JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A Survey on Network Intrusion Detection System using Machine Learning

[1]Pratham Doke, [2]Nikita Gokhale, [3]Samruddhi Kadam, [4]Akash Kalme, [5]Preeti Jain

[1]Student, [2]Student, [3]Student, [4]Student, [5]Assistant Professor
[1-5]Department of Computer Engineering
[1-5]Pune Institute of Computer Technology, Pune, India

Abstract: The rapid evolution of the internet and communication technologies has led to a significant expansion in network size and the associated data volume. Consequently, this surge has given rise to new and sophisticated forms of cyber-attacks, which present considerable challenges for maintaining the security and integrity of networks. In this context, the presence of intruders seeking to launch various malicious attacks within networks cannot be underestimated. To counteract these threats, Intrusion Detection Systems (IDS) play a pivotal role by scrutinizing network traffic to ensure the confidentiality, integrity, and availability of data. However, despite extensive research efforts, IDS still struggles with the need to enhance detection accuracy while reducing false alarm rates and addressing emerging forms of intrusions. Recently, the application of machine learning (ML) and deep learning (DL) techniques has emerged as a promising avenue to strengthen network-based IDS (NIDS) systems, aiming to detect intrusions efficiently. Machine Learning enhances NIDS by improving accuracy, adaptability to new threats, reducing false positives, and enabling the detection of complex and subtle anomalies. It also automates threat detection, reduces manual rule maintenance, and provides real-time monitoring, ultimately enhancing an organization's overall security posture.

Index Terms **- Deep Learning, Machine Learning, Network Anomaly Detection, Network Intrusion Detection System, Network Security.**

I. INTRODUCTION

In today's digital age, where information technology plays an indispensable role in our daily lives, ensuring the security and integrity of computer networks is of paramount importance. The interconnected nature of these networks and the rapid growth of internet usage have brought about significant vulnerabilities, making them potential targets for malicious activities. As a result, Network Intrusion Detection Systems (NIDS) have become a critical component in safeguarding the digital infrastructure of organizations and individuals alike.

NIDS is a vital cybersecurity tool designed to monitor and analyze network traffic, seeking signs of unauthorized access, malware, data breaches, and other malicious activities. Its primary purpose is to detect and respond to potential security threats swiftly and effectively. NIDS operates by examining packets of data flowing through a network and comparing them against predefined patterns, signatures, or behavior profiles indicative of intrusions. When suspicious or malicious activities are identified, NIDS can trigger alerts or automated actions to mitigate the threat.

Cyber threats are evolving at an alarming pace, with attackers continuously developing new and sophisticated methods to breach network defenses. NIDS is essential to identify emerging threats and vulnerabilities proactively. With the proliferation of sensitive data stored and transferred through networks, protecting this information from unauthorized access, data breaches, or leaks is paramount. NIDS plays a crucial role in detecting and preventing data compromises. Many industries and organizations are subject to regulatory compliance standards that mandate robust network security measures. NIDS helps organizations meet these requirements by providing continuous monitoring and threat detection capabilities. NIDS not only detects intrusions but also allows for swift response and mitigation, minimizing the potential damage from cyberattacks. Early detection is vital in preventing widespread network compromises. NIDS can help organizations optimize resource allocation by focusing on real threats, reducing false positives, and automating certain response actions. This can lead to cost savings in the long run. NIDS provides organizations with a comprehensive view of their network traffic, helping them understand usage patterns, vulnerabilities, and potential points of entry for attackers. This visibility can lead to enhanced accountability in network security management.

II. NIDS

NIDS stands for Network Intrusion Detection System. It is a security tool used to monitor and analyze network traffic for signs of malicious activities or unauthorized access attempts. NIDS aims to detect and respond to potential security threats by examining network packets and identifying patterns or anomalies that could indicate cyber-attacks.[1]

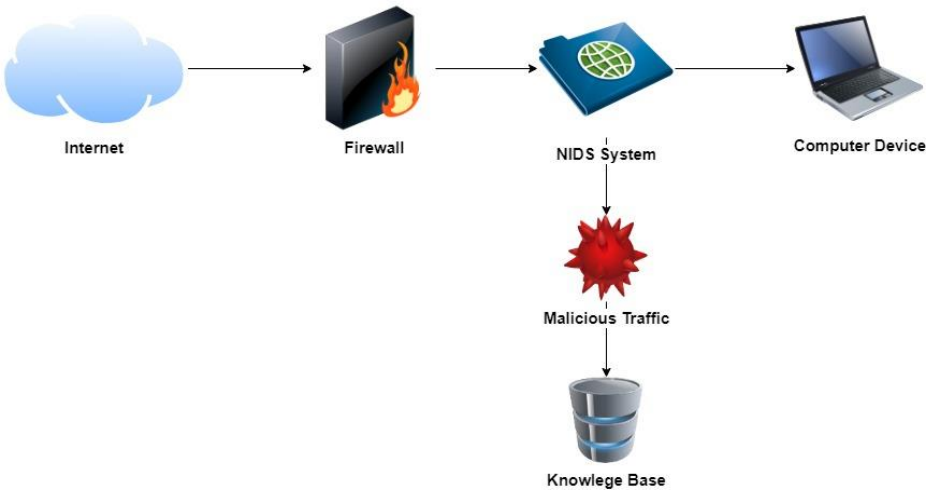

Fig. 1: Network Intrusion Detection System (NIDS)

