Detecting Network Attacks Using Machine Learning

Priya Yadav[1], Trisha Sharma[2], Himanshi Naval[3]
[1,2,3]*Department of Computer Science & Engineering, Rajasthan College of Engineering for Women, Jaipur (RAJ), India*

Abstract- As technology continues to evolve, the number of devices connected to the internet is increasing at an exponential rate. This rapid growth has made modern networks more complex and, consequently, more vulnerable to cyberattacks.

Ensuring network security has therefore become a critical research priority. Traditional Intrusion Detection Systems (IDS) depend heavily on predefined rules and signature-based methods, which struggle to detect novel or previously unseen attacks. To overcome these limitations, Machine Learning (ML) offers an intelligent and adaptive approach capable of identifying hidden patterns, learning from network behavior, and detecting anomalies that may indicate malicious activity. This paper explores the application of various ML techniques in detecting network attacks, comparing the performance of algorithms such as Decision Trees, Random Forests, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Deep Neural Networks. It also analyzes how supervised, unsupervised, and hybrid learning models can enhance detection accuracy and reduce false alarms. Emphasis is placed on critical factors like dataset quality, feature selection, model training, and real-time adaptability to ensure robust performance in dynamic environments.

Furthermore, the study discusses the challenges faced by ML-based IDS, including data imbalance, high computational costs, and the need for continuous learning to handle evolving attack patterns. The paper concludes with recommendations for future research, highlighting the potential of deep learning, ensemble methods, and federated learning in developing more scalable, autonomous, and efficient network security systems. By leveraging ML, this research aims to pave the way for next-generation intrusion detection systems capable of defending against the ever-changing landscape of cyber threats.

Keywords ***Intrusion Detection System, Network Security, Machine Learning, Anomaly Detection, Deep Learning, Cybersecurity.***

I.INTRODUCTION

The increasing use of online systems for communication, banking, shopping, education, and data storage has made computer networks an indispensable part of modern society. As digital connectivity continues to grow, both individuals and organizations are becoming increasingly dependent on secure and reliable network infrastructures. However, this dependence has also given rise to a surge in cyber threats, where malicious actors attempt to exploit vulnerabilities to steal sensitive information, disrupt critical services, or compromise system integrity.

These attacks can range from simple phishing attempts to complex, large-scale intrusions targeting enterprise or government networks.

Detecting and preventing such attacks has become a major challenge due to the enormous volume, velocity, and variability of network data. Traditional Intrusion Detection Systems (IDS) based on fixed rules or known attack signatures are often unable to detect previously unseen or evolving attack patterns. As a result, there is a pressing need for more adaptive and intelligent security solutions capable of analyzing network traffic in real time and identifying suspicious behaviors that deviate from normal activity.

Machine Learning (ML) offers a powerful solution to this challenge by enabling systems to automatically learn from data, identify hidden patterns, and make predictions without being explicitly programmed. ML-based intrusion detection can recognize both known and unknown threats by continuously updating its models as new attack patterns emerge. Furthermore, the integration of Deep Learning (DL) and hybrid approaches combining anomaly and signature-based detection has shown great promise in improving detection accuracy and reducing false positives.

The study also examines current challenges in ML-

based intrusion detection— such as data imbalance, model interpretability, and scalability—and discusses future directions, including federated learning, real-time IDS deployment, and adaptive cybersecurity frameworks designed to protect against ever-evolving digital threats.

II. LITERATURE REVIEW

Researchers have been investigating methods to detect network intrusions since the early days of the internet. One of the earliest conceptual frameworks was proposed by Dorothy Denning in 1987, which introduced the idea of anomaly detection in computer security monitoring. This foundational concept inspired decades of research into systems that could automatically identify deviations from normal network behavior. Over time, a variety of algorithms and analytical approaches have been developed to improve the accuracy, efficiency, and adaptability of Intrusion Detection Systems (IDS).

