International Journal of Interdisciplinary Innovative Research & Development (IJIIRD)

ISSN: 2456-236X

Vol. 10 Issue 01 | 2025

A Hybrid Approach for Network Intrusion Detection

M.R. Rajput[1], A.S.Narkhede[2], P. B. Patthe[3], Prof. V. S. Chopade[4]

[1,2,3,4] *Assistant Professor, CSE, Padm.Dr. VBKCOE, Maharastra, India*

DOI: 10.5281/zenodo.15751522

ABSTRACT

Due to the widespread use of the internet and smart devices, var- ious attacks like intrusion, zero-day, Malware, and security breaches are a constant threat to any organization's network infrastructure. Thus, a Network Intrusion Detection System (NIDS) is required to detect attacks in network traffic. This paper proposes a new hybrid method for intrusion detection and attack categorization. The proposed approach comprises three steps to address high false and low false-negative rates for intrusion detection and attack categorization. In the first step, the dataset is preprocessed through the data transformation technique and min-max method. Secondly, the random forest recursive feature elimination method is applied to identify optimal features that positively impact the model's performance. Next, we use various Support Vector Machine (SVM) types to detect intrusion and the Adaptive Neuro- Fuzzy System (ANFIS) to categorize probe, U2R, R2U, and DDOS attacks. The validation of the proposed method is calculated through Fine Gaussian SVM (FGSVM), which is 99.3% for the binary class. Mean Square Error (MSE) is reported as 0.084964 for training data, 0.0855203 for testing, and 0.084964 to validate multiclass categorization.

Keyword : *- Network security; intrusion detection system; machine learning; attacks; data mining; classification; feature selection*

1.INTRODUCTION

Due to deep integration between the world and the internet, the network framework always experiences various kinds of attacks. Identification of these attacks is a technical issue and currently the area of concern these days. Intrusion violates fundamental privacy conditions, e.g., confidentiality, integrity, accessibility, denial of services [1–3]. The purpose of NIDS is to identify an intrusion on networks. They detect misuse of attempts either by a legal person or by third parties [4]. Break-in security vulnerabilities, misuse of the system are attacks that the IDS can identify [5–9]. Analysis of the transmitted packet in a network and through the collection of data, IDS worked [10–12]. IDS is a classification problem to detect the behavior of data, either it is

 normal or anomalous [13,14]. If network traffic is malicious, it will lie in any few of four attack categories: Denial of services, User-to-root, root-to-local, probing.

In the research related to the intruder's signature-based detection, activities that are not intrusive are also marked as anomalous behavior and generate false and false-positive results . Data set preprocessing is vital to achieving the maximum success rate through the classifier. Classifier results depend on input data.If input data is preprocessed, the results will be more accurate. However, it is a complex task and takes and more time. Pre-processing reduces data complexity, so it is easy to understand the nature of data, and data analysis will be performed more accurately. Several tools and methods are available to preprocess data, including sampling, normalization, discretization data transformation, and feature extraction. Problems that cater to data preprocessing are removing noise, replacing missing values, and inconsistency in data.

The utilization of various data mining techniques and efficient NIDS is usually developed because they improve prediction performance, reduce computation time, and better understand data. The number of input variables is reduced to achieve desirable results to reduce compu- tational modeling cost and increase performance. NSL KDD dataset has a large no of data that is needed to be filtered. Supervised and unsupervised are the two main feature selection techniques. The unsupervised feature selection technique works on finding the correlation between

input variables [18,19] and ignores the target variable. Supervised feature selection removes the irrelevant features with the help of the target variable.

Various machine learning methods are used to identify anomalies in networks, and, as a result, it helps the network administrator take the required precautions to avoid intrusion over its network. Machine learning conventional strategies are part of shallow learning and rely on inputs. Since the data requirements for classification methods differ from one another. The two categories of machine learning methods are supervised and unsupervised. Unsupervised machine learning methods have been proven to be the most powerful in an anomaly-based intrusion detection system. Machine learning systems are composed of several computing layers and learn various data representations with several abstraction levels that fall in deep learning classes.

SVM1 can be used for classification and works by finding a hyperplane in N-dimensional space that separates n classes. Hyperplanes are of any possible type chosen to separate two different class data points a finding a hyperplane with maximum margin results in better accuracy. Data points closer to the hyperplane influence its position and orientation and are named support vectors. In the NSL-KDD dataset, there are 42 features, so it is pretty complex to draw it. The classifier's margin is maximized with these support vectors' help, and they helped build an SVM.

SVM works on the output of a linear function. If the output of a function is 1, it will classify to one class; if it is −1, then it will classify to other classes.