2.1. Functions of NIDS

***i.* Packet Analysis:** NIDS examines network traffic by analyzing individual packets to identify suspicious and malicious activities based on predefined signature or behavioral pattern.[3]

***ii.* Signature-based Detection:** NIDS used predefined signatures or patterns of known attacks to identify similar patterns in network traffic. When a match is found, it triggers an alert or takes predefined action to mitigate the threat.[7]

***iii.* Anomaly-based Detection:** NIDS establishes a baseline of normal network behavior and flags any deviations from this baseline as potential threats. This helps to detect unknown attacks or evolving attacks.[3]

***iv.* Alter generations:** NIDS establishes alerts or notifications when it detects suspicious activities. These alerts are then sent to the security administrator.[3]

2.2. Types of Attacks

***i.* DoS (Denial of Service) and DDoS (Distributed Denial of Service) Attacks:** DoS attack overwhelms a network system with traffic, causing it to become unavailable or slow legitimate users. DDoS attack utilizes multiple compromised devices to flood the target with a massive volume of traffic, making it inaccessible.[4]

***ii.* R2L-Attacks:** Stands for 'Remote-to-Local' attack is a type of attack where an adversary attempts to gain unauthorized access to a system from a remote location, typically over a network, and escalate their privileges to gain local access or control on a targeted system.[4]

***iii.* Probing Attacks:** A probing attack, also known as networks probing, is an attempt by an unauthorized user or system to gather information about a targeted network, system, or infrastructure. The objective of a probing attack is to identify vulnerabilities, weakness or potential entry point into a network for further exploitation.[4]

2.3. Enhancement of NIDS using Machine Learning

Machine Learning (ML) and Network Intrusion Detection Systems (NIDS) can work together to enhance the detection and response capabilities against evolving cyber threats. ML algorithms can significantly improve the effectiveness and efficiency of NIDS by enabling more dynamic and adaptive threat detection. Here's a detailed explanation of how ML and NIDS can collaborate.[2]

***i.* Data Collection and Preprocessing:** NIDS collects network traffic data, which includes packets, protocols, and other relevant information. ML models require preprocessed and structured data. Data preprocessing involves cleaning, normalization, and feature extraction from the raw network data.[2]

***ii.* Feature Extraction:** ML algorithms require specific features to make predictions. NIDS extracts relevant features from the network data, such as packet size, source/destination IP addresses, protocol type, and timestamps.[2]

***iii.* Training Machine Learning Models:** NIDS utilizes labeled network traffic data to train ML models. Labels indicate whether the network traffic is benign or malicious. ML models, such as supervised learning classifiers (e.g., decision trees, support vector machines, neural networks), are trained using these features and labels to learn patterns and characteristics of normal and malicious traffic.[2]

***iv.* Model Training and Tuning:** ML models are trained and tuned using various algorithms to achieve optimal performance in terms of accuracy, precision, recall, and false positives/negatives. Hyperparameters are adjusted through techniques like cross-validation to enhance the model's ability to generalize and detect intrusions accurately.[2]

***v.* Anomaly Detection:** Anomaly-based NIDS utilizes ML models for anomaly detection. These models learn normal network behavior during the training phase. ML algorithms identify deviations from this learned behavior, flagging them as potential intrusions or anomalies.[2]

***vi.* Signature-based Detection:** ML can enhance signature-based NIDS by improving the accuracy and efficiency of signature matching.ML algorithms can learn and optimize signatures from historical data, adapting and updating them to new attack patterns.[2]

***vii.* Real-time Detection and Alerting:** ML-powered NIDS continuously monitors network traffic in real-time, classifying packets and identifying potential threats based on the trained models. Alerts are generated when the ML model detects deviations from normal behavior or matches against known attack patterns.[2]

***viii.* Adaptive Learning and Improvement:** ML models can be designed to adapt and evolve over time, improving their detection capabilities as they encounter new attack variants and network patterns. Continuous monitoring and feedback mechanisms allow the ML model to be updated and refined, ensuring it remains effective against evolving threats.[2]

