International Scientific Journal Of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 03 | March – 2026 DOI:10.55041/ISJEM05906
An International Scholarly || Multidisciplinary || Open Access || Indexing In All Major Database & Metadata

Intelligent Network Intrusion Detection System using Ml

Guide: **M. Sowjanya (Phd)**

1. **V. Vasu** (5A7) 2. **M. Naveen** (578) 3. **R. Hareesh** (596) 4. **M. Afroz** (577)

Abstract - Network intrusion detection systems, or NIDS, are essential for defending computer networks from several types of security threats and assaults. There is a growing need within network attacks become more complicated and frequent, sophisticated analytics methods to improve NIDS's capacity for detection and reaction. The creation and application of analytics techniques for network intrusion detection systems is the main topic of this study. Utilizing these methods can help NIDS be more accurate, efficient, and effective at detecting and mitigating security breaches. The first part of the study looks at the foundational ideas of network intrusion detection, including the various kinds of attacks and the difficulties in detecting them. The advantages of several NIDS systems, such as signature-based and anomaly-based systems, are highlighted and restricted. The overall goal of this research is to enhance the field of network intrusion detection by using analytics approaches to increase NIDS's capabilities. The suggested techniques can decrease false positives, enhance automated incident response, process large amounts of data more effectively, and increase the accuracy of attack detection. In the face of constantly changing cyber threats, the research findings will aid in the creation of stronger and more efficient network security solutions.

Keywords - Network intrusion detection systems, Analytics techniques, Security breaches, False positives, Cyber threats

INTRODUCTION

In the contemporary digital landscape, the widespread proliferation of cyber threats poses significant challenges to security of information systems and digital assets. Cyberattacks come in various forms, ranging from malware and phishing to sophisticated network intrusions, necessitating robust defence mechanisms for threat detection and classification. ML algorithms have become essential assets for automating the assessment and categorization of cyber threats, offering the potential to enhance cybersecurity posture and mitigate risks effectively.

This study focuses on applying ML techniques to classify attack methods using a provided dataset. The goal is to develop accurate and reliable models capable of identifying and categorizing different attack vectors, thereby enabling proactive threat mitigation strategies. The research is structured into several modules, beginning with data preprocessing to handle missing values and encode categorical variables, followed by data visualization to gain insights into the dataset's characteristics and patterns.

Subsequently, three prominent ML algorithms, namely Random Forest, AdaBoost, and Extra Trees Classifier are implemented and assessments of their effectiveness in classifying attack methods involve selecting algorithms based on their capacity to effectively handle diverse datasets and their demonstrated performance in classification tasks across various domains. The performance of the models is assessed using evaluation metrics like accuracy, precision, recall, and F1-score alongside confusion matrices to analyze classification errors and biases.

LITERATURE SURVEY

[1] "An Efficient Hybrid Deep Learning-Based Intrusion Detection System": This paper likely presents a novel approach to intrusion detection by combining deep learning techniques with traditional methods. It may introduce a hybrid system that leverages the strengths of both approaches to enhance the efficiency and accuracy of intrusion detection. [2] "Network intrusion detection system: A systematic study of machine learning and deep learning approaches": The paper examines the efficacy of machine learning algorithms in network intrusion detection, presenting experimental evaluations across various metrics. It discusses algorithmic performance, dataset characteristics, and implications for enhancing cybersecurity

measures, culminating in recommendations for future research directions. Authored by Saddam Hossen and Anirudh Janagam, it contributes to the ongoing discourse on bolstering network security through advanced analytical techniques.

[3] "State-of-the-art in intrusion detection and future directions": This article likely offers an overview of the current state of intrusion detection systems and explores potential future directions in the field. It may discuss emerging trends, challenges, and advancements in intrusion detection technology. [4] "Network anomaly detection: Methods, systems, and tools": This paper surveys methods, systems, and tools for detecting network anomalies. It likely covers various approaches to anomaly detection in network traffic, including statistical methods, machine learning techniques, and system architectures. [5] "A survey on intrusion detection system and its data mining techniques": This survey paper likely provides an overview of intrusion detection systems and the data mining techniques employed in them. It may discuss different types of intrusion detection systems, data mining algorithms, and their applications in network security. [6] "A data mining framework for building intrusion detection models": This paper presents a data mining framework specifically designed for constructing intrusion detection models.

It likely discusses methodologies for gathering and preprocessing data, selecting appropriate features, and building models using data mining techniques to detect network intrusions effectively. [7] "Anomaly intrusion detection system: Techniques and challenges": This article likely provides an overview of anomaly-based intrusion detection systems, focusing on the techniques employed and the challenges faced in their implementation. It may discuss various anomaly detection methods, such as statistical approaches, machine learning algorithms, and hybrid models, along with the inherent difficulties in accurately detecting anomalous behavior.