Machine Learning (ML) techniques have significantly transformed this field by enabling systems to learn from historical data and generalize to unseen patterns. Classical ML models such as Decision Trees, Random Forests, Support Vector Machines (SVM), Naïve Bayes, and K-Nearest Neighbors (KNN) have been successfully applied to distinguish between normal and malicious traffic. Studies have shown that these models achieve high detection rates when properly trained on representative datasets. However, their performance can be limited by feature selection quality, data imbalance, and the inability to capture complex attack relationships.

In recent years, the focus has shifted toward Deep Learning (DL) approaches, which can automatically extract hierarchical and abstract features from raw network data. Models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have shown remarkable performance in detecting sophisticated attacks, particularly in scenarios involving temporal or sequential data patterns. Advanced architectures like Long Short-Term Memory (LSTM) networks and Autoencoders have further enhanced anomaly detection by modeling the temporal dependencies and reconstructing normal traffic patterns to identify outliers. Several benchmark datasets are widely used to evaluate and compare IDS models, including KDD Cup 99, NSL-KDD, UNSW-NB15, and CICIDS2017. These datasets contain a mixture of normal and attack records, covering multiple intrusion categories such as Denial of Service (DoS), Probe, Remote-to-Local (R2L), and User-to-Root (U2R) attacks. Researchers have also explored ensemble learning techniques that combine multiple classifiers to enhance robustness and accuracy.

III. METHODOLOGY

The process of detecting network attacks using machine learning generally involves several key steps that ensure data reliability, model accuracy, and efficient detection performance.

1.Data Collection: Network traffic data is collected from publicly available benchmark datasets such as NSL-KDD, CICIDS2017, or from real-time network monitoring systems. These datasets contain features like protocol type, service, flag, packet length, and connection duration, which help differentiate between normal and malicious traffic patterns.

2.Data Preprocessing: Since raw network data often contains missing, redundant, or irrelevant values, preprocessing is a crucial step. It includes cleaning and filtering the data, encoding categorical variables, normalizing numerical attributes, and selecting the most relevant features for model training. Proper preprocessing enhances both model performance and detection accuracy.

3.Model Selection and Training: Various machine learning algorithms are trained to recognize patterns that distinguish normal activity from potential attacks. Models such as Decision Tree, Random Forest, Support Vector Machine (SVM), and Neural Networks are implemented and compared based on their predictive capabilities and computational efficiency.

Evaluation Metrics: The performance of each model is assessed using metrics such as Accuracy, Precision, Recall, F1-Score, and Receiver Operating Characteristic (ROC). A high Recall value ensures that most attacks are correctly detected, while strong Precision helps prevent normal traffic from being

misclassified as malicious, ensuring balanced and reliable detection.

IV.RESULTS AND DISCUSSION

Experiments conducted on benchmark datasets such as NSL-KDD and CICIDS2017 demonstrate that machine learning algorithms substantially enhance detection performance compared to traditional rule-based intrusion detection systems. The experimental outcomes indicate that models like Random Forest and Neural Networks achieve the highest detection accuracy, often exceeding 95%, whereas algorithms such as Support Vector Machines (SVM) and Decision Trees yield slightly lower yet consistent results. The Random Forest model performs particularly well because it integrates multiple decision trees through ensemble learning, thereby minimizing overfitting and improving generalization across different attack categories.

Deep learning approaches, including Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, demonstrate even greater potential by automatically learning complex spatial and temporal dependencies in network traffic data. These models can effectively capture subtle attack patterns that may go unnoticed by shallow classifiers. However, their high computational requirements, longer training times, and need for large labeled datasets remain key challenges for real-time deployment.

The evaluation metrics highlight the efficiency of ML-based IDS, with high Precision and Recall demonstrating accurate attack detection and minimal false positives.

