Network Intrusion Detection and Forensic Logging System Using Machine Learning

Mr. R. Alex Giftson,
Associate Professor, CSE
N.S.N CET, Karur

Raja Sekar M,
Student, CSE
N.S.N CET, Karur

Yuvaraj V,
Student, CSE
N.S.N CET, Karur

Yogesh I
Student, CSE
N.S.N CET, Karur

Vasanth R
Student, CSE
N.S.N CET, Karur

I. ABSTRACT

In recent years, the rapid expansion of digital technologies, cloud computing, and internet-based services has significantly increased the vulnerability of network systems to a wide range of cyber threats. Modern organizations rely heavily on interconnected systems for data storage, communication, and business operations, making them attractive targets for cyber attackers. As a result, the frequency and sophistication of cyberattacks such as Distributed Denial of Service (DDoS), phishing, ransomware, and advanced persistent threats (APTs) have increased dramatically.

Traditional Intrusion Detection Systems (IDS), which primarily rely on predefined signatures or static rule-based mechanisms, are no longer sufficient to handle these evolving threats. While signature-based systems are effective in detecting known attack patterns, they fail to identify unknown or zero-day attacks. Similarly, anomaly-based systems often suffer from high false positive rates, leading to unnecessary alerts and reduced system reliability. These limitations highlight the need for more intelligent and adaptive security solutions.

To address these challenges, this research proposes a Machine Learning-based Network Intrusion Detection and Forensic Logging System (NIDS) that enhances the detection and classification of cyber threats in real time. The proposed system leverages multiple machine learning algorithms, including Random Forest, Decision Tree, K-Nearest Neighbors (KNN), Naive Bayes, and Logistic Regression, to analyze network traffic and accurately classify it as normal or malicious. By utilizing supervised learning techniques, the system is capable of learning complex patterns and detecting previously unseen attacks.

The system is trained and evaluated using the NSL-KDD dataset, which is widely recognized as a benchmark dataset in intrusion detection research. The dataset includes various types of network traffic along with labeled attack categories such as DoS, Probe, Remote-to-Local (R2L), and User-to-Root (U2R). To ensure effective model performance, data preprocessing techniques such as feature encoding, normalization, and noise removal are applied. The performance of the models is evaluated using standard metrics including accuracy, precision, recall.

II. INTRODUCTION.

The increasing reliance on digital communication and online services has made cybersecurity a critical concern for individuals, organizations, and governments worldwide. Network systems serve as the backbone of modern infrastructure, enabling data exchange, communication, and business operations. However, this dependency has also made them prime targets for cyberattacks.

Cyber threats such as Distributed Denial of Service (DDoS), malware infections, phishing attacks, and unauthorized access attempts have become more frequent and sophisticated. These attacks can lead to severe consequences, including data breaches, financial losses, and damage to organizational reputation.

Intrusion Detection Systems (IDS) are essential tools used to monitor network traffic and detect suspicious activities. Traditional IDS techniques can be categorized into:

- **Signature-Based Detection:** Identifies known threats using predefined patterns
- **Anomaly-Based Detection:** Detects deviations from normal network behavior

While these methods have been widely used, they suffer from several limitations:

- Inability to detect unknown or zero-day attacks
- High false positive and false negative rates
- Limited adaptability to new attack patterns

Machine learning has emerged as a promising solution to overcome these limitations. By analyzing large volumes of network data, machine learning models can learn patterns and identify anomalies more effectively.

This research aims to develop an intelligent intrusion detection system that combines machine learning techniques with forensic logging capabilities to provide a comprehensive cybersecurity solution.

Moreover, the system is designed with a real-time monitoring capability through a web-based dashboard, allowing administrators to visualize network activity, detect anomalies instantly, and respond to threats proactively. This integration of detection, classification, logging, and visualization makes the system a comprehensive solution for modern cybersecurity challenges.

Overall, this research aims to bridge the gap between traditional intrusion detection methods and modern intelligent security systems by leveraging machine learning and secure logging techniques.

