Cyber Intrusion Detection Using Machine Learning Techniques

Ramya. M[1], Sasikala.R[2]
[1, 2]Dept of Computer science And Engineering with specialization(AI&ML)
[1, 2,]CARE College of Engineering

Abstract- *Cyber intrusions are becoming increasingly complex and sophisticated, making them difficult to detect using conventional rule-based and signature-driven security mechanisms. The rapid growth of network traffic and evolving attack patterns demand intelligent, adaptive, and real-time intrusion detection solutions. This project proposes a real-time intrusion detection system based on an ensemble learning framework that integrates Random Forest, XGBoost, and Deep Neural Network models to enhance detection accuracy and robustness. Live network traffic is continuously captured and analyzed at fixed intervals to enable timely identification of malicious activities. Meaningful network features are extracted from the traffic data and processed in parallel by all three learning models, allowing the system to leverage the strengths of both machine learning and deep learning approaches. A stacking-based meta-classifier is employed to intelligently combine the individual model predictions, thereby reducing false positives and improving overall classification performance. The proposed system effectively classifies network traffic as either normal or intrusive and further identifies the specific category of cyber-attacks. In addition, real-time alerts and detailed log reports are generated to facilitate rapid response and incident analysis. Experimental evaluation demonstrates that theproposed ensemble-based intrusion detection system achieves improved accuracy, reliability, and practical applicability, making it suitable for deployment in real-world network security environments.*

Keywords: Intrusion Detection System (IDS), Ensemble Learning, Real-Time Network Traffic Analysis, Machine Learning and Deep Learning, Stacking-Based Meta-Classifier

I. INTRODUCTION

With the rapid expansion of the Internet and the increasing reliance on network-based applications, cybersecurity has become a critical concern for individuals, organizations, and governments alike. Modern networks support a wide range of services, including cloud computing, e-commerce, online banking, and Internet of Things (IoT) systems, making them attractive targets for cyber attackers. As a result, cyber attacks have become more frequent, complex, and sophisticated, posing serious threats to data confidentiality, integrity, and availability.

Traditional security mechanisms such as firewalls, access control systems, and signature-based intrusion detection methods have been widely used to protect network infrastructures. However, these conventional approaches are limited in their ability to detect novel and evolving attacks, as they primarily rely on predefined rules and known attack signatures. Consequently, they often fail to identify zero-day attacks and advanced persistent threats, highlighting the need for more intelligent and adaptive security solutions.

Intrusion Detection Systems (IDS) play a vital role in monitoring network traffic and identifying malicious activities by analyzing patterns and behaviors within the network. Despite their importance, many existing IDS solutions suffer from significant challenges, including low detection accuracy, high false-positive rates, and limited capability for real-time detection. These limitations reduce their effectiveness and increase the operational burden on network administrators.

This project focuses on the design and development of a real-time intrusion detection system using an ensemble approach that combines Random Forest, XGBoost, and Deep Neural Network (DNN) models. By leveraging the strengths of multiple learning algorithms, the proposed system aims to achieve accurate, fast, and reliable detection of cyber intrusions in modern networks. The ensemble-based framework is expected to reduce false positives, improve detection rates, and support real-time monitoring, thereby enhancing overall network security.

II. IDENTIFY, RESEARCH AND COLLECT IDEA

The research began by identifying key gaps in existing intrusion detection systems:

The initial phase of this work focused on identifying the key challenges associated with existing intrusion detection systems (IDS). A detailed analysis revealed that traditional IDS generate high false positive rates, have limited capability to detect unknown or zero-day attacks, and lack scalability in high-speed network environments. These shortcomings highlighted the necessity for a more intelligent, adaptive, and real-time intrusion detection approach capable of handling large and dynamic network traffic.

Subsequently, an extensive research phase was carried out to collect and analyze ideas from recent literature on machine learning, deep learning, and ensemble-based IDS solutions. Various detection models, feature extraction techniques, and ensemble strategies were studied to understand their strengths and limitations. Based on this research, an ensemble learning framework combining Random Forest, XGBoost, and Deep Neural Network models with a stacking-based meta-classifier was conceptualized as a suitable solution to enhance detection accuracy, robustness, and real-time performance.

