iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** V **Month of publication:** May 2024

DOI: https://doi.org/10.22214/ijraset.2024.61565

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

Intrusion Detection System Using Probabilistic Adaptive Learning

Ms. Siva Priya[1], Udaya Kiran[2], Nanthine Sri[3], Bhavani[4]
Assistant Professor, CSE-SRMIST RMP Chennai India
CSE-CS, SRMIST RMP Chennai India

Abstract: ***Machine learning and deep learning techniques are widely used to evaluate intrusion detection systems (IDS) capable of rapidly and automatically recognizing and classifying cyber-attacks on networks and hosts. However, when destructive attacks are becoming more extensive, more challenges develop, needing a comprehensive response. Numerous intrusion detection datasets are publicly accessible for further analysis by the cybersecurity research community. The Intrusion Detection System (IDS) is an effective tool utilized in cybersecurity systems to detect and identify intrusion attacks. With the increasing volume of data generation, the possibility of various forms of intrusion attacks also increases. Feature selection is crucial and often necessary enhance performance. The structure of the dataset can impact the efficiency of the machine learning model. Furthermore, data imbalance can pose a problem, but sampling approaches can help mitigate it. With technological advancements, machine learning-based methods have emerged as the cornerstone of modern intrusion detection, enabling more precise identification of abnormal behaviors and potential intrusions by learning the patterns of normal network traffic. In response to these challenges, this paper introduces an innovative intrusion detection model that amalgamates the Probabilistic Adaptiv eLearning Network architecture.***

I. INTRODUCTION

Intrusion detection systems (IDS) stand as stalwarts in safeguarding network security, evolving from traditional methods to harness the power of machine learning (ML). Initially, reliance on encryption-decryption, protocol control, and firewalls proved inadequate against a diverse array of cyber threats. The rise of sophisticated attacks, notably Denial-of-Service (DoS) assaults, underscored the limitations of these approaches, often leading to high false positive rates. Enter ML, offering a paradigm shift in IDS design. ML algorithms, such as Support Vector Machines (SVMs), Knearest neighbor (KNN), and deep learning architectures, promise enhanced detection rates while mitigating false positives. This shift is propelled by ML's ability to discern patterns in vast amounts of network data, adapting to emerging threats and minimizing false alarms.

Key to ML-based IDS success is the availability of robust datasets encompassing normal and malicious network traffic. Feature selection and extraction techniques play a pivotal role in distilling pertinent information from network data, enabling ML algorithms to effectively differentiate between benign and intrusive activities. Anomaly-based and misuse-based detection methods represent two pillars of IDS strategy. Anomaly-based approaches compare network traffic against baseline norms, identifying deviations indicative of potential threats. Conversely, misuse-based methods rely on predefined signatures of known attacks to flag malicious activity. Each method has its strengths and weaknesses, with anomaly-based detection excelling at identifying unknown threats but often suffering from high false positive rates.

II. LITERATURE SURVEY

Overview of Network Intrusion Detection Systems (NIDS): Explanation of NIDS and its importance in cybersecurity. Overview of traditional methods used in NIDS, such as rule-based systems and statistical approaches.

A. Challenges with Traditional Approaches

Limitations of rule-based and statistical methods in detecting complex and evolving threats. Difficulty in extracting meaningful features from raw network data.

B. Introduction to Deep Learning

Explanation of deep learning concepts, including neural networks, convolutional neural networks (CNNs), recurrent neural networks (RNNs), and autoencoders.

Advantages of deep learning in handling complex data and extracting intricate patterns.

C. Deep Learning for Network Intrusion Detection

Review of recent studies and projects that have applied deep learning techniques to NIDS.
Discussion on the effectiveness of deep learning models in detecting various types of network intrusions.

D. Feature Extraction Methods

Overview of feature extraction techniques used in NIDS, such as statistical features, frequency-based features, and protocol-specific features.

E. Temporal and Spatial Features in NIDS

Explanation of temporal features related to time-based patterns in network traffic.
Discussion on spatial features concerning spatial relationships between network entities.
Examples of how temporal and spatial features can enhance intrusion detection capabilities.

F. State-of-the-Art Techniques

Review of cutting-edge deep learning models and algorithms used in NIDS, such as Deep Convolutional Generative Adversarial Networks (DCGANs) and Ensemble of Extreme Sparse Neural Networks (EESNNs).
Analysis of the strengths and weaknesses of these techniques in the context of network intrusion detection.