[8] "A review on intrusion detection system using machine learning techniques": This paper likely offers a comprehensive review of intrusion detection systems (IDS) that leverage machine learning techniques. It may cover different machine learning algorithms employed in IDS, their advantages, limitations, and performance evaluations. Additionally, it may discuss recent advancements and challenges in the field of intrusion detection using machine learning. [9] N. Sharma, A. Sharma, and A. K. Singh, "Intrusion detection using hybrid machine learning techniques: A systematic review," in International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMIT on). This paper likely presents a systematic review of intrusion detection systems that employ hybrid machine learning techniques.

It may discuss the combination of different machine learning algorithms or approaches to enhance the accuracy and effectiveness of intrusion detection. [10] S. B. Zargar, J. Joshi, and D. Tipper, "A survey of defense mechanisms against distributed denial of service (DDoS) flooding attacks," IEEE Communications Surveys & Tutorials, vol. 15, no. 4, pp. 2046-2069. While not exclusively focused on intrusion detection systems, this survey paper likely covers defense mechanisms against various network attacks, including Distributed Denial of Service (DDoS) attacks. It may discuss intrusion detection as a component of broader network security strategies and highlight the challenges and approaches in mitigating DDoS attacks.

EXISTING SYSTEM

Conventional cybersecurity systems primarily rely on rule-based and signature-based methodologies for attack detection and classification. While effective against known threats, these systems encounter challenges in identifying novel or previously unseen attack vectors. Manual intervention is often required for rule updates and maintenance, resulting in delays and increased operational overhead. In contrast, machine learning-based systems leverage automated feature extraction and model training techniques. They exhibit the capability to detect new and emerging threats without explicit rule definitions, adapting and evolving. Machine learning based systems offer scalability by analyzing large datasets and learning from historical incidents. However, they face challenges such as data imbalance, adversarial attacks, and model drift, necessitating ongoing research and development efforts to address these issues. Integrating machine learning-based systems with existing security infrastructures and workflows may require careful planning and deployment strategies.

PROPOSED SYSTEM

The proposed system aims to enhance attack method classification by leveraging advanced machine learning algorithms and techniques. Building upon the limitations of existing rule-based and signature-based approaches, the proposed system introduces a more adaptive and scalable framework for detecting and classifying cyber threats.

At the core of the proposed system is the utilization of machine learning algorithms for automated feature extraction, model training, and attack classification. Supervised learning algorithms such as Random Forest, Extra tree classifier, and Ada boost will be employed to train classification models using labeled datasets of historical attack data. These models will learn to differentiate between various attack methods based on input features extracted from network traffic, system logs, and other relevant sources.

In addition to supervised learning, the proposed system will explore the use of unsupervised learning techniques such as clustering and anomaly detection to identify novel attack patterns and outliers in the data. Unsupervised learning algorithms will help detect previously unseen attack vectors and adapt the classification models to evolving threats without the need for labeled training data.

METHODOLOGY

To develop a Network Intrusion Detection System (NIDS), a systematic approach is essential. It begins with defining the objectives and scope of the NIDS, identifying the types of threats it aims to detect, and the network environment it will monitor. Data collection follows, involving the gathering of relevant network traffic data. This data undergoes preprocessing to handle missing values and format it appropriately for analysis. Then identifies key indicators of malicious activity, followed by the selection and training of suitable machine learning or statistical models for intrusion detection. Evaluation of these models' performance is crucial, utilizing metrics such as accuracy and precision. Upon deployment, the NIDS is integrated into the production network for real-time monitoring. Continuous maintenance and improvement are necessary, involving regular updates to address new threats and refine detection capabilities based on evolving network behavior and attack patterns. Through this methodology, organizations can develop robust NIDS tailored to their network security needs.