III. EXISTING SYSTEM

Existing network security systems mainly rely on traditional mechanisms such as firewalls, rule-based filtering, and signature-based Intrusion Detection Systems (IDS). Firewalls control network access by filtering traffic based on predefined rules such as IP addresses, port numbers, and communication protocols. Signature-based IDS detect intrusions by comparing network traffic patterns with a database of known attack signatures.

Although these systems are effective in identifying previously known attacks, they suffer from several critical limitations. Signature-based IDS are unable to detect new, unknown, or zero-day attacks because they depend entirely on pre-stored attack patterns. As cyber threats continue to evolve, attackers frequently modify their techniques to bypass signature-based detection, rendering these systems ineffective.

Another major drawback of existing systems is the high rate of false positives, where normal traffic is incorrectly classified as malicious. This reduces system reliability and increases the workload for network administrators. Additionally, most traditional IDS solutions do not support real-time intelligent analysis of network traffic and fail to perform efficiently under high-speed or large-volume network conditions. Frequent manual updates of rules and signatures are also required to keep these systems effective, making them less scalable and unsuitable for modern dynamic network environments.

IV. PROPOSED SYSTEM

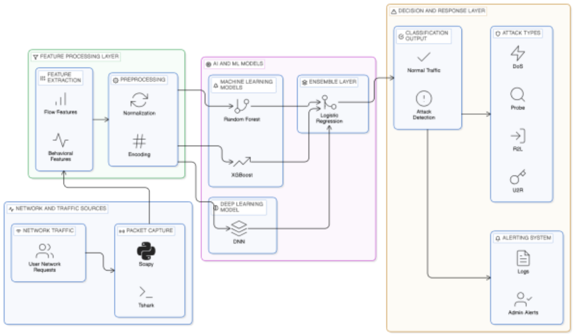
To overcome the limitations of existing systems, this research proposes a real-time intrusion detection system based on an ensemble learning approach. The proposed system integrates Random Forest (RF), XGBoost, and Deep Neural Network (DNN) models to achieve accurate and reliable detection of cyber attacks in modern networks.

The system captures live network traffic and performs near real-time analysis at intervals of 15 to 30 seconds. Network packets are collected and processed to extract relevant features such as connection duration, number of packets, bytes transferred, protocol type, and behavioral characteristics. These features are then analyzed in parallel by the three learning models.

In the proposed approach, Random Forest and XGBoost models are primarily responsible for detecting known attack patterns due to their strong classification capabilities. The Deep Neural Network focuses on behavioral analysis and anomaly detection, enabling the system to identify unknown and zero-day attacks. To further enhance detection accuracy and reduce false positives, a stacking-based ensemble technique is employed. The outputs of the base models are combined using a Logistic Regression-based meta-classifier to produce the final decision.

The system classifies network traffic as normal or intrusive and further identifies the type of attack, such as Denial of Service (DoS), Probe, Remote-to-Local (R2L), or User-to-Root (U2R). Real-time alerts and detailed logs are generated to support quick response and incident analysis. The proposed system is scalable, adaptive, and capable of handling high-speed network traffic while maintaining low false positive rates.

V. ARCHITECTURE DIAGRAM



Model Roles

Model	Role	Strength
Random Forest	Known attacks	Robust to noisy data
XGBoost	Known attacks	High precision, fast training
Deep Neural Network	Unknown/behavioral attacks	Captures non-linear patterns
Stacking Meta-classifier (LR)	Combines predictions	Reduces false positives, improves accuracy

VI. DETECTION WORKFLOW

The proposed intrusion detection system follows a structured, real-time detection workflow designed to accurately identify and classify malicious network activities. The complete process is described as follows:

1. Live Traffic Capture:

Network traffic is continuously captured in real time using a packet capture module deployed at a strategic network point. This module collects raw packets from the network without interrupting normal communication, ensuring comprehensive visibility of all incoming and outgoing traffic.