2.4. Use Cases and Applications

i. **Anomaly Based Detection in Network Traffic:** ML-based NIDS can analyze normal network behavior and detect anomalies that deviate from the established baseline. Unusual patterns, such as a sudden increase in data transfer or irregular access times, could indicate a potential security threat.[8]

ii. **Network Forensics:** ML algorithms can assist in network forensics by analyzing historical network data to trace the origin and impact of a security incident. This information is crucial for understanding the nature of the attack, attributing responsibility, and preventing future occurrences.[8]

iii. **Cloud Security Monitoring:** As organizations migrate to cloud environments, ML-based NIDS can monitor and analyze network traffic in the cloud, ensuring the security of data and applications hosted on cloud platforms.

iv. **Behavioral Analysis for IoT Devices:** ML can be applied to analyze the behavior of devices in Internet of Things (IoT) networks. NIDS using ML can detect abnormal activities or unauthorized access in IoT environments, safeguarding against potential security risks.[6]

v. **Automated Incident Response:** ML-enabled NIDS can not only detect intrusions but also automate incident response processes. This includes isolating compromised systems, blocking malicious traffic, and alerting security teams for further investigation, reducing response time in the event of a security incident.[7]

III. Survey and Findings

In this section, we provide a comprehensive overview of research papers that delve into the convergence of network intrusion detection. These papers cover a wide spectrum of topics, including the fusion of machine learning and deep learning techniques with the network intrusion detection systems and innovative approaches for recognizing and countering various network threats. The primary aim of this survey is to provide valuable perspectives on the key findings, contributions, and common trends evident in these research papers.

3.1. Network Intrusion Detection Using Machine Learning Techniques

In the research [10], Yasmeen S. Almutairi et. al. dive into various methodologies that include evolutionary, information security, statistical and machine learning approaches that improve the performance and efficiency of the traditional NID systems. The dataset used, NSL-KDD dataset, has data related to DoS, R2L, U2R and Probing attacks that involve 41 features for each attack. The author used two ML approaches one is binary classification and other is multi-class classification. Out of both the approaches Random forest came with highest accuracy whereas Naïve Bayesian (NB) classifier showed less accuracy. Hence, Random forest proved to be more reliable in detecting attacks like DoS, R2L, U2R and Probs attacks.[10]

3.2. An Intrusion Detection System Integrating Network-Level Intrusion Detection and Host-Level Intrusion Detection

In the paper [6], Biswanath Mukherjee Et al. propose a Neural Network based hybrid framework which combines the properties of the Network-based and Host-based IDSs. Autoencoders are used for automatic feature extraction while neural network algorithms are used for classification. While for HIDS, word embedding methods from NLP which are then fed to one-dimensional convolutional neural networks to extract temporal features for classification.[6]

3.3. Network Intrusion Detection and its Strategic Importance

In the research [8], Muhammad K. Asif Et al. discuss the challenges of network security due to the rapidly increasing illegal activities. The concepts of anomaly-based and signature-based intrusion detection, and various methods and techniques used in intrusion detection, including neural networks, predictive pattern generation, genetic algorithms, fuzzy logic, immune systems, Bayesian methods, and decision trees are discussed in depth along with their advantages and disadvantages. IDS based on data mining techniques, where hidden predictive information is extracted from large and complex datasets are employed. Data fusion techniques are also integrated, emphasizing the integration of data from multiple sources and sensors to make inferences about network events, activities, and situations.[8]

3.4. Intelligent Intrusion Detection System Using Clustered Self Organized Map

According to the [7], Muder Almi'ani Et al. explored an unsupervised learning approach for intrusion detection using the NSL-KDD dataset. A hierarchical agglomerative clustered Self-Organizing Map (SOM) network to balance sensitivity and computational efficiency was employed to enhance the efficiency of NIDS. The clustering step reduced processing time and computational effort while maintaining a high detection rate. However, the system showed weaker sensitivity to normal activities, affecting overall accuracy.[7]

3.5. Machine Learning based Intrusion Detection System

According to the study [4], Anish Halimaa A Et al. mainly focus on the application of machine learning techniques, specifically Support Vector Machine (SVM) and Naive Bayes for intrusion detection. The authors analyze the performance of the above two algorithms using the NSL-KDD dataset and ultimately the SVM exhibits superior performance than Naive Bayes, particularly for large volumes of data. The evaluation is done through an analysis of accuracy and misclassification rates. Future work endeavors to improve the precision with extensive datasets and construct more efficient models.[4]

