INTRUSION DETECTION SYSTEM USING MACHINE LEARNING FOR NETWORK TRAFFIC ANALYSIS

K. PAVANI

UG Student, Dept. Computer Science and Engineering Vignan's Institute of Management and Technology for Women email: konampavani93@gmail.com

M. KEERTHANA,

UG Student, Dept. Computer Science and Engineering Vignan's Institute of Management and Technology for Women email: keerthanamushke@gmail.com

M. SATHWIKA

UG Student, Dept. Computer Science and Engineering Vignan's Institute of Management and Technology for Women email: msathwika57@gmail.com

M. MANASA

(UG student department of computer science and engineering Vigan's institute of management and technology for women) email: medharimetlamanasa@gmail.com

GUIDE-AMULYA RACHANA

Assistant Professor, Dept. Computer Science and Engineering Vignan's Institute of Management and Technology for Women, email: amulya@vmtw.in

(Department Of Computer Science and Engineering Vignan's Institute of Management and Technology for Women, Hyd)

ABSTRACT

Intrusion Detection Systems (IDS) play a vital role in safeguarding networks by monitoring and analysing traffic to identify potential attacks. This research focuses on the development of an IDS using machine learning techniques for network traffic analysis. By extracting and preprocessing features from network data, various supervised and unsupervised learning algorithms are applied to classify normal and anomalous traffic patterns. With the rapid growth of computer networks and internet usage, cybersecurity threats have become increasingly sophisticated, making traditional security mechanisms insufficient for detecting malicious

activities. Intrusion Detection Systems (IDS) play a vital role in safeguarding networks by continuously monitoring and analyzing network traffic to identify potential security breaches. Machine learning models, such as Decision Trees, Random Forests, Support Vector Machines, and Deep Neural Networks, are evaluated for accuracy, detection rate, and false alarm rate. Experimental results demonstrate that machine learning significantly enhances the capability of IDS to detect known and unknown intrusions compared to traditional rule-based approaches. The study highlights the importance of feature selection, dataset quality, and model optimization in building robust intrusion detection systems. This work contributes to advancing intelligent, adaptive, and scalable security solutions for modern network environments.

1. INTRODUCTION

With the rapid expansion of the internet and interconnected digital systems, cybersecurity has become a major concern for individuals, organizations, and governments. Network systems are continuously exposed to various cyber threats such as denial-of-service (DoS) attacks, phishing, malware injection, ransomware, and unauthorized access attempts. Traditional security mechanisms like firewalls and signature-based Intrusion Detection Systems (IDS) are no longer sufficient to handle the complexity and evolving nature of modern attacks.

An Intrusion Detection System (IDS) is a security mechanism designed to monitor network traffic and detect suspicious activities that may indicate a cyberattack or policy violation. IDS can be classified into two main types: Host-based IDS (HIDS) and Network-based IDS (NIDS). In this research, the focus is on Network-based IDS, which analyzes real-time network traffic data to identify malicious patterns.

Machine Learning (ML) has emerged as a powerful approach to improve IDS performance. Unlike traditional rule-based systems, ML-based IDS can learn from historical data and adapt to new and unseen attack patterns. By training models on labeled network traffic datasets, the system can effectively classify traffic as normal or malicious.

This study aims to develop an intelligent IDS using machine learning algorithms for network traffic analysis. Various supervised and unsupervised learning techniques such as Decision Trees, Random Forest, Support Vector Machines (SVM), and Deep Neural Networks are applied and evaluated. The system focuses on improving detection accuracy, reducing false alarm rates, and enhancing adaptability to evolving cyber

threats. The proposed approach contributes to building a more secure, scalable, and intelligent cybersecurity framework for modern networks

2. EXISTING SYSTEM

Traditional Intrusion Detection Systems primarily rely on signature-based and rule-based approaches to identify malicious activities in network traffic. In these systems, predefined rules or known attack signatures are used to detect intrusions. When incoming traffic matches an existing signature in the database, it is flagged as malicious.

One of the main advantages of existing systems is their high accuracy in detecting previously known attacks. They are also computationally efficient and easy to implement in real-time environments. However, these systems suffer from significant limitations when dealing with modern and evolving cyber threats.

The most critical drawback of traditional IDS is their inability to detect zero-day attacks—new and unknown threats that do not match any existing signature. Since attackers continuously modify their techniques, rule-based systems fail to adapt dynamically. Additionally, maintaining and updating large signature databases is time-consuming and resource-intensive.

Another major issue is the high false negative rate, where malicious activities go undetected because they do not match predefined rules. This creates serious security vulnerabilities in critical systems. Moreover, these systems often generate a large number of alerts, including false positives, leading to alert fatigue among security analysts.

Traditional IDS also lacks scalability in handling large volumes of high-speed network traffic generated in modern cloud-based and distributed systems. As network environments become more complex, these systems struggle to maintain performance and accuracy.