G. Evaluation Metrics

Overview of common evaluation metrics used to assess the performance of NIDS, including accuracy, precision, recall, F1-score, and area under the curve (AUC).
Discussion on the importance of choosing appropriate evaluation metrics based on the characteristics of the dataset and the goals of the detection system.

H. Conclusion and Future Directions

Summary of key findings from the literature survey. Identification of gaps in existing research and opportunities for future enhancements in NIDS using deep learning techniques.

III. PROPOSED WORK

The project proposes Using Probabilistic adaptive learning for precise classification of different types of attacks. To avoid data imbalance while data preprocessing DSSTE and DCGAN techniques are used. Spatial and temporal features are used for the process of feature selection. Spatial feature -The DenseNet 169 model is used to extract input data and helps in extracting characteristics through multi-level learning using deep learning properties. Temporal feature -SAT net is utilized to extract features automatically through repetitive multi-level learning. EESNN is employed for the classification of the attack categories in the Network Intrusion detection System. The raw data is initially processed using hybrid sampling to achieve balance.

Then, it goes through data normalization and other preprocessing to tackle the issue of the flow of network data and the extensive nature of the characteristics.

A deep hierarchical network model is finally employed for categorization. The suggested network intrusion detection model's specifics are divided into four phases. The effectiveness of the categorization model is impacted by the unbalanced distribution of network traffic data.

As a result, this paper uses the Difficult Set Sampling Technique (DSSTE) algorithm to reduce the majority sample size while removing noise, followed by the Deep Convolutional Generative Adversarial Networks (DCGANs) algorithm to enhance the minority sample size.

These two approaches are combined to create a balanced data collection. Due to the ntricate structure of network transmission data's features, the spatial and temporal characteristics of the data are extracted using a deep hierarchical network model developed by DenseNet 169 and 10 SAT-Net to increase classification efficiency. Finally, the Enhanced Elman Spike Neural Network (EESNN) is employed to classify the attack categories.

IV. MODULES AND FEATURES

A. Module 1: Exploratory Data Analysis

Initial step in data analysis, identifies anomalies, patterns, and relationships. Ensures data cleanliness and prepares it for analysis. Exploratory Data Analysis is a data analytics process to understand the data in depth and to learn the different characteristics of data, often with visual means. This allows us to get a feel for our data better and find useful patterns in it. The whole aim is to understand the data; understanding the data can be a lot of things when we are exploring the data. Few things we have to keep in mind while exploring the data, we have to make sure that the data is clean and does not have redundancy or missing values, or even null values on the dataset.

B. Module 2: Feature Selection

Chooses relevant features, improves data quality, and enhances model performance. Uses filter-based methods to rank features based on importance. Feature selection methods are categorized into wrapper-based and filter-based methods. Filter method uses variable ranking techniques to rank the features where the highly ranked features are selected and applied to the learning algorithm. In this study, we applied filter method using information gain based selection algorithm to evaluate the feature ranks, checking which features are most important to build performance model

C. Module 3: Model Creation and Prediction

Long short-term memory is a sort of Recurrent Neural Network(RNN), that makes recalling past knowledge easier. Here, LSTM overcomes the RNNs vanishing gradient problem. The LSTM algorithm is highly suitable for discovering, analysing and forecasting time series with uncertain length. The model is trained via backpropagation. The Components of LSTM are discussed below.

1) *Input Gate:* With an input gate, it determines which value from the input should be utilised to change the memory. The sigmoid function specifies which numbers are allowed to pass between 0 and 1, while the tanh function gives the data input weight, indicating its relevance on a scale of -1 to 1.
2) *Forget Gate:* This Gate defines which block information should be deleted. The sigmoid function determines this. It creates a number between 0 and 1 for each number in the cell state. It looks at the previous state and the content input for each number in the cell state.
3) *Output Gate:* The input and memory of the block defines the output gate. The sigmoid function determines which values are permitted to pass through numbers between 0 and 1. The tanh function multiplies Sigmoids output by the weightage provided to input values, establishing their relevance level, which range from -1 to 1. LSTMs are subject to the same hyper-parameters as ANNs

D. Module 4: Performance Metrics

Various metrics are specified for evaluating the performance of the models based on the confusion matrix. A total of five metrics, including Accuracy Score, Precision, Recall, Error Rates, F1 Score, and Overall Accuracy, has been used to evaluate and compare classification performance. For every class in the models, the value of these metrics is estimated. The five-performance metrics were obtained as follows:

1) *Accuracy:* Accuracy can be used to judge a models classification ability, but it cannot reflect specific details. When the classification model makes predictions, the confusion matrix indicates the prediction details of each category by comparing the predicted result with the actual value. Accuracy is the ratio of the total number of positive observations to the total number of all observations.
2) *Recall:* Recall evaluates a models ability to correctly anticipate genuine instances, computed as the ratio of true positive predictions to the combined count of true positive and false negative predictions where T indicates true positive cases and F shows false negative.
3) Error Rate Error rate is simply one minus the accuracy. If the accuracy of a model is 90%, the error rate would be 10%.