Despite their widespread use, these traditional approaches suffer from several significant limitations. One major drawback is their inability to detect zero-day attacks, which exploit previously unknown vulnerabilities. Additionally, high false positive and false negative rates

III. LITERATURE REVIEW.

Numerous studies have explored the use of machine learning techniques in intrusion detection systems. These approaches have shown significant improvements over traditional methods.

Decision Tree and Random Forest algorithms are widely used due to their high accuracy and ability to handle complex datasets. Random Forest, in particular, reduces overfitting by combining multiple decision trees, making it more robust.

K-Nearest Neighbors (KNN) is a simple yet effective algorithm that classifies data based on similarity measures. However, it requires high computational resources, especially for large datasets.

Naive Bayes is a probabilistic classifier that performs well in scenarios with independent features. It is computationally efficient but may not capture complex relationships between variables.

Logistic Regression is commonly used for binary classification tasks and provides interpretable results. However, it may not perform well with highly nonlinear data.

Recent advancements have introduced deep learning techniques such as Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN), which offer improved performance but require significant computational power.

Despite these advancements, most existing systems focus only on detection and lack proper forensic logging mechanisms. Forensic logging is crucial for analyzing past attacks, identifying attackers, and providing legal evidence.

IV. PROBLEM STATEMENT

- Despite the widespread deployment of Intrusion Detection Systems (IDS) in modern network infrastructures, several critical limitations reduce their effectiveness in handling today's complex cybersecurity threats.
- One of the primary challenges is the **lack of real-time intelligent detection capabilities**. Traditional IDS solutions rely heavily on predefined signatures or static rule sets. While effective against known attacks, these systems fail to detect new and evolving threats, commonly referred to as zero-day attacks. As cyber attackers continuously modify their techniques, static detection mechanisms become outdated and ineffective.
- Another major issue is the **inability to accurately classify different types of attacks**. Many existing systems can identify whether traffic is malicious but struggle to categorize the nature of the attack, such as Denial of Service (DoS), Probe, Remote-to- Local (R2L), or User-to-Root (U2R). Accurate classification is essential for taking appropriate countermeasures and improving response strategies.

- Despite their widespread use, these traditional approaches suffer from several significant limitations. One major drawback is their inability to detect zero-day attacks, which exploit previously unknown vulnerabilities. Additionally, high false positive and false negative rates reduce the reliability of these systems, making it difficult for security administrators to distinguish between genuine threats and normal activities. Furthermore, these systems lack adaptability and struggle to keep up with the rapidly evolving nature of cyber threats.
- To overcome these challenges, machine learning has emerged as a powerful and intelligent solution in the field of cybersecurity. Machine learning models can analyze vast amounts of network data, identify hidden patterns, and make accurate predictions without being explicitly programmed for specific attack signatures. By leveraging supervised and unsupervised learning techniques, these models can continuously improve their performance and adapt to new types of threats.
- In this context, the proposed research focuses on developing an advanced Network Intrusion Detection System (NIDS) that integrates multiple machine learning algorithms to enhance detection accuracy and efficiency. Unlike traditional systems, the proposed approach not only detects malicious activities but also classifies different types of attacks, enabling more precise and effective responses.
- Furthermore, this research incorporates a forensic logging mechanism to address another critical gap in existing systems. Most intrusion detection systems focus solely on detection and lack the capability to securely store and manage evidence of cyber incidents. The absence of proper logging mechanisms makes it difficult to perform post-attack analysis, trace the origin of attacks, and support legal investigations.

V. METHODOLOGY

The proposed system follows a structured approach consisting of data collection, preprocessing, model training, and real-time detection.

- Data Collection
- The NSL-KDD dataset is used, which contains 41 features representing various network traffic attributes.
- Data Preprocessing
- Preprocessing steps include:
- Handling missing values
- Encoding categorical features
- Feature scaling using StandardScaler
- The following machine learning algorithms are implemented:
- Model Training

- Random Forest
- Decision Tree
- K-Nearest Neighbors (KNN)
- Naive Bayes
- Logistic Regression
- Forensic Logging
- A secure logging mechanism is implemented using:
- SHA-256 hashing for data integrity
- UUID for unique event identification

VI. SYSTEM DESIGN

The system follows a three-tier architecture consisting of frontend, backend, and database layers.

Architecture

Frontend: React / Next.js

Backend: Flask API

Database: SQLite

Results

The system achieved approximately 76% accuracy and successfully classified multiple types of attacks in real time.

Advantages

- Detects unknown and zero-day attacks
- Enables real-time monitoring
- Ensures secure forensic logging
- Provides a scalable architecture

This paper presents a machine learning-based intrusion detection system integrated with forensic logging. The proposed system enhances cybersecurity by enabling real-time threat detection and secure evidence storage.

VII. RESULTS AND DISCUSSION

The system was evaluated using the NSL-KDD dataset.

Table I: Performance Comparison

Model	Accuracy	Remarks
Random Forest	76%	Best performance
Decision Tree	72%	Good interpretability
KNN	70%	High computation cost
Naive Bayes	68%	Fast but less accurate
Logistic Regression	69%	Baseline model

VIII. CONCLUSION

- This paper presents a comprehensive Machine Learning-based Network Intrusion Detection System (NIDS) integrated with a secure forensic logging mechanism to address the growing challenges in modern cybersecurity. The proposed system effectively overcomes the limitations of traditional intrusion detection approaches by leveraging intelligent algorithms capable of learning from data and adapting to new and evolving threats.
- By utilizing multiple machine learning models such as Random Forest, Decision Tree, K-Nearest Neighbors, Naive Bayes, and Logistic Regression, the system demonstrates improved accuracy in detecting and classifying various types of cyberattacks. The use of the NSL-KDD dataset ensures a standardized evaluation, and performance metrics such as accuracy, precision, recall, and F1-score validate the effectiveness of the proposed approach.
- One of the key contributions of this research is the integration of a **forensic logging framework**, which enhances the reliability and trustworthiness of the system. The implementation of SHA-256 hashing ensures that all logged events remain tamper-proof, while the use of unique identifiers enables efficient tracking and analysis of security incidents. This feature not only supports real-time monitoring but also plays a crucial role in post-attack investigation and digital forensics.
- Furthermore, the inclusion of a web-based dashboard provides administrators with real-time visibility into network activities, enabling faster detection and response to potential threats. The combination of intelligent detection, secure logging, and interactive visualization makes the system a holistic solution for network security management.

IX. REFERRENCE

[1] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani,

"NSL-KDD: A New Intrusion Detection Dataset,"

Proceedings of the 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA), 2009.

[2] T. M. Mitchell,

Machine Learning,

New York, NY, USA: McGraw-Hill, 1997.

[3] L. Breiman,

"Random Forests,"

Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[4] J. R. Quinlan,

"C4.5: Programs for Machine Learning,"

San Francisco, CA, USA: Morgan Kaufmann, 1993.

[5] D. Dua and C. Graff,

"UCI Machine Learning Repository,"

University of California, Irvine, 2019. [Online]. Available: http://archive.ics.uci.edu/ml

[6] F. Chollet,

Deep Learning with Python,

Shelter Island, NY, USA: Manning Publications, 2017.

[7] W. Stallings,

Network Security Essentials: Applications and Standards, 6th ed., Pearson, 2017.

[8] A. Scikit-learn Developers,

"Scikit-learn: Machine Learning in Python,"

Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[9] S. Axelsson,

"Intrusion Detection Systems: A Survey and Taxonomy,"

Technical Report, Chalmers University, Sweden, 2000.

[10] National Institute of Standards and Technology (NIST),

"Guide to Intrusion Detection and Prevention Systems (IDPS),"

Special Publication 800-94, 2007.