3.6. Network Intrusion Detection System: A Systematic Study of Machine Learning and Deep Learning Approaches

In the research [3], Zeeshan Ahmad Et al. provide a comprehensive review of recent NIDS-based articles along with their strengths and limitations. The purpose is to provide an overview of recent trends and advancements in ML and DL-based approaches for NIDS. Authors discuss the importance of training NIDS using suitable datasets, with a preference for Deep Learning (DL) over Machine Learning (ML). In essence, future trends focus on better NIDS frameworks, exploring less common DL algorithms and hybrid approaches for feature extraction and classification. This review is beneficial for researchers aiming to develop lightweight and efficient ML and DL based NIDS.[3]

3.7. Enhancement of Intrusion Detection System using Machine Learning

In the paper [2], Mukesh Kumar Yadav Et al. proposes an ensemble learning model which combines all the weak classifiers and creates a strong classifier which can detect different types of attacks more precisely. This paper uses some common supervised and unsupervised learning algorithms like- Support Vector Machine, Logistic Regression, K-Means Clustering, etc. The dataset used, NSL-KDD dataset, has data related to DoS, R2L, U2R and Probing attacks. The results show that the proposed model which is AdaBoost with Logistic Regression provides a solution with 99.39% accuracy which is far better than the traditional NID systems.[2]

Table 3.1: Survey Findings with respect to Models

Model	Findings	Identified Gaps
Random Forest Classifier	High accuracy for DOS Attacks	Lower Accuracy for U2R Attacks
Clustering Self Organized Map (SOM)	High speed detection of attacks compared to supervised learning	Limited anomaly detection and not adaptable to dynamic network
Decision Tree	High accuracy for detection of known attacks	Not suitable for detection of new attacks
Support Vector Machine (SVM)	Gives good accuracy to wide range of attacks	Not capable to detect zero-day attack
Naïve Bayesian (NB) Classifier	Computationally efficient	High misclassification rate

IV. Conclusion

The importance of intrusion detection in network security cannot be overstated, with a particular focus on leveraging Machine Learning to enhance its efficacy. Existing systems encounter various challenges, such as speed, high false alarm rates, and real-time accuracy, which highlight the need for improvement in addressing low-frequency attacks in practical settings and mitigating damage during intrusions. It is evident that there is potential for enhancing the performance of intrusion detection models by employing simpler Deep Learning algorithms in the development of a more efficient Network Intrusion Detection System (NIDS). Our research aims to concentrate on zero-day attacks, where current systems often exhibit reduced effectiveness.

References

[1] Bane Raman Raghunath, Shivsharan Nitin Mahadeo, 'Network Intrusion Detection System (NIDS)', First International Conference on Emerging Trends in Engineering and Technology.

[2] Mukesh Kumar Yadav, Mahaiyo Ningshen, 'Enhancement of Intrusion Detection System using Machine Learning', International Journal of Engineering Research and & Technology (IJERT), January-2023.

[3] Zeeshan Ahmad, Adnan Shahid Khan, Cheah Wai Shiang, Johari Abdullah, Farhan Ahmad, 'Network intrusion detection system: A systematic study of machine learning and deep learning approaches', May 2020.

[4] Anish Halimaa A, Dr. K. Sundarakantham, 'MACHINE LEARNING BASED INTRUSION DETECTION SYSTEM', Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019) IEEE Xplore Part Number: CFP19J32-ART; ISBN: 978-1-5386-9439-8.

[5] Dikshant Gupta, Suhani Singhal, Shamita Malik, Archana Singh, 'Network Intrusion Detection System Using various data mining techniques', International Conference on Research Advances in Integrated Navigation Systems (RAINS - 2016), April 06-07, 2016, R. L. Jalappa Institute of Technology, Doddaballapur, Bangalore, India.

[6] Biswanath Mukherjee, L. Todd Heberlein, and Karl N. Levitt, 'Network Intrusion Detection Intrusion detection is a new, retrofit approach for providing a sense of security in existing computers and data networks, while allowing them to operate in their current "open" mode.', IEEE Network MayiJune 1994.

[7] Muder Almi'ani, Alia Abu Ghazleh, Amer Al-Rahayfeh, Abdul Razaque, 'Intelligent Intrusion Detection System Using Clustered Self Organized Map', 2018 Fifth International Conference on Software Defined Systems (SDS).

[8] Muhammad K. Asif, Talha A. Khan, Talha A. Taj, Umar Naeem, Sufyan Yakoob, Network Intrusion Detection and its Strategic Importance, 2013 IEEE Business Engineering and Industrial Applications Colloquium (BEIAC).

[9] Satish Kumar, Sunanda Gupta, Sakshi Arora, 'Research Trends in Network-Based Intrusion Detection Systems: A Review', November 2021.

[10] Yasmeen S. Almutairi, Bader Alhazmi, Amr A. Munshi, 'Network Intrusion Detection Using Machine Learning Techniques', Advances in Science and Technology Research Journal 2022, 16(3), 193–206.