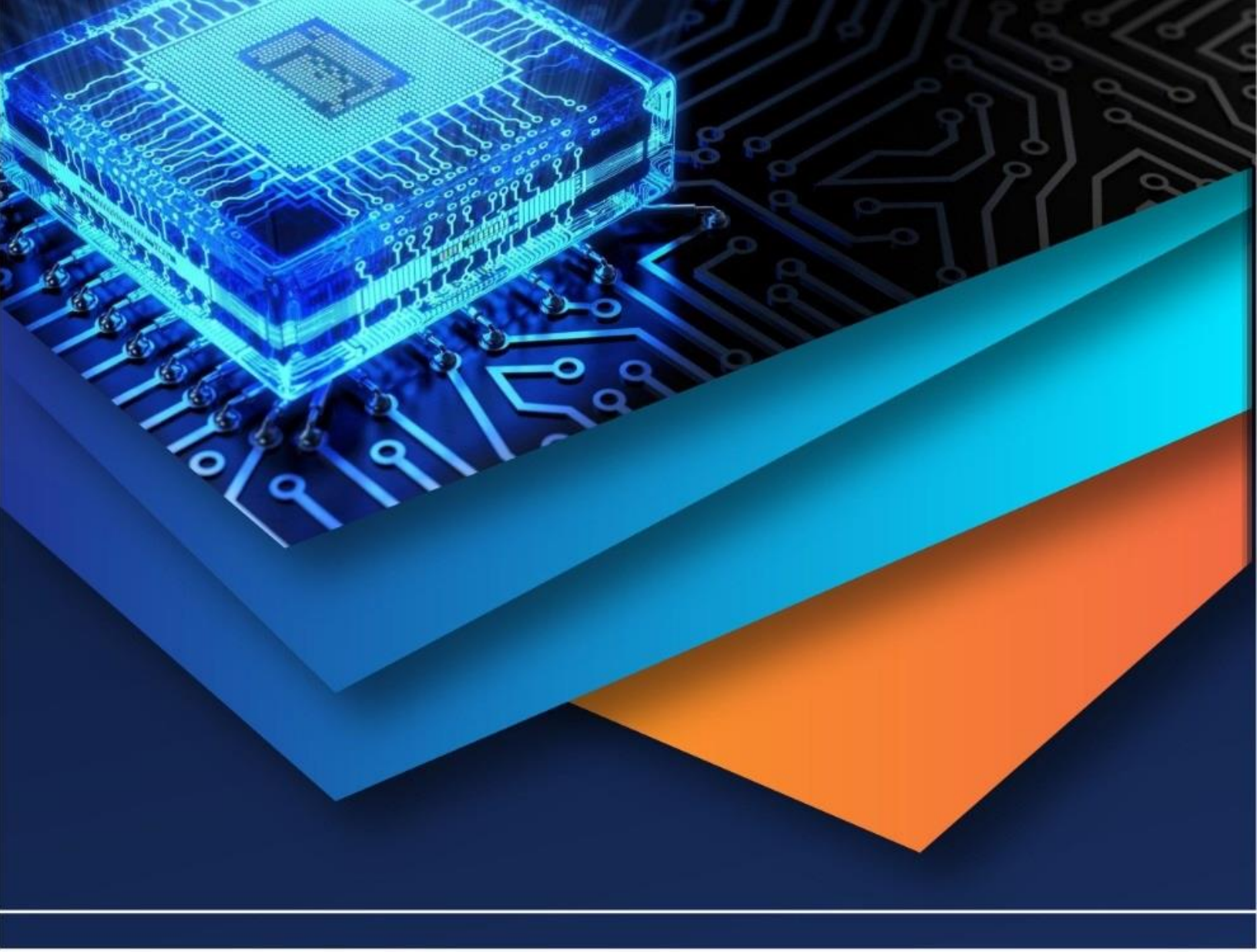
V. CONCLUSION

Explored various machine learning and deep learning models for network traffic classification. Preprocessed dataset by encoding and normalizing features. Trained classifiers like Logistic Regression, Random Forest, etc., and a deep learning Autoencoder. Evaluated models using accuracy, precision, recall, and F1-score metrics.