Due to these limitations, there is a strong need for intelligent and adaptive security solutions. This has led to the adoption of machine learning-based approaches, which can analyze patterns in data, learn from experience, and detect both known and unknown attacks more effectively than traditional methods.

3. PROPOSED SYSTEM

The proposed Intrusion Detection System (IDS) leverages machine learning techniques to improve the accuracy and adaptability of network traffic analysis. Unlike traditional systems that rely on fixed rules, the proposed model learns from data

patterns and continuously improves its detection capability.

The system is designed to classify network traffic into two categories: normal and malicious. It uses a combination of supervised and unsupervised machine learning algorithms such as Decision Trees, Random Forest, Support Vector Machines (SVM), and Deep Neural Networks. These algorithms are trained using labeled datasets containing various types of network traffic.

The proposed system includes several key components: data collection, preprocessing, feature extraction, model training, and classification. Raw network traffic data is first collected and cleaned to remove noise and inconsistencies. Relevant features such as protocol type, packet size, source IP, destination IP, and traffic duration are extracted for analysis.

One of the major improvements in the proposed system is feature selection, which helps in identifying the most relevant attributes for better model performance. This reduces computational complexity and increases prediction accuracy.

The trained models are evaluated based on performance metrics such as accuracy, precision, recall, F1-score, and false alarm rate. Among the tested models, ensemble methods like Random Forest often show superior performance due to their ability to reduce overfitting and handle large datasets effectively.

The proposed system is scalable and adaptable to evolving cyber threats. It is capable of detecting both known and unknown attacks, making it more robust compared to traditional IDS. Overall, the system provides an intelligent, automated, and efficient solution for modern network security challenges.

4. METHODOLOGY

The methodology of the proposed Intrusion Detection System involves several systematic stages to ensure accurate classification of network traffic. The process begins with dataset collection, where network traffic data is obtained from publicly available datasets such as KDD Cup 99, NSL-KDD, or real-time network logs.

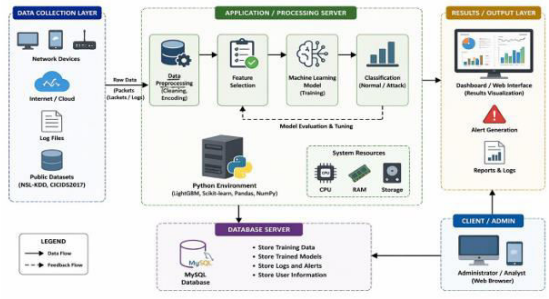
The next stage is data preprocessing, which plays a crucial role in improving model performance. This step involves handling missing values, removing duplicates, normalizing data, and converting categorical variables into numerical formats using encoding techniques. Data balancing techniques such as SMOTE may also be used to handle class imbalance between normal and attack data.

After preprocessing, feature extraction and selection are performed. Relevant features such as protocol type, service type, packet duration, and byte count are selected. Feature selection techniques like correlation analysis or recursive feature elimination are applied to improve efficiency and reduce redundancy.

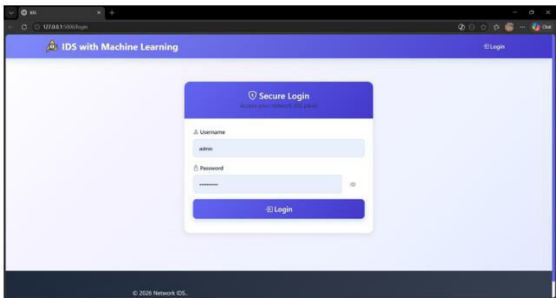
The dataset is then split into training and testing sets, typically in a 70:30 or 80:20 ratio. Machine learning models such as Decision Tree, Random Forest, SVM, and Neural Networks are trained using the training data. Each model learns patterns that distinguish normal traffic from malicious activity.

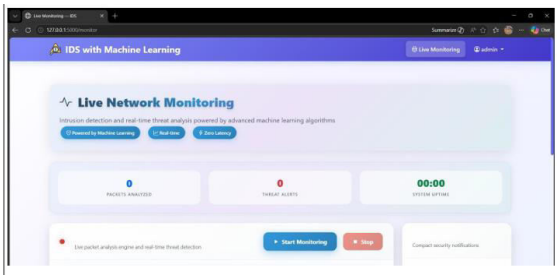
Once trained, the models are tested using unseen data. Performance evaluation is conducted using metrics such as accuracy, precision, recall, F1-score, and confusion matrix analysis. This helps in identifying the most effective model for intrusion detection.