Fuzzy logic works on the methodology of human decision-making. As the word fuzzy refers to vague things, so it deals with vague and imprecise information. Fuzzy logic2 is based on fuzziness instead of Boolean Logic that only results in true or false. The adaptive neuro-fuzzy inference system resembles artificial neural networks (ANN). For capturing the benefits of ANN and fuzzy in a single framework, it integrates principles of both techniques. Fuzzifier takes input and assigns a linguistic variable. If-then rules are defined in the rule base block

2. PROPOSED FRAMEWORK

This section describes the proposed approach for network intrusion detection. In this work, the hybrid approach SVM with ANFIS is applied for effective detection. Fig. 1 shows the proposed approach.

*2.*1 Dataset Selection

DARPA99 is the first data set created for the intrusion detection system at Lincoln laboratory in 1998. KDD99is the enhanced version of DARPA. Since the elimination of duplicate records in the KDD99 set of data has a substantial effect on the output of NSL-KDD [26] systems, the data set is considered a standard for the identification of Trespass in organizational structures. The usage of the NSL-KDD data set has the following benefits. Fig. 2 is a Visualization of some attributes of the NSL-KDD Data Set.

- Results are not biased because the train data set have no redundant records.
- Significant degradation rates owing to the lack of redundant information in the test sample.
- The chosen record is inversely proportional to each particular class category in the original KDD data collection.

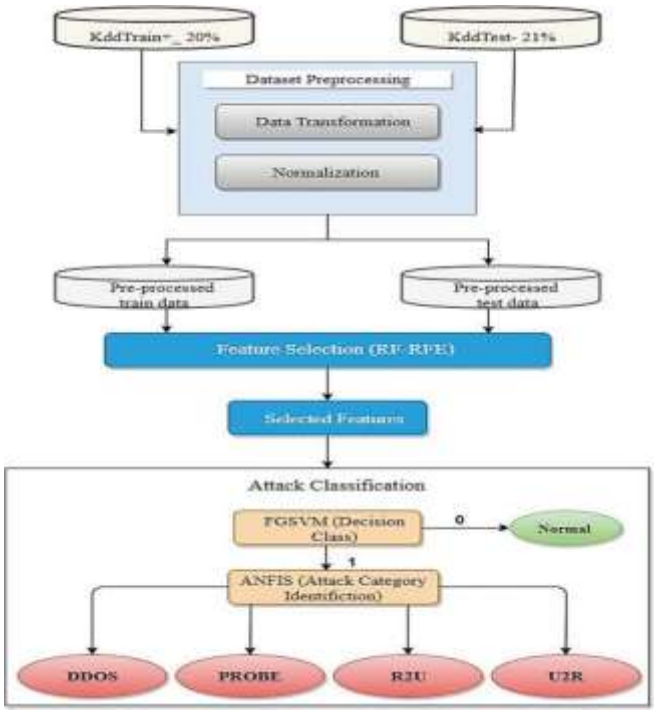

Fig -Proposed approach for intrusion detection

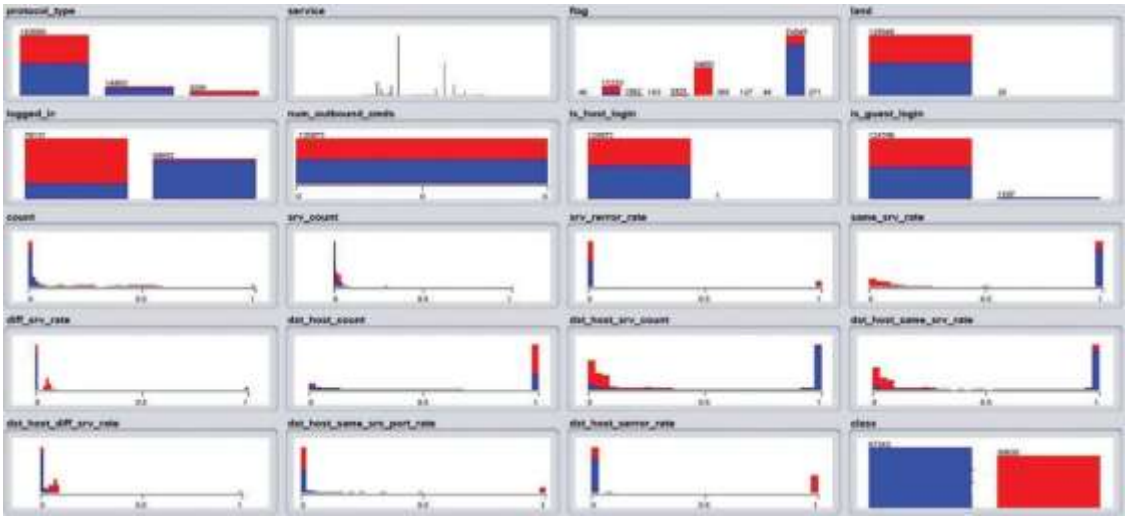

Fig 2: Visualization of some attributes in NSL-KDD data

2.2 Dataset Preprocessing

Data preprocessing is an important task to achieve accurate results through classifiers. Pre- processing of the original NSL-KDD data set is performed in MATLAB, which renders it an acceptable input to the classifier. Various data preprocessing techniques are available, but we use the following: Dataset Transformation and Dataset Normalization.

a).Data Transformation

In data transformation of nominal features to numeric value has been done. All the cate- gorical values of attributes are transformed into numeric form. We assign random numbers to them. NSL-KDD Train 20% data set to have approximately 25192 Connection instances. Each connection instance has 42 features, including attack type. Value to the target class has also been assigned in the same way. Zero is set for the normal class, and one will be for attack or any other deviation.

b).Data Normalization

Data set normalization is necessary when we have a large dataset that includes thousands of rows in training and testing files. The NSL-KDD dataset is large, so for accurate results of the classifier for intrusion detection. In this paper, we used the Min-Max method for the normalization of the dataset.