Further optimization could enhance performance for real-world applications. This paper showcases the impressive performance of machine learning (ML) models in detecting attacks, specifically in Anomaly Detection (binary classification) and Anomaly Classification (multi-class problems). It compares model performance using three widely-used datasets and outlines the entire operational flow of dataset management and manipulation, emphasizing techniques to optimize model performance. The paper evaluates classification metrics and computational complexity, highlighting that Decision Trees (DT) have the lowest computational burden. It extends its scope to evaluating ML models for embedded data traffic safety applications in mechatronic systems and explores the potential of deep learning models, emphasizing their effectiveness in various tasks. Integrating deep learning models into intrusion detection systems has the potential to improve accuracy, handle complex data, reduce false alarms, enhance detection rates, and enable real-time monitoring for anomaly detection, particularly crucial for high-traffic networks. The main objective of this review article is to examine the state of the art of ML in computer network security, with future developments expected to increasingly focus on applying deep learning (DL).

REFERENCES

[1] T. Ma, F. Wang, J. Cheng, Y. Yu and X. Chen, A hybrid spectral clustering and deep neural network ensemble algorithm for intrusion detection in sensor networks, Sensors, vol. 16, no. 10, pp. 1701, 2016.
[2] S. Leyi, Z. Hongqiang, L. Yihao and L. Jia, Intrusion detection of industrial control system based on correlation information entropy and CNN-BiLSTM, J. Comput. Res. Develop., vol. 56, no. 11, pp. 2330-2338, 2019.
[3] Y. Ding and Y. Zhai, Intrusion detection system for NSL-KDD dataset using convolutional neural networks, Proc. 2nd Int. Conf. Comput. Sci. Artif. Intell. (CSAI), pp. 81-85, 2018.
[4] C. Liu, Y. Liu, Y. Yan and J. Wang, An intrusion detection model with hierarchical attention mechanism, IEEE Access, vol. 8, pp. 67542-67554, 2020.
[5] D. E. Denning, An intrusion-detection model, IEEE Trans. Softw. Eng., vol. SE-13, no. 2, pp. 222-232, Feb. 1987.
[6] M. Akbanov, V. G. Vassilakis and M. D. Logothetis, Ransomware detection and mitigation using softwaredefined networking: The case of WannaCry, Comput. Electr. Eng., vol. 76, pp. 111-121, Jun. 2019.
[7] S. Pan, T. Morris and U. Adhikari, Developing a hybrid intrusion detection system using data mining for power systems, IEEE Trans. Smart Grid, vol. 6, no. 6, pp. 3104-3113, Nov. 2015.
[8] M. Aloqaily, S. Otoum, I. A. Ridhawi and Y. Jararweh, An intrusion detection system for connected vehicles in smart cities, Ad Hoc Netw., vol. 90, Jul. 2019
[9] A. Azab, M. Alazab and M. Aiash, Machine learning based botnet identification traffic, Proc. 15th IEEE Int. Conf. Trust Secur. Privacy Comput. Commun. (Trustcom), pp. 1788-1794, Aug. 2016. 31
[10] M. Tang, M. Alazab, Y. Luo and M. Donlon, Disclosure of cyber security vulnerabilities: time series modelling, Int. J. Electron. Secur. Digit. Forensics, vol. 10, no. 3, pp. 255-275, 2018.
[11] K. Zheng, Z. Cai, X. Zhang, Z. Wang and B. Yang, Algorithms to speedup pattern matching for network intrusion detection systems, Comput. Commun., vol. 62, pp. 47-58, May 2015.
[12] D. Papamartzivanos, F. G. Marmol and G. Kambourakis, Introducing deep learning self-adaptive misuse network intrusion detection systems, IEEE Access, vol. 7, pp. 13546-13560, 2019.
[13] S. M. Kasongo and Y. Sun, A deep learning method with filter based feature engineering for wireless intrusion detection system, IEEE Access, vol. 7, pp. 38597-38607, 2019.
[14] X. Yang and Z. Hui, Intrusion detection alarm filtering technology based on ant colony clustering algorithm, Proc. 6th Int. Conf. Intell. Syst. Design Eng. Appl. (ISDEA), pp. 470-473, Aug. 2015.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY