Secure Intrusion Detection System using Advanced Technology

Prof.M.D.Ingle
Department of Computer Engineering
JSPM's JSCOE
ingle.madhav@gmail.com

Swapnil Bhima Bansode
Department of Computer Engineering
JSPM's JSCOE
swapnilbansode401@gmail.com

Yogesh Sandip Bhongale
Department of Computer Engineering
JSPM's JSCOE
yogeshbhongale45091@gmail.com

Rutwik Sanjay Ingawale
Department of Computer Engineering
JSPM's JSCOE
rutwika.872000@gmail.com

Abstract - **The increasing sophistication of cyber threats poses significant challenges to maintaining secure network infrastructures. Traditional Intrusion Detection Systems (IDS) often struggle to balance detection accuracy, computational efficiency, and adaptability to emerging threats. This document presents an advanced IDS framework leveraging ensemble learning techniques, including Bagging (Random Forest, Extra Trees), Boosting, and Voting Classifiers, to enhance intrusion detection capabilities. The proposed system is designed to identify and classify various types of network intrusions, such as Denial-of-Service (DoS) attacks and malware, with improved accuracy and robustness. The integration of multiple machine learning algorithms (Logistic Regression, Support Vector Machine (SVM), Decision Tree) ensures that the model effectively captures diverse attack patterns while mitigating overfitting and variance. Experimental evaluations on the dataset demonstrate the effectiveness of the system, achieving high detection accuracy and outperforming individual machine learning models. The system's scalability makes it suitable for dynamic and large-scale network environments, providing a reliable solution for safeguarding critical digital assets.**

Keywords **– NIDS, Data analysis, methodology, Cybersecurity, Logistic Regression, Support Vector Machine (SVM), Decision Tree, Network Security, Anomaly Detection.**

I. INTRODUCTION

Due to the rapid rise in cyberattacks aimed at critical infrastructure, the demand for sophisticated and effective Intrusion Detection Systems (IDS) is at an all-time high. Conventional IDS models, usually depending on signature-based or anomaly-based detection techniques, frequently fall short in identifying complex and advancing threats. These traditional systems encounter challenges regarding scalability, and data privacy, particularly in decentralized or cloud-based settings.

This study presents an IDS framework leveraging machine learning techniques to enhance intrusion detection accuracy. Using the NSL-KDD dataset, we apply feature selection methods such as Variance Inflation Factor (VIF) and SelectKBest to identify the most relevant network traffic features. Machine learning classifiers, including Logistic Regression, are trained on this optimized feature set to improve detection performance. Additionally, statistical techniques are used to analyze network traffic patterns, aiding in the differentiation of normal and malicious activities.

The main contributions of this study are outlined as:

- Feature Engineering: Applying statistical techniques for optimal feature selection.
- Machine Learning Models: Utilizing supervised learning methods to classify network traffic.
- Performance Evaluation: Analyzing detection accuracy and effectiveness using evaluation metrics.

Overview: This introduces a new Intrusion Detection System (IDS) utilizing multinomial classification techniques to improve network security. By analyzing network traffic features and leveraging statistical modeling, the suggested IDS identifies intricate patterns across diverse intrusion types, including DOS, PROBE, R2L, and U2R attacks, enhancing the detection of advanced threats. The solution was evaluated using a real-world intrusion detection dataset, demonstrating excellent detection accuracy and showing significant advancements over conventional approaches, establishing it as a strong method for contemporary cybersecurity challenges.

II. LITERATURE SURVEY

In recent years, various techniques have emerged to enhance the accuracy and privacy of Network Intrusion Detection Systems (NIDS). A 2024 study introduced a Graph Neural Network (GNN) model equipped with privacy-enhancing mechanisms specifically for detecting network intrusions. This model effectively balances the need for privacy protection with the necessity for detection accuracy. However, its implementation can be computationally intensive and complex, largely due to its dependence on high-quality graph data [1].

Another approach focuses on power systems, employing a Random Forest algorithm alongside clustering techniques to bolster detection accuracy. While this method shows promise, it necessitates careful management of data imbalance and presents challenges in deployment complexity [2].

For securing Internet of Things (IoT) environments, a model that integrates Conditional Random Fields (CRF), Spider Monkey Optimization (SMO), and Convolutional Neural Networks (CNN) has been developed. This model enhances detection accuracy and allows for advanced feature selection. However, it requires a high degree of machine-learning expertise and involves significant computational resources [3].