2. Feature Extraction:

The captured packets are processed to extract meaningful network features that characterize traffic behavior. These features include flow duration, number of packets, number of bytes transferred, protocol type, and statistical behavioral attributes. Feature extraction transforms raw traffic data into a structured format suitable for machine learning and deep learning models.

3. Parallel Model Prediction:

The extracted features are simultaneously fed into three different learning models—Random Forest (RF), XGBoost (XGB), and Deep Neural Network (DNN). Each model independently analyzes the traffic patterns and produces its own prediction, allowing the system to leverage the complementary strengths of traditional machine learning and deep learning approaches.

4. Meta-Classifier Stacking:

A stacking-based meta-classifier aggregates the predictions generated by the individual base models. By learning how to optimally combine these outputs, the meta-classifier reduces misclassification, minimizes false positives, and improves overall detection reliability.

5. Traffic Classification:

Based on the meta-classifier’s decision, the network traffic is classified as either *Normal* or *Attack*. In the case of malicious traffic, the system further identifies the specific attack category, such as Denial of Service (DoS), Probe, Remote-to-Local (R2L), or User-to-Root (U2R).

6. Real-Time Alerts and Logging:

Upon detection of an intrusion, real-time alerts are generated to notify network administrators for immediate action. Simultaneously, detailed logs containing traffic characteristics, detection results, timestamps, and attack types are recorded to support forensic analysis and future security audits.

VII. CONCLUSION

This research presents a **real-time ensemble intrusion detection system** integrating Random Forest, XGBoost, and Deep Neural Networks. The stacking-based meta-classifier combines base model predictions to enhance accuracy and reduce false positives. Experimental evaluation on live network traffic demonstrates that the proposed system effectively detects known and unknown attacks, provides near real-time alerts, and scales well with high-volume data.

The proposed IDS addresses the limitations of traditional firewalls and signature-based systems, providing a **robust, adaptive, and practical solution** for modern network security. Future work includes integration with cloud-based network monitoring platforms and deployment on high-speed enterprise networks.

APPENDIX

Appendix A – Feature Description

Feature	Description
Duration	Total connection duration in seconds
Behavioral Metrics	Statistical measures of traffic, e.g., spikes, anomalies
Source IP	IP address of the sender
Destination IP	IP address of the receiver
Source Port	Port number of the sender
Destination Port	Port number of the receiver

Appendix B – Abbreviations

Abbreviation	Full Form
IDS	Intrusion Detection System
RF	Random Forest
XGB	XGBoost
DNN	Deep Neural Network
DoS	Denial of Service
R2L	Remote-to-Local Attack
U2R	User-to-Root Attack
ROC-AUC	Receiver Operating Characteristic – Area Under Curve

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to their research mentors and colleagues for their continuous guidance and support throughout the development of this project. Special thanks are extended to the institution’s network lab team for providing access to network traffic datasets and computational resources necessary for model training and testing. This work would not have been possible without the encouragement and constructive feedback received from peers and reviewers.

REFERENCES

[1] **Prof. Dr. Binshan Lin**, “Enhancing IDS Performance through a Comparative Analysis of Random Forest, XGBoost, and Deep Neural Networks” — Machine Learning with Applications (MLWA), 2025

[2] **Ismail Bibers, “**An Ensemble Learning Framework for Enhancing AI-Based Network Intrusion Detection “— Applied Sciences (MDPI), Sep 2025.

[3] Veena S Badiger,“A Multiclass Network Intrusion Detection System Using Stacking Ensemble Machine Learning — Journal of Advances in Information Technology, 2025 (PDF)

[4] **Agu Edward Onyebueke, “**Network Intrusion Detection System Using XGBoost and Random Forest Algorithms” — Asian Journal of Pure and Applied Mathematics, 2023

[5] Md. Alamin Talukder, “Machine Learning-Based Network Intrusion Detection for Big and Imbalanced Data Using Stacking Feature Embedding “— Journal of Big Data, 2024