2.3. Feature Selection

Feature selection techniques are built to eliminate the number of input variables that are considered most appropriate for the model to predict the target variable. Random Forest (RF- RFE) is a machine-learning approach that could be perfect for integrating omics data commonly works well for high-dimensional problems and can classify strong predictors of a given outcome without making assumptions about the underlying model. Tree considers a different subset of randomly selected predictors, of which the best predictor is selected and split on at each node. In recursive features, elimination features are rank according to their importance score in a particular context. The significance of each feature is calculated at each iteration, and features with less importance are eliminated. Over a different subset of features, the importance of the same 46 feature can significantly vary while f evaluating highly correlated features. Protocol-type, logged-in, error-rate, srv-s error-rate, same-srv-rate are the features that are selected after applying the RF-RFE method.

2.4. Classification Process

SVM is a statistically supervised machine learning method that has been commonly used in the last few years. Classifies data inputs to various groups by creating a hyperplane. To conduct binary classification on data, SVM used a large dimensional space to figure out the hyperplane. The goal of optimizing the separation margin between the two groups is to find a hyperplane. SVM is used to identify input features into two distinct normal and attack classes. A line hyperplane separates a two-dimensional linearly separable input training data.

4. CONCLUSION

In this research, a detection system is projected on the NSLKDD dataset by applying data transformation and maximization and minimization methods. FGSVM is used to classify the NSLKDD dataset into two classes normal class and attack class. Substantial results obtained. from FGSVM have shown 99.03% to identify DDOS, probe, U2R, and R2L. FGSVM identified abnormal patterns are stimulated through ANFIS. According to their importance scores and prediction role, five features are selected for training, testing, and validation procedures of ANFIS. Error tolerance and epochs are set from zero and 100. During training, the MSE is 0.08523, and on testing and validation, the MSE is 0.08496, which reflects reasonable accuracy rates for the precise identification of DDOS, Probe, R2U, and U2R. To find out the intrusion and to prevent the network from it, ANFIS quickly build previously extracted connections and record-based measures as it is a rule-based method. FGSV‘M and ANFIS can be robust potential solutions that significantly boost their efficiency as more machine learning. In the future, deep learning-based classifiers are available to classify data and make systems more accurate and cost-effective. Advanced machine learning and AI-based systems are improving in intrusion detection with a higher accuracy rate.

5. REFERENCES

1. R. Heady, G. Luger, A. Maccabe and M. Servilla, “The architecture of a network level intrusion detection system,” Tech. Rep., Los Alamos National Lab, New Mexico University, Albuquerque, NM (United States), 1990.
2. S. Bhattacharya, P. K. R. Maddikunta, R. Kaluri, S. Singh, T. R. Gadekallu et al. “A novel PCA-firefly based XGBoost classification model for intrusion detection in networks using GPU,” Electronics, vol. 2, no. 19, pp. 219, 2020
3. RM, S. Priya, P. K. R. Maddikunta, M. Parimala and S. Koppu, T. R. Gadekallu et al. “An effective feature engineering for DNN using hybrid PCA-GWO for intrusion detection in IoMT architecture,” Computer Communications, vol. 160, pp. 139–149, 2020.
4. H. Debar, “An introduction to intrusion-detection systems,” Proc. of Connect, vol. 2000, 2000.

ISSN: 2456-236X

Vol. 10 Issue 01 | 2025

5. S. Namasudra, “Fast and secure data accessing by using dna computing for the cloud environment,” IEEE Transactions on Services Computing, 2020.
6. S. Namasudra, R. Chakraborty, A. Majumder and N. R. Moparthi, “Securing multimedia by using DNA-based encryption in the cloud computing environment,” ACM Transactions on Multimedia Com- puting, Communications and Applications (TOMM), vol. 16, no. 3s, pp. 1–19, 2020.
7. S. Kumari and S. Namasudra, “System reliability evaluation using budget constrained real d-mc search,” Computer Communications, vol. 171, pp. 10–15, 2021.
8. S. Kumari, R. J. Yadav, S. Namasudra and C.-H. Hsu, “Intelligent deception techniques against adversarial attack on the industrial system,” International Journal of Intelligent Systems, vol. 36, no. 5,pp. 2412–2437, 2021.