Additionally, Deep Convolutional GANs (DCGANs) have been utilized to improve detection accuracy, particularly in scenarios involving imbalanced data. Despite their advantages, the performance of these GANs is highly sensitive to dataset quality biased data can lead to unreliable results [4]. An improved GAN model enhances multi-class classification capabilities, increasing the number of labeled samples and overall detection rates.

Nevertheless, challenges with parameter tuning and the generalization to new attack patterns persist [5].

Furthermore, a Long Short-Term Memory (LSTM) model optimized with Quantum Particle Swarm Optimization (QPSO) has demonstrated high accuracy when processing time-series data, making it well-suited for NIDS applications. However, its complex implementation and stringent data quality requirements can limit its accessibility [6]. In a novel approach, integrating Blockchain technology with WLAN Intrusion Detection Systems (IDS) has enhanced real-time monitoring capabilities and ensured secure, tamper-proof records. Nonetheless, this method can be computationally demanding and may suffer from high false-positive rates [7].

A real-time intrusion detection system based on deep learning architectures, such as AE-AlexJNet, has shown adaptive learning capabilities that are particularly beneficial for handling big data. However, this system may encounter issues related to extended training times and challenges with portability [8]. Traditional machine learning methods that combine Support Vector Machines (SVM) and Naïve Bayes classifiers are also effective in processing large datasets. These approaches are characterized by their high accuracy but require significant computational resources and careful management to mitigate the risk of false positives [9].

In the context of managing imbalanced network traffic, CNN models enhanced with SMOTE (Synthetic Minority Over-sampling Technique) have been shown to improve dataset quality. However, there is a risk of inadvertently removing useful samples during the preprocessing phase [10]. A comprehensive model that employs Convolutional, Recurrent, and LSTM networks provides enhanced feature extraction capabilities and outperforms traditional methods. Yet, it demands considerable computational power and may struggle to adapt to evolving datasets [11].

Data mining techniques, including C4.5 Decision Trees and k-means clustering, have proven effective in detecting both known and unknown threats. However, these methods require security expertise for rule creation and can be challenging to maintain [12]. In specialized environments like Tor networks, Neural Networks (NN) utilizing back-propagation have achieved high accuracy in recognizing patterns of malicious activity. Nonetheless, the complexity of data analysis and the need for large datasets present substantial challenges [13].

Finally, an adaptive NIDS that combines machine learning algorithms with network protection tools such as iptables facilitates real-time intrusion prevention. This approach is contingent upon careful feature selection and the availability of extensive labeled data to ensure success [14]. Additionally, a neural network-based model for intrusion detection enhances accuracy in identifying complex attack patterns, although it remains vulnerable to overfitting if the training data is insufficient or lacks diversity [15].

III METHODOLOGY

The suggested Intrusion Detection System (IDS) integrates privacy-protecting technologies to provide a strong solution for network security. The architecture of the system is designed to identify intrusions while maintaining the privacy of sensitive information. To protect data privacy, particularly in situations requiring joint detection across various nodes, the following methods were employed: Logistic Regression, a baseline classifier for binary and multi-class classification tasks, is integrated into a Voting Classifier ensemble to enhance performance; Decision Tree, a versatile model used to identify intrusions by learning decision rules, is optimized through hyperparameter tuning and ensemble methods to improve accuracy and reduce overfitting; and Support Vector Machine (SVM), a powerful classifier for distinguishing normal traffic from attack types, utilizes the radial basis function (RBF) kernel. Additionally, the Voting model combines predictions from multiple classification algorithms to improve accuracy and robustness.