Ensemble and hybrid models that combine anomaly and signature-based techniques offer greater adaptability against both known and zero-day attacks. Overall, ML-based intrusion detection systems enhance accuracy, responsiveness, and adaptability to evolving network threats. With proper optimization and infrastructure, they can support real-time monitoring and automated threat response. Future improvements such as federated learning and lightweight deep models may further increase scalability and efficiency in large-scale network environments.

V. CHALLENGES AND FUTURE SCOPE

Although Machine Learning (ML) techniques offer numerous advantages in intrusion detection, several challenges still hinder their full-scale implementation in real-world environments. Data imbalance remains a major concern, as normal network traffic records greatly outnumber attack samples. This imbalance often leads to biased models that overlook rare but critical intrusions. Another significant challenge is the high computational cost of training and deploying complex models, especially deep learning architectures that require substantial processing power and memory resources. Implementing such systems on high-speed or large-scale networks demands efficient optimization, hardware acceleration, and resource management. Moreover, many ML models operate as “black boxes”, providing limited insight into their internal decision-making process, which reduces trust and interpretability. Future research can address this by integrating Explainable Artificial Intelligence (XAI) methods to make model predictions more transparent and understandable for security analysts. Federated Learning also presents a promising direction, enabling multiple organizations to collaboratively train models without exchanging sensitive data, thus preserving privacy while improving detection accuracy.

Additionally, combining ML-based IDS with real-time threat intelligence, adaptive learning frameworks, and edge computing can enhance responsiveness and scalability. As cyber threats continue to evolve, the integration of lightweight deep learning models, self-learning systems, and continuous model updating will play a vital role in developing more secure, efficient, and intelligent network defense mechanisms for the future.

VI.CONCLUSION

Machine Learning (ML) has revolutionized the field of network security by introducing intelligent systems capable of learning from data and adapting to emerging threats. Unlike traditional signature-based intrusion detection systems, ML-based approaches can detect both known and unknown attack patterns with higher accuracy and efficiency. By continuously analyzing network behavior, these models enhance the ability to identify evolving and sophisticated cyber threats, offering stronger and more proactive protection for organizations.

Despite these advancements, challenges such as data imbalance, model complexity, and lack of interpretability continue to pose limitations. However, the integration of Deep Learning (DL), Explainable AI (XAI), and adaptive learning frameworks shows great promise in addressing these issues and improving trust, transparency, and scalability. As cyber threats become more dynamic, continuous research and real-world implementation of these intelligent systems will be essential to building resilient, autonomous, and secure network infrastructures.

REFERENCES

[1] D. E. Denning, "An Intrusion Detection Model," *IEEE Transactions on Software Engineering*, vol. 13, no. 2, pp. 222–232, 1987.

[2] A. Kumar, et al., "Network Intrusion Detection Using Random Forest Algorithm on NSL-KDD Dataset," *International Journal of Computer Applications*, vol. 182, no. 32, pp. 1–5, 2019.

[3] R. Sommer and V. Paxson, "Outside the Closed World: On Using Machine Learning for Network Intrusion Detection," in *IEEE Symposium on Security and Privacy*, 2010, pp. 305–316.

[4] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization," in *Proceedings of the International Conference on Information Systems Security and Privacy (ICISSP)*, 2018, pp. 108–116.

[5] A. L. Buczak and E. Guven, "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1153–1176, 2016.

[6] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set," in *IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA)*, 2009, pp. 1–6.

[7] N. Moustafa and J. Slay, "UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems (UNSW-NB15 Network Data Set)," in *Military Communications and Information Systems Conference (MilCIS)*, 2015

[8] W. Wang, Y. Sheng, J. Wang, X. Zeng, and X. Ye, "HAST-IDS: Learning Hierarchical Spatial-Temporal Features Using Deep Neural Networks to Improve Intrusion Detection," *IEEE Access*, vol. 6, pp. 1792–1806, 2018.

[9] S. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, "A Deep Learning Approach to Network Intrusion Detection," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, no. 1, pp. 41–50, 2018.