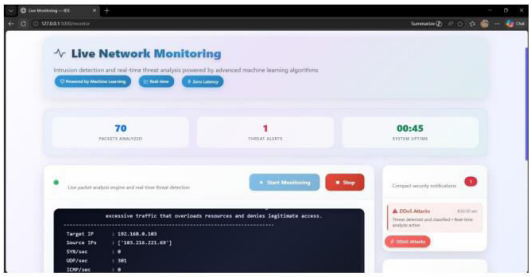
Finally, hyperparameter tuning techniques like grid search or random search are applied to optimize model performance. The methodology ensures that the system is efficient, accurate, and capable of detecting both known and unknown network intrusions effectively.



5. RESULTS AND DISCUSSION







6. CONCLUSION

In conclusion, The Intrusion Detection System using Machine Learning for Network Traffic Analysis presents an effective and modern approach to securing computer networks. By combining techniques such as data preprocessing, feature selection, and advanced machine learning algorithms like LightGBM, the system is able to accurately identify and classify network traffic as normal or malicious. This significantly improves detection performance compared to traditional rule-based systems, which are limited to identifying only known attacks. The system is designed to handle large volumes of data efficiently and supports real-time monitoring, making it suitable for practical deployment in modern network environments. Through proper training and evaluation using standard datasets, the model achieves high accuracy while minimizing false positives and false negatives. The use of optimization techniques and validation methods ensures that the system remains reliable and consistent. Various testing methods, including unit testing, integration testing, functional testing, system testing, and acceptance testing, were performed to validate the system. These tests confirm that all modules work correctly both individually and as a complete system. The system also provides timely alerts, enabling administrators to respond quickly to potential threats and prevent damage. In addition, the system is flexible and can be extended to support new datasets, updated algorithms, and evolving cyber threats. Future improvements may include the integration of deep learning models, cloud-based deployment, and enhanced visualization dashboards for better monitoring and analysis. Overall, this project demonstrates the importance of machine learning in cybersecurity and provides a scalable, efficient, and intelligent solution for intrusion detection in modern networks.

7. FUTURE WORK

Deep Learning-Based Enhancement: This type focuses on improving the system using advanced deep learning models such as CNN, RNN, and LSTM. These models can automatically learn complex patterns from large datasets and provide higher accuracy in detecting unknown and sophisticated cyber-attacks. Real-Time Detection Systems: This type aims to implement the system for real-time network monitoring. It continuously analyzes live network traffic and detects intrusions instantly, enabling quick response and reducing potential damage. Cloud-Based Intrusion Detection: In this type, the system is deployed on cloud platforms. It allows scalable storage, faster processing, and remote access. Cloud integration helps in handling large-scale

network environments efficiently. Hybrid Detection Models: This approach combines multiple machine learning algorithms to improve performance. By using more than one model, the system can achieve better accuracy and reduce false positives and false negatives. IoT-Based Security Systems: This type extends intrusion detection to Internet of Things (IoT) devices. Since IoT devices are highly vulnerable to attacks, this enhancement helps in securing smart homes, industries, and connected devices. Automated Response Systems: This type focuses on automatic actions after detecting an intrusion. The system can block malicious IP addresses, isolate affected systems, or trigger security protocols without human intervention. Encrypted Traffic Analysis: This enhancement deals with analyzing encrypted network data. Since many modern attacks use encrypted channels, this feature helps in detecting hidden threats without compromising data privacy.

REFERENCE

[1] Sumaiya Thaseen, I., & Kumar, C. A., "Intrusion Detection Model Using Fusion of Chi square Feature Selection and Multi-class SVM," Journal of King Saud University – Computer and Information Sciences.

[2] Hamamoto, A. H., Carvalho, L. F., Sampaio, L. D., Abrão, T., & Proença, M. L., "Network Anomaly Detection System Using Genetic Algorithm and Fuzzy Logic," Expert Systems with Applications.

[3] Hamed, T., Dara, R., & Kremer, S. C., "Network Intrusion Detection System Based on Recursive Feature Addition and SVM," IEEE Transactions.

[4] Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A., "A Detailed Analysis of the KDD Cup 99 Data Set," IEEE Symposium on Computational Intelligence for Security and Defense Applications.

[5] NSL-KDD Dataset, "Improved Version of KDD Cup 99 Dataset for Intrusion Detection," Available Online.

[6] Dua, D., & Graff, C., "UCI Machine Learning Repository," University of California, Irvine.

[7] Scikit-learn Documentation, "Machine Learning in Python," Available at: https://scikit learn.org

[8] Ke, G., Meng, Q., Finley, T., et al., "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," NeurIPS Conference.

[9] Goodfellow, I., Bengio, Y., & Courville, A., "Deep Learning," MIT Press.

[10] Stallings, W., “Network Security Essentials: Applications and Standards,” Pearson Education.H. ErdincKocer and K. KursatCevik, "Artificial neural networks-based vehicle license plate recognition," Procedia Computer Science, vol. 3, pp. 1033-1037, 2011