A. ARCHITECTURE SECURE INTRUSION DETECTION SYSTEM USING ADVANCED TECHNOLOGY

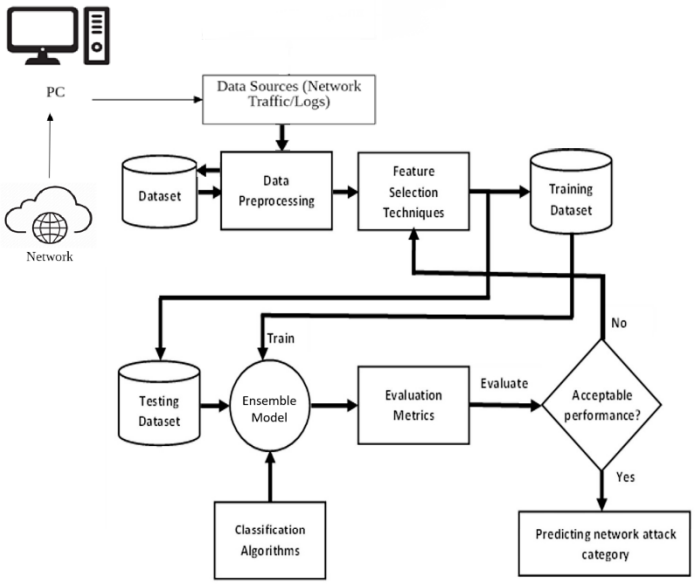

Fig. 1: Architecture diagram

1. **Data Sources (Network Traffic/Logs)**
   - This represents the raw data collected from the network, including network traffic logs or packet data. It serves as the input for the intrusion detection system.

2. **Dataset**
   - The raw data collected from the network is stored in a structured format, such as a dataset, for further processing.

3. **Data Preprocessing**
   - This block involves cleaning and transforming the raw dataset. It may include handling missing values, removing irrelevant features, encoding categorical data, and normalizing the data to prepare it for feature extraction and model training.

4. **Feature Selection Techniques**
   - This step selects the most relevant features from the dataset, reducing dimensionality and improving the model's efficiency and accuracy. Techniques like correlation analysis, mutual information, or feature importance rankings can be used.

5. **Training Dataset**
   - A portion of the pre-processed data is allocated for training the model. This dataset is used to teach the ensemble model the patterns and characteristics of different types of intrusions.

6. **Testing Dataset**
   - Another portion of the pre-processed data is reserved for testing the trained model to evaluate its performance. This dataset contains unseen data to validate the model's

generalization capabilities.

7.**Classification Algorithms**

- These algorithms form the core of the ensemble model. Examples include Decision Trees, Random Forests, and Gradient Boosting, which are used to classify network traffic into normal behavior or different attack categories.

8.**Ensemble Model**

- The ensemble model combines predictions from multiple classification algorithms to improve accuracy and robustness. Logistic Regression, Support Vector Machine (SVM) and Decision Tree are typically used for this purpose.

9.**Evaluation Metrics**

- Metrics such as accuracy, precision, recall, F1-score, and confusion matrix are used to assess the model's performance. These metrics help identify whether the model is effective in detecting intrusions.

10.**Acceptable Performance?**

- This decision block evaluates whether the model meets the desired performance thresholds. If the performance is unacceptable, the model is retrained with improved preprocessing, feature selection, or hyperparameter tuning.

11.**Predicting Network Attack Category**

- Once the model achieves acceptable performance, it is deployed to predict network attack categories in real-time, enabling the detection of intrusions and quick response to threats.

B. MATHEMATICAL MODEL

The main used algorithms in models are Logistic Regression, Support Vector Machine (SVM), and Decision Tree as applied in a Secure Intrusion Detection System (IDS):

1. Logistic Regression

Hypothesis:

$$h_\theta(x) = 1 / (1+e^{-\theta^{\wedge}(T)x1})$$

2. Support Vector Machine (SVM)

Decision Function:

$$f(x) = w^T x + b$$

3. Decision Tree

Splitting Criterion (Information Gain):

$$IG = H(parent) - \sum_k [(n_k / n) H(k)]$$

4.Voting Model

Voting Classifier:

Combines predictions from M models hm(x) using majority voting:

$$H(x) = \text{mode}(h1(x), h2(x), \ldots, h_M(x))$$

or for soft voting:

$$H(x) = \arg\max_k \; M\sum_{m=1} P_m(y=k|x)$$

where Pm(y = k | x) is the probability of class k from model m.

C. ALGORITHM

1) Logistic Regression:

Logistic regression is a widely used method for binary and multi-class classification. In this study, the lbfgs solver was employed to handle multi-class tasks efficiently. The model predicts intrusion types by learning the relationship between features and class likelihoods. It was integrated into a Voting Classifier ensemble to enhance performance. Its simplicity and interpretability make it a valuable baseline for intrusion detection.

2) Decision Tree for Intrusion Detection:

The Decision Tree classifier is an interpretable model that learns decision rules based on feature data. In this study, Decision Trees were configured with hyperparameter tuning using Grid Search to optimize performance. Their ability to capture nonlinear relationships and intuitive structure makes them effective for detecting complex attacks. They were also incorporated into ensemble methods like bagging and voting classifiers to improve accuracy and reduce overfitting.

