this project highly relevant to industry needs [3].

V. Feasibility Analysis

A. Technical Feasibility

Component	Tool/Technology	Justification
Log Analysis	Splunk Enterprise / Splunk Cloud	Industry standard SIEM tool
Anomaly Detection	Splunk MLTK, Python	Enables advanced anomaly scoring
Deployment	AWS EC2 / On-prem VM	Scalable and affordable setup

Splunk's ecosystem provides both free-tier and enterprise features, making it technically feasible for academic research while offering real-world relevance.

B. Economic Feasibility

- **Splunk Free License:** Supports up to 500 MB/day log ingestion at no cost.
- **Cloud Costs:** Approx. USD 50–100 per month for compute instances (AWS/Azure).
- **Additional Costs:** Minimal, as datasets will be synthetic or anonymized. This ensures affordability for a student-led project.

C. Ethical Considerations

- **Data Privacy:** Sensitive identifiers (usernames, IPs) anonymized.
- **Consent:** Only synthetic or public datasets used to avoid legal risks.
- **Fairness:** Models tested to minimize biases in anomaly classification.

VI. Market and User Needs Analysis

The growing shortage of skilled cybersecurity professionals makes **automated monitoring systems** increasingly valuable. Splunk-based solutions are widely adopted by enterprises, but SMEs (Small and Medium Enterprises) often lack the resources to implement them. This project fills that gap by offering a **cost-effective, student-friendly SIEM prototype**.

- **Target Users:** SMEs, educational institutions, and security analysts.
- **Market Demand:** By 2027, SIEM adoption will increase by 40%, driven by AI integration [3].
- **Supporting Studies:** Research indicates anomaly-based IDS outperforms static rule-based systems in adapting to new attack vectors [4][5].

VII. Literature Review

Previous research emphasizes the limitations of **signature-based IDS**, which can only detect known threats [6]. Machine learning models applied to IDS have shown improved detection accuracy but often lack transparency and scalability [7].

This project distinguishes itself by integrating:

1. **Visualization with anomaly scoring**, bridging the gap between raw data and analyst decision-making.
2. **Scalability across hybrid infrastructures**, enabling deployment for SMEs.
3. **Ethical data handling**, ensuring user privacy—a critical factor often overlooked in academic IDS studies.

VIII. Conclusion

This project proposal outlines the development of a **Splunk-based anomaly detection and visualization system**, designed to enhance cybersecurity visibility. By addressing false positives, timeliness, and contextualization, the project not only advances academic understanding but also offers practical value to SMEs and training environments. With strong technical feasibility, affordable costs, ethical safeguards, and a clear alignment to ICT trends, the project is both innovative and achievable.

References

[1] IBM Security, *Cost of a Data Breach Report 2024*, IBM Corporation, 2024.
[2] A. Patcha and J. Park, "An overview of anomaly detection techniques: Existing solutions and latest technological trends," *Computer Networks*, vol. 51, no. 12, pp. 3448–3470, 2007.
[3] Gartner, *Market Guide for Security Information and Event Management*, Gartner Inc., 2024.
[4] S. M. Bridges and R. B. Vaughn, "Fuzzy data mining and genetic algorithms applied to intrusion detection," in *Proc. National Information Systems Security Conf.*, 2000, pp. 13–31.
[5] J. Kim, J. Kim, H. Choi, and J. Kim, "Deep learning-based anomaly detection in cybersecurity: A survey," *IEEE Access*, vol. 9, pp. 140–156, 2021.
[6] D. E. Denning, "An intrusion-detection model," *IEEE Transactions on Software Engineering*, vol. SE-13, no. 2, pp. 222–232, 1987.
[7] Y. Xin et al., "Machine learning and deep learning methods for cybersecurity," *IEEE Access*, vol. 6, pp. 35365–35381, 2018.