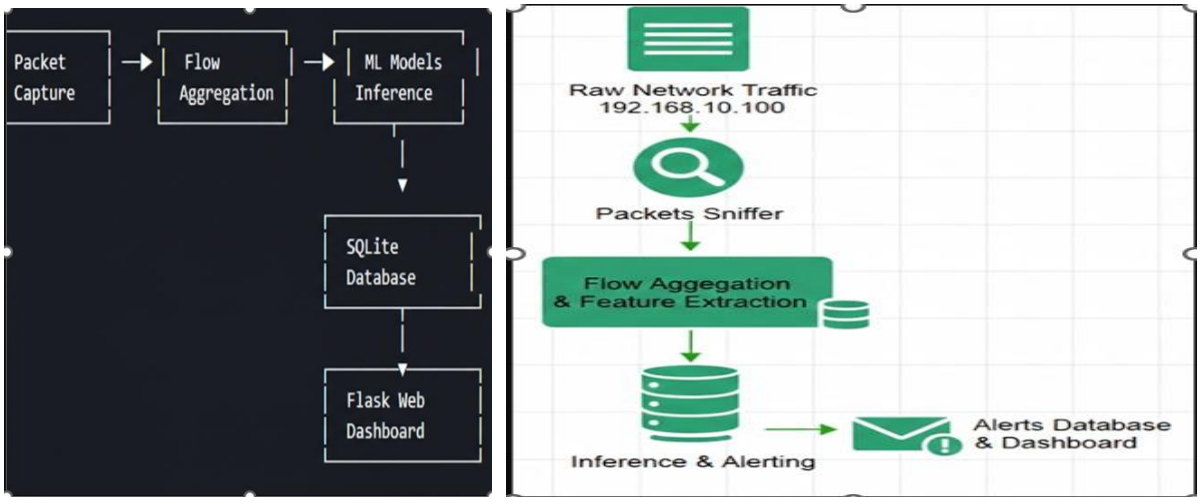

Fig 1 Architecture of the proposed system

Data Pre-processing:

Getting ready the data for algorithmic processing is crucial, especially concerning Intrusion Detection Systems (IDS). The primary aim of data preparation is to ensure that the dataset is clear and accurate, enabling the effective operation of IDS algorithms. This phase includes feature selection and normalization to optimize the performance of the IDS.

Datasets often contain symbolic attributes like flags and protocol types, which possess categorical values. To enhance algorithm performance, these nominal values need to be transformed into numerical representations. In scenarios involving

multiclass classification, which distinguishes between different attack types and normal traffic, and binary classification, which categorizes instances as either normal or attack, the dataset needs to be transformed using discretization.

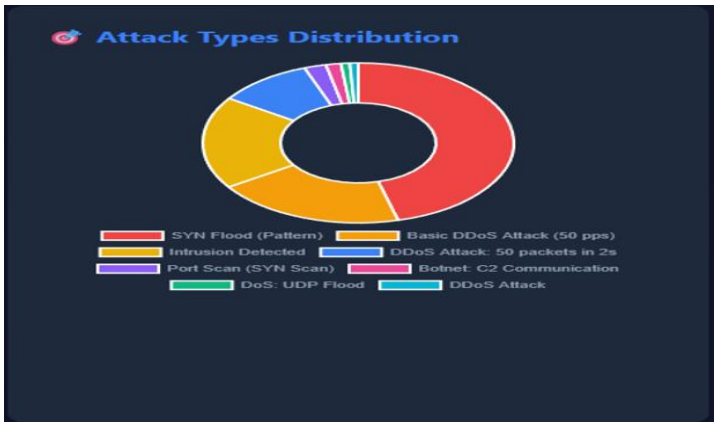

Fig 2. Data distribution

Discretization is essential for improving classification accuracy while conserving memory space. Before classification, it's important to assess the effectiveness of various classifiers using discretized data and evaluate the impact of discretization.

Numeric attributes undergo discretization using unsupervised techniques such as 10-bin discretization filters, implemented. Additionally, splitting the dataset into training and testing subsets is an essential stage in model construction.

In this research, the dataset is partitioned into two sections: 80% allocated for training, and the remaining 20% for testing. Attack labels are renamed to facilitate binary and multi-class classification, categorizing instances as normal traffic or attacks. Furthermore, the main attack categories, including Denial of Service (DoS), Probe, Remote to Local (R2L), and User to Root (U2R) intrusion attempts, are identified to refine the classification process.

Table 1: Listing the dataset labels.

Label (category)	Number of occurrences
Regular	67343
Intrusion	58675

Table 2: The dataset labels intended for multiclass classification.

Category	Number of occurrences
Regular	67343
DOS	45927
PROBE	11656
U2R	995
R2L	52

Fig. 2. Pie chart for the distribution of attack types

Training:

The training dataset, generated from the preprocessing phase, is used to train the machine learning algorithm. During this phase, a model is developed, and the training procedure is completed. The algorithms being assessed comprise the Extra Tree Classifier, AdaBoost, and Random Forests.

Decision tree classifier:

The Extra Trees Classifier, a form of collaborative learning method is initialized and trained using the training data. Once the model has been trained, the classifier makes predictions on the test set, and the model's accuracy is assessed utilizing the accuracy score. This process ensures robustness by training and testing the model on separate datasets, thereby providing insights into its generalization performance.