3) Support Vector Machine (SVM):

SVM is a powerful classifier used for detecting intrusions by identifying optimal decision boundaries. In this study, the RBF kernel was utilized to capture complex patterns, with hyperparameters (C and gamma) tuned via Grid Search with cross-validation for optimal performance. Integrated into a Voting Classifier, SVM enhances overall detection accuracy. Despite higher computational costs, its ability to handle high-dimensional data makes it a valuable addition to the intrusion detection system.

IV. RESULTS AND ANALYSIS

Result :

1) Fill the form as shown below.

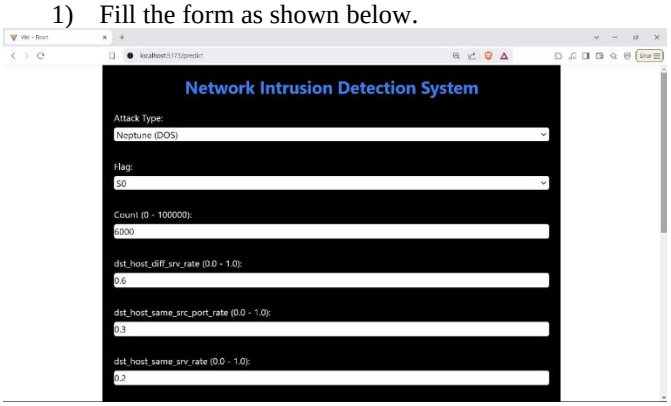

Fig. 2: Insert Input (Image 1)

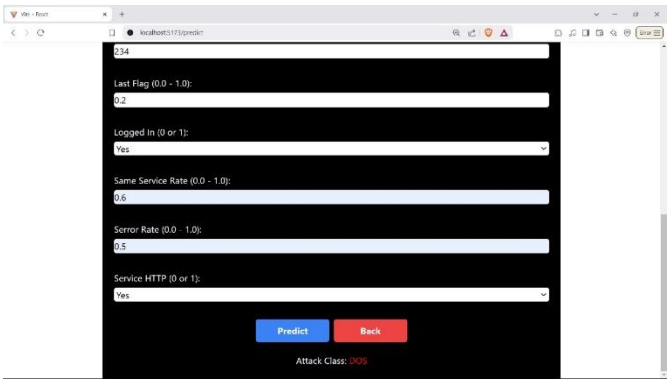

Fig. 3:. Insert Input (Image 2)

2) OUTPUT

Then Click on Predict and you get the predicted attack class.

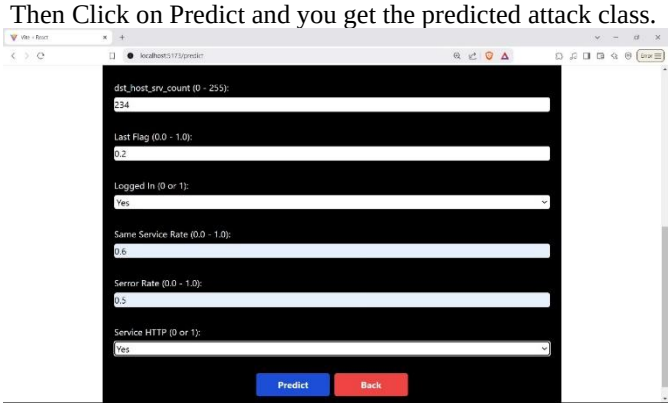

Fig. 4: Output

3.Results:
Shows the history of all the recorded results.

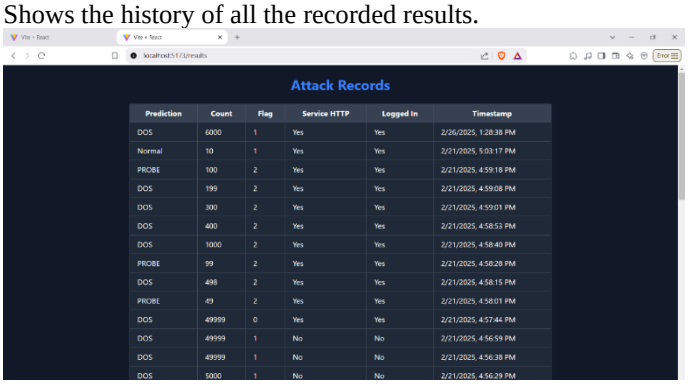

Fig. 5: Results

V. CONCLUSION

In conclusion, the proposed Secure Intrusion Detection System (NIDS), which integrates advanced ensemble learning techniques with privacy-preserving methods such as differential privacy, federated learning, and encryption, offers an innovative solution to contemporary network security challenges. By leveraging models like Random Forest and Gradient Boosting, the system accurately detects a wide range of cyber threats, including malware and DoS attacks, while ensuring sensitive data remains protected throughout the detection process. Experimental results, particularly with the dataset, demonstrate superior performance compared to traditional machine learning models in terms of detection accuracy, and scalability. This IDS framework not only enhances cybersecurity but also adheres to privacy regulations, providing a proactive, resilient, and adaptable solution for modern network infrastructures.

