IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 5, Issue 10, April 2025

Impact Factor: 7.67

Machine Learning Based Intrusion Detection System

Dr. Deepika Ajalkar[1], Vrushali Chavan[2], Pratiksha Bhosle[3]

Professor, Department of Computer Science and Engineering (Cyber Security)[1]
Student, Department of Computer Science and Engineering (Cyber Security)[2,3]
G H Raisoni College of Engineering and Management ,Wagholi, Pune, Maharashtra, India

Abstract: *With the exponential growth in the size and sophistication of cyber attacks, the security of digital infrastructures has emerged as a critical issue for organizations and individuals. Conventional Intrusion Detection Systems (IDS) mainly rely on signature-based methods, which are plagued by the inability to detect unknown or zero-day attacks. To address these limitations, this paper introduces a Machine Learning-Based Intrusion Detection System (MLA IDS) that can intelligently scan network traffic and classify it into normal or malicious categories. The system uses supervised learning algorithms — Random Forest, Decision Tree, and Support Vector Machine (SVM)—trained on benchmark datasets like NSL-KDD and CICIDS2017. The project involves several phases: data collection, preprocessing, model training, evaluation, and real-time detection. The experimental results show high accuracy, lower false positives, and good adaptability to new intrusion types. It is scalable to fit deployment across enterprise, cloud, or IoT networks and is a cutting-edge approach to defending against cybersecurity.*

Keywords: Intrusion Detection System, Machine Learning, Cybersecurity, Random Forest, SVM, NSL-KDD, CICIDS2017

I. INTRODUCTION

In today's age of modern technology, security has become a critical issue for individuals, companies, and governments around the world. With the ongoing growth of the internet and the use of digital technologies in nearly every domain, networks have grown increasingly complex and vulnerable. This development has been matched by a sudden increase in the severity, magnitude, and complexity of cyberattacks. From high-profile data breaches and ransomware attacks to targeted espionage and denial-of-service (DoS) attempts, the threat environment is quickly evolving, rendering conventional security measures inadequate. An Intrusion Detection System (IDS) has a vital position in a cybersecurity design by being able to monitor network or system activities and whether any unauthorized access or abnormal activity. Traditional IDS techniques depend primarily on signature-based or rule-based approaches, with signatures of attacks known to the system pre-stored in a database and compared to incoming traffic. While this method is good at identifying known threats, it fails to handle them well when confronted with zero-day attacks, unknown intrusion patterns, or evasive attacks with time-varying mutation. Furthermore, static IDS models usually generate elevated false positive levels, overwhelming the security analysts and diminishing the level of trust in the system. To overcome such constraints, researchers and practitioners have increasingly looked towards artificial intelligence (AI) and machine learning (ML) methods for developing smart IDS. Machine Learning-based Intrusion Detection Systems (MLA IDS) can also learn, recognize patterns that are not obvious, and update itself because there are dynamic threat profiles. These systems are not based on hand-crafted rules but rather employ training datasets to develop predictive models that can generalize effectively to new, unseen inputs. The purpose of this study is to create a lightweight, modular, and efficient machine learning-driven IDS that can identify malicious traffic patterns by applying common supervised learning algorithms like Random Forest, Support Vector Machine (SVM), and Decision Tree. The system proposed here is centered on public datasets such as NSL-KDD for training and validation, and adheres to a systematic pipeline, beginning from data preprocessing to model evaluation, and ultimately deployment and real-time detection

DOI: 10.48175/IJARSCT-25679

II. MODULE IDENTIFICATION

Module 1: Registration & Login

Provides access control and allows only authorized administrators to configure and operate the system. This ensures the integrity of IDS functionalities.

Module 2: Data Collection

This module retrieves labeled datasets from public repositories or live traffic sources. Common datasets used include NSL-KDD and CICIDS2017, which contain examples of both normal and attack traffic.

Module 3 : Data Preprocessing

- Raw data is cleaned, normalized, and transformed. This includes:
- Missing value treatment .
- Feature selection and dimensionality reduction
- Label encoding of categorical data
- Data balancing using techniques like SMOTE

Module 4 : Model Training and Evaluation

This module involves splitting the dataset into training and testing sets and training supervised learning models such as:

- Random Forest -Decision Tree
- Support Vector Machine (SVM)

The models are evaluated using key metrics: Accuracy, Precision, Recall, F1-Score, and Confusion Matrix.