The presented methodology offers a systematic and reproducible framework for developing and accessing machine learning models for classification tasks, which could be particularly relevant in the context of Network Intrusion Detection Systems (NIDS).

These techniques provide a foundation for research aimed at enhancing the detection and mitigation of security threats in network environments, with potential implications for improving cybersecurity measures.

XG boost algorithm:

The AdaBoost algorithm is implemented for model training, leveraging its ability to iteratively improve classification performance by emphasizing misclassified instances. The trained model's accuracy is evaluated, providing insights into its effectiveness in classifying network intrusion attacks based on various features.

Furthermore, detailed analyses are conducted, including producing a classification report and constructing a confusion matrix to evaluate the model's performance across various attack categories. Visualization techniques, such as plotting the confusion matrix and comparing anticipated versus actual attack types, offer intuitive insights into the model's estimated behavior and potential areas needing enhancement.

Overall, the study offers a comprehensive framework for leveraging machine learning techniques, specifically XGBoost, in the domain of NIDS. The methodology and findings contribute to the ongoing efforts aimed at enhancing network security by effectively detecting and mitigating malicious activities in network traffic.

Random Forest Algorithm:

Initially, the dataset is loaded and pre-processed to handle missing values and categorical variables using Label Encoder. Subsequently, the dataset is partitioned into training and testing subsets, ensuring stratified representing class labels to maintain the dataset's balance.

The Random Forest algorithm is then utilized for model training, leveraging its ensemble learning technique to build a robust classifier. The accuracy of the trained model is assessed using diverse metrics, such as accuracy score, classification report, and confusion matrix, providing insights into its performance across different attack methods.

Furthermore, visualizations such as confusion matrices and plots comparing predicted versus actual attack methods offer intuitive insights into the performance of the model and potential avenues for enhancement.

Overall, the study offers a systematic approach to leveraging machine learning techniques, specifically Random Forest, for enhancing network security through effective detection and classification of network intrusion attacks. The methodology and findings contribute to advancing the understanding and development of robust NIDS systems, with potential implications for bolstering cybersecurity measures in today's interconnected digital landscape.

Testing:

The testing phase is crucial for assessing the performance of the machine learning algorithms trained during the training phase. In this phase, the trained models are subjected to unseen data to assess their generalization ability and predictive accuracy.

The testing process involves feeding the testing dataset, which was set aside during the data preprocessing phase, into the trained models. The models then make predictions on this unseen data, classifying instances into appropriate categories based on their learned patterns.

Metrics like accuracy, precision, recall, and F1 score are calculated to gauge the performance of the models. These metrics offer understanding into the models' capability to accurately classify instances, detect false positives and false negatives, and strike a balance between precision and recall.

Deployment:

The deployment phase involves the integration of the trained machine learning models into operational environments, where they can actively contribute to real-time intrusion detection and response activities.

During deployment, the trained models are embedded within existing cybersecurity infrastructures, such as Intrusion Detection Systems (IDS) platforms. This integration allows the models to consistently observe network activities, system records, and other pertinent data outlets for indications of suspicious activity or potential security breaches.

The deployment process also includes configuring the models to interact seamlessly with other components of the cybersecurity infrastructure, such as data collection agents, alerting systems, and incident response workflows. Integration with these systems ensures that alerts generated by the models are promptly acted upon by cybersecurity personnel, enabling rapid incident response and mitigation.

EXPERIMENTAL RESULTS

Table 3: Accuracies measured for the three ML classification algorithms

Classifier	Accuracy	Precision	Confidence	F1 score
Decision tree	99.912	100	100	100
XG boost	82.87	84	83	83
Random forest	99.920	100	100	100

The next phase involves implementing and training the model based on the conclusions drawn from earlier stages. Subsequently, the model will undergo validation to ensure it meets the specified criteria and assess its accuracy in predicting outcomes with new data. Through these evaluations, any flaws or limitations in the model are identified, enabling necessary actions to mitigate them.

Table 3 illustrates assessing the effectiveness of four classifiers across different evaluation criteria, such as accuracy, precision, recall, and F1 score. The Random Forest classifier emerges as the top performer, closely trailed by the Extra Tree Classifier. Both models attain an impressive accuracy level of 99.92%.

CONCLUSION AND FUTURE WORK

The performance evaluation of NIDS utilizing machine learning techniques underscores their effectiveness in enhancing network security and mitigating cyber threats. While ML-based NIDS exhibit superior detection capabilities and adaptability, addressing challenges such as false positives and computational efficiency remains paramount.

Future research directions include exploring advanced ML algorithms, leveraging anomaly detection techniques, and integrating threat intelligence for proactive intrusion prevention.