VI. ACKNOWLEDGMENT

We would like to express our special thanks of gratitude to our guide Prof. M. D. Ingle for their able guidance and support. We would also like to extend our gratitude to the Principal Mr. R. D. Kanphade for providing us with all the facilities that were required. We would also like to acknowledge with much appreciation the role of our staff. We sincerely thank our parents and our friends who have always helped and encouraged us.

VI. REFERENCES

[1] X. Pei, X. Deng, S. Tian, P. Jiang, Y. Zhao, and K. Xue, "A Privacy-Preserving Graph Neural Network for Intrusion Detection," IEEE Transactions on Dependable and Secure Computing, DOI: 10.1109/TDSC.2024.3417853.
[2] G. Zhu, H. Yuan, Y. Zhuang, Y. Guo, X. Zhang, and S. Qiu, "Network Intrusion Detection Method for Power Systems Using Random Forest Algorithm," in Proceedings of the 2021 13th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Wuhan, China.
[3] G. Parimala and R. Kayalvizhi, "An Effective Intrusion Detection System for Securing IoT Through Feature Selection and Deep Learning," in Proceedings of the 2021 International Conference on Computer Communication and Informatics (ICCCI-2021), Jan. 27–29, 2021, Coimbatore, India, SRM Institute of Science and Technology.
[4] W. Chao, W. Wenhui, J. Dong, and G. Guo, "Investigation on Network Intrusion Detection Technology Using DCGAN," in 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC).
[5] D. Zhang, K. Hao, L. Li, et al., "Network Intrusion Detection Utilizing GAN Model," in 2020 International Conference on Computer Communication and Network Security (CCNS).
[6] L. Zhang, H. Yan, and Q. Zhu, "Enhanced LSTM Network Method for Intrusion Detection," in 2020 IEEE 6th International Conference on Computer and Communications (ICCC), Chongqing, China.
[7] Z. Xin and Y. Huabing, "Study on Blockchain Network Intrusion Detection System," in 2019 International Conference on Computer Network, Electronic and Automation (ICCNEA), Xi'an, China.
[8] Y. Dong, "Real-Time Network Intrusion Detection System Utilizing Deep Learning," Huazhong University of Science & Technology, Wuhan, China.
[9] X. Zhang, J. Ran, and J. Mi, "A Convolutional Neural Network-Based Intrusion Detection System for Imbalanced Network Traffic," in 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT), Oct. 2019, Dalian, China. Department of Computer Engineering, JSCOE Hadapsar
[10] Secure Intrusion Detection System using Advanced Technology 48 T. Ishitaki, D. Elmazi, Y. Liu, T. Oda, L. Barolli, and K. Uchida, "Utilizing Neural Networks for Intrusion Detection in Tor Networks," in 2015 29th International Conference on Advanced Information Networking and Applications Workshops (WAINA), Mar. 2017.
[11] Y. Zhao, "Network Intrusion Detection System Model Utilizing Data Mining," School of Computer Engineering, Weifang University, Weifang, China.
[12] V. R. Vinayakumar and K. P. Soman, "Implementing Convolutional Neural Networks for Network Intrusion Detection," Centre for Computational Engineering and Networking (CEN), Amrita School of Engineering, Coimbatore, India.
[13] N. Wattanapongsakorn, S. Srakaew, E. Wonghirunsombat, C. Sribavon mongkol, T. Junhom, P. Jongsubsook, and C. Charnsripinyo, "A Practical Network Based Intrusion Detection and Prevention System," in 2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications, Jun. 2012, Bangkok, Thailand.
[14] J. Zhao, M. Chen, and Q. Luo, "Study on Neural Network-Based Intrusion Detection System," Department of Computer Science and Technology, Hunan Institute of Technology, Hunan, China.
[15] J. Zhao, M. Chen, and Q. Luo, "Study on Neural Network-Based Intrusion Detection System," Department of Computer Science and Technology, Hunan Institute of Technology, Hunan, China.