Module 5 : Intrusion Detection & Classification

The best-performing model is deployed to analyze real-time or batch network traffic and classify each instance as normal or malicious.

Module 6 : Alert Generation and Logging

Whenever a malicious activity is detected, the system generates an alert. Logs are maintained with timestamps, IP addresses, type of attack, and system response.

III. ARCHITECTURE DIAGRAM

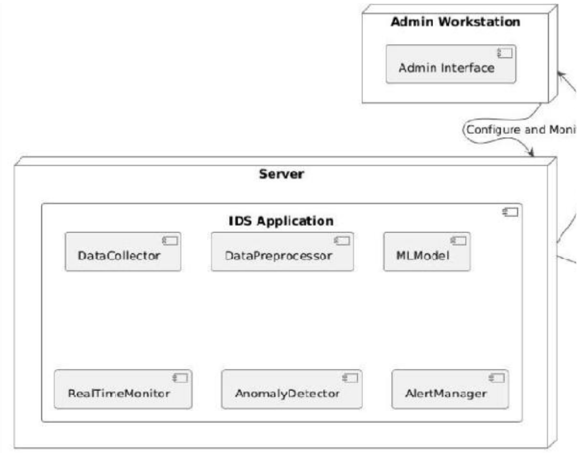

Fig. 1 System Architecture of Machine Learning Based Intrusion Detection System

IV. SCOPE

- Multi-attack detection: Can identify different attacks such as DoS, Probe, R2L, and U2R.
- Real-time adaptability: Can be used in real-time network environments with real-time monitoring.
- Scalability: Most suitable for big organizations, cloud servers, and IoT networks.
- Modular Enhancement: Scalable with deep learning (CNNs, RNNs), online learning, or federated learning.

V. RELATED WORK

Kumar et al. (2020) proved that feature selection enhances detection precision and reduces computation.

Vinayakumar et (2019) also proposed a hybrid IDS based on supervised and unsupervised approaches to improve the detection.

Shone et al. (2018) employed autoencoders for feature extraction, which enhanced the effectiveness of deep learning models. Zhang et al. (2022) suggested federated learning towards distributed IoT security with privacy and performance.

VI. PROBLEM DEFINITION

Conventional IDS systems suffer from the following problems:

- Failure to identify new attacks
- High rates of false positives/false negatives
- Limited scalability and flexibility
- Manual rule-based tuning requires

The MLA IDS system described here resolves these issues via automated learning, high detection, and capacity to handle varying attack types by strong ML models.

VII. CONCLUSION

Machine Learning-Based Intrusion Detection System that utilizes the strength of smart algorithms for security network infrastructures. Shaking loose the rigid signature-based approach, the MLA IDS delivers dynamic, scalable, and highly accurate identification of harmful activity. Dynamic real-time tuning, eliminating false positives, and modular design make this system a top contender to be implemented in any setting, such as enterprise networks, cloud computing, and IoT systems. Future work can involve the application of deep learning models, self-learning capabilities, federated deployment, and integration into automated response systems. This paper not only contributes to the literature understanding of ML in cybersecurity but also gives a real-world foundation for building next-generation intrusion detection platforms.

REFERENCES

[1]. Kumar, S., Raj, P., & Rathore, H. (2020). "Anomaly-based intrusion detection using feature selection and machine learning algorithms." Journal of Cybersecurity Research, 5(2),85-99.

[2]. Shone, N., Ngoc, T. N., Phai, V. D., & Shi, Q. (2018). "A deep learning approach to network intrusion detection." IEEE Transactions on Emerging Topics in Computational Intelligence, 2(1),41-50.

[3]. Vinayakumar, R., Alazab, M., Srinivasan, S., Pham, Q. V., & Simran, K. (2019). "A hybrid deep learning approach for network intrusion detection." Future Generation Computer Systems, 100,334-352.

[4]. Li, W., He, X., Zhang, X., & Chen, Y. (2021)."Real-time intrusion detection system using deep learning for network security." Computers & Security, 110,102433.

[5]. Zhang, J., Wang, H., & Liu, Y. (2022). "Federated learning-based IDS for distributed networks: A comprehensive study." IEEE Internet of Things Journal, 9(4), 2897-2909

DOI: 10.48175/IJARSCT-25679