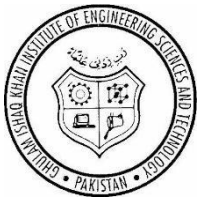

Ghulam Ishaq Khan Institute of Engineering Sciences and Technology (GIKI)

Project Scope Document

For

Project: Machine Learning-based Network Monitoring System

By

Muhammad Umar Maqsood	**2022447**
Shamina Durrani	**2022543**
Muhammad Younas	**2022456**

Supervisor

Dr. Muhammad Zain Siddiqi

Co-Supervisor

Dr. Khurram Jadoon

Madam Beenish, Lecturer

Bachelor of Science in Cyber Security (2022-2026)

Faculty of Computer Science and Engineering (FCSE)

Table of Contents

Contents

Abstract

This study presents a machine learning-based intrusion detection system (IDS) to detect anomalies' behavior in network traffic. Unlike traditional signature-based IDS reliant on predefined attack patterns, the proposed system employs behavioral analysis and advanced pattern recognition to identify irregular attacks in real time. The objective is to develop a scalable system that integrates seamlessly with modern network environments to achieve reliability of the system. The system implements continuous learning through model training and adaptive decision rules to tackle emerging cyber threats, which can contribute to enhancing security of the system. By addressing the limitations of conventional approaches, this ML- based IDS can offer a robust solution for real-time threat detection, contributing to enhanced security across all network environments.

1. Introduction

The rapid growth of cyber threats poses a significant challenge to traditional intrusion detection systems (IDS), which often rely on predefined attack signatures. As cyber-attacks become more sophisticated, traditional intrusion detection systems struggle to detect novel and polymorphic threats, rendering them ineffective against zero-day attacks. Machine learning (ML) techniques, however, have the potential to address these limitations by learning from data and detecting previously unknown attack patterns.

The system will be designed to detect several types of attacks, such as Distributed Denial of Service (DDoS) attacks, port scans, web application attacks, and others. By applying multiple preprocessing techniques, we aim to enhance the quality of the data and ensure the robustness of the system. The primary goal of the system will be to provide real-time detection and prevention, ensuring that the network can respond promptly to attacks while minimizing resource usage.

2. Problem Statement

Traditional IDS are primarily signature-based, relying on predefined patterns of attacks. This is a critical limitation, especially in the face of rapidly developing cyber-attacks. Furthermore,

traditional IDS often struggle with handling large amounts of network traffic, which makes real-time detection challenging.

As cyber threats become more advanced, it is essential for IDS to be scalable, adaptable, and capable of detecting a broad range of attack types. The challenge is compounded by imbalanced and high-dimensional datasets, which often result in poor performance for detecting minority attacks. Additionally, the ability to process large volumes of network traffic in real time while maintaining high detection accuracy is an ongoing challenge for the development of effective intrusion detection systems.

This project proposes a solution by developing an ML-based intrusion detection and prevention system that addresses these limitations, focusing on the detection of both known and unknown attacks. The system aims to provide real-time threat detection, scalability, and adaptability, overcoming the challenges of traditional signature-based IDS.

3. Objectives

The primary objective of the project is to design and develop a machine learning based intrusion detection system for real-time network threat detection using ML techniques, with a focus on metadata analysis and feature selections. The system aims to enhance the security of high-traffic networks by achieving high detection accuracy, minimizing false positives, and ensuring computational efficiency.

The key objectives and contributions of the project are as follows:

- **Enhance Feature Selection:** Identify and select optimal features to reduce dimensionality, improve model accuracy, and minimize computational overhead in resource-constrained environments.
- **Compare Machine Learning Algorithms:** Evaluate the performance of multiple ML algorithms based on key performance indicators (KPIs) such as detection accuracy, false positive rate, and processing latency etc.
- **Minimize False Positives**: Optimize model sensitivity and specificity to reduce false alarms while maintaining high detection rates for genuine threats, addressing data imbalance through synthetic data generation techniques.

- **Deliver an Open-Source Software Prototype:** Develop a user-friendly system that serves as a practical tool for network administrators and a benchmark for future research in network security.

4. Related System Analysis/Literature Review

Application Name	Features	Weakness	Relevance with the Proposed System – How these weaknesses are addressed
• **Snort (Signature-Based)** • **Cloud-Based IDS** • **Basic ML-IDS**	• Real-time analysis, Rule-based, Signature DB • Scalable monitoring, Centralized logging • Traditional classifiers	• No zero-day detection, Signature-dependent • Slow adaptation, Latency • Misses rare attacks	• Unsupervised ML for unknown threats • Federated learning + dynamic updates • Synthetic data + class balancing

Table 1: Related System Analysis with proposed project solution

5. Vision Statement

Instead of relying on outdated methods that only recognize known threats, we envision a future where technology actively learns from network behavior to spot suspicious activity even if it's never been seen before. This means building tools that work in real time, analyze encrypted traffic without needing to decrypt it, and help organizations stop attacks before they cause harm.

We aim to make security systems smarter, simpler, and more accessible. By focusing on patterns rather than predefined rules, our approach will protect everyone from small businesses to large

corporations without requiring constant manual updates. It will reduce false alarms, handle massive amounts of data efficiently, and stay effective as networks grow more complex.

We want to empower people and organizations to trust their digital environments fully. This means creating solutions that are not just powerful but also easy to understand and use, bridging the gap between cutting-edge technology and everyday security needs. By working together, machines learning from data and humans guiding the process, we can build a safer internet where innovation thrives, and threats are neutralized before they escalate.

6. Scope

The project aims are to deliver a robust, machine learning-powered system for real-time network threat detection. The system will operate efficiently in high-traffic network environments and adapt to emerging threats. The scope of the project includes the following core functionalities:

- Perform feature selection based on metadata (e.g., flow duration, packet patterns, inter-arrival times) to reduce dimensionality and enhance model performance.
- Implement multiple ML algorithms for threat detection, with a focus on real-time processing in high-traffic networks.
- Enable adaptive learning to continuously update models with new threat patterns, ensuring resilience against evolving cyberattacks.

7. Project Stakeholders and Roles

Project Sponsor	
Stakeholder	• [PROJECT MEMBER] Muhammad Umar Maqsood Muhammad Younas Shamina Durrani • [PROJECT SUPERVISOR] Dr. Zain Siddiqui

	• [PROJECT CO-SUPERVISOR] Miss Beenish Urooj Dr. Khurram Jadoon • Final Year Project Evaluation Panel

Table 2: Project Stakeholders for Machine Learning-based Network Monitoring System

8. References

- J. Dromard et al., "Online Detection of Network Attacks with Autoencoders," 2018.
- F. A. A. Azar et al., "Machine Learning for Cybersecurity: Techniques and Applications," Springer, 2021.
- Sommer, R., & Paxson, V. (2010). Outside the Closed World: On Using Machine Learning for Network Intrusion Detection. IEEE Symposium on Security and Privacy.
- Discusses challenges and opportunities of ML in intrusion detection, emphasizing generalization beyond labeled datasets.
- Ahmad, Z., Shahid Khan, A., Wai Shiang, C., Abdullah, J., & Ahmad, F. (2021). Network intrusion detection system: A systematic study of machine learning and deep learning approaches. Transactions on Emerging Telecommunications Technologies.
- Reviews ML/DL techniques for intrusion detection, including real-time and encrypted traffic analysis.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. JAIR.