An Efficient Network Intrusion Detection System Using ANN and SVM with Feature Selection

Y.Nagamalleswararao [1],K.Pavani[2],S.Dharani Sri Lakshmi Bashitha[3]

#1 Assistant Professor in the Department Of MCA, SRK Institute Of Technology,Vijayawada.

#2 Assistant Professor & Head of Department of MCA, SRK Institute of Technology, Vijayawada.

#3 Student in the Department of MCA, SRK Institute of Technology, Vijayawada

Abstract: Due to cyberattacks, network security has become a major issue as internet-based services proliferate. This research introduces an efficient Network Intrusion Detection System (IDS) that uses supervised machine learning and feature selection to categorize network traffic as normal or malicious. The suggested system uses SVM and ANN algorithms to choose the most relevant features, reduce dimensionality, and enhance model performance utilizing filter and wrapper techniques.

The labeled NSL-KDD dataset trains and evaluates the system on network assaults and regular traffic. Before model training, data is cleaned, encoded, and normalized. Experimental findings show that the ANN model with wrapper-based feature selection outperforms the SVM model in detection and classification.

Improved accuracy and fewer false positives make the proposed intrusion detection method more useful for real-world network security applications. Detecting unknown or zero-day assaults is difficult, requiring significant deep learning research.

Index terms *- — Network Intrusion Detection System (IDS), Supervised Machine Learning, Feature Selection, Artificial Neural Network (ANN), Support Vector Machine (SVM), NSL-KDD Dataset, Cyber Security, Classification, Anomaly Detection, Wrapper Method, Filter Method*

1. INTRODUCTION

With the rapid expansion of internet technologies and online services, network security has become a major concern for organizations and individuals. Cyber-attacks such as denial-of-service (DoS), probing, and unauthorized access are increasing in both frequency and complexity. Traditional security mechanisms like firewalls and encryption are not sufficient to detect all types of attacks, especially new and unknown threats. Hence, Intrusion Detection Systems (IDS) play a vital role in identifying malicious activities by continuously monitoring network traffic.

IDS can be broadly classified into signature-based and anomaly-based detection systems. Signature-based systems are effective in detecting known attacks but fail to identify new or zero-day attacks. On the other hand, anomaly-based systems can detect unknown attacks but often suffer from high false

positive rates. To overcome these limitations, machine learning techniques have been widely adopted in IDS to automatically learn patterns from network data and improve detection accuracy.

Supervised machine learning algorithms, such as Support Vector Machine (SVM) and Artificial Neural Network (ANN), are particularly effective in classification problems where labeled data is available. However, network datasets often contain a large number of features, many of which may be irrelevant or redundant. This can negatively impact model performance and increase computational complexity. Therefore, feature selection techniques are essential to identify the most relevant attributes, reduce dimensionality, and enhance the efficiency of the learning process.

In this paper, a network intrusion detection system is proposed using supervised machine learning techniques combined with feature selection methods. The system is evaluated using the NSL-KDD dataset, a widely used benchmark dataset for intrusion detection research. A comparative analysis is conducted between SVM and ANN models, and the results show that ANN with wrapper-based feature selection achieves better performance in terms of detection accuracy.

The main contribution of this work is the development of an efficient IDS model that integrates feature selection with supervised learning algorithms to improve classification performance. The proposed system not only enhances detection accuracy but also reduces false alarms, making it suitable for practical network security applications.

2. LITERATURE SURVEY

a) Testing Gottfredson and Hirschi's "low self-control" stability hypothesis: An exploratory study:

According to Gottfredson and Hirschi's (1990) A General Theory of Crime, "low self-control," a consistent individual psychological feature, is the cause of criminal and similar acts. The idea that a lack of self-control is a stable trait is tested in this article. We examine college students' self-reported self-control at two reasonably near time periods using a two-wave panel research. We use four different data analysis (t-tests, correlations, individual change ratings, and HLM analyses) to examine the stability hypothesis. Overall, we discover that the majority of the self-control construct's characteristics seem to have remained largely consistent throughout this little time frame. However, the strength of the stability is not always evident because the measurements were made in close proximity to one another and we do not uncover correlations as strong as may be anticipated.

b) Detection of Network Intrusion by using Supervised Machine Learning Technique with Feature Selection:

To identify network traffic and categorize it as benign or harmful It makes use of a revolutionary supervised machine learning system.Both the supervised learning algorithm and the feature selection approach have been combined to produce the optimal model while taking detection success rate into account. The Artificial Neural Network (ANN) with wrapper feature selection support vector machine (SVM) approach is used to categorize network traffic. SVM and ANN supervised machine learning algorithms are used to categorize network traffic using the NSL-KDD dataset in order to assess performance. Comparative analysis reveals that the

suggested model outperforms other current models in terms of intrusion detection success rate.

c) Modeling and Implementation Approach to Evaluate the Intrusion Detection System:

Systems known as intrusion detection systems (IDSs) attempt to identify assaults either while or after they take place. Reducing the effect of assaults and evaluating the system IDS were the two goals of this field's research. In fact, on the one hand, intrusion detection systems (IDSs) gather network traffic data from various sources within the network or computer system and utilize this information to improve system security. On the other hand, evaluating IDS is an essential responsibility. In reality, it's critical to recognize the distinction between assessing a system's overall efficacy and assessing the features of its individual components. In this study, we propose an IDS evaluation method based on component performance measurements. First, we have suggested a hardware platform based on embedded systems to properly integrate the IDS SNORT components. Next, we evaluated it using a traffic and attack generator based on the Metasploite 3 Framework and Linux KALI (Backtrack). The findings obtained indicate a tight relationship between the properties of these components and the IDS performance.

d) Toward Credible Evaluation of Anomaly-Based Intrusion-Detection Methods

The field has expanded significantly since anomaly-based intrusion detection was originally introduced to the scientific community in 1987. Numerous strategies and tactics were created that added new skills for identifying fresh attackers. At a low false alarm rate of 1%, the majority of these methods indicate a high detection rate of 98%. The industry often prefers signature-based detection for the widespread deployment of intrusion-detection systems, notwithstanding the attractiveness of the anomaly-based method. Although several anomaly-detection strategies have been put forth, it is challenging to compare their advantages and disadvantages in a way that may result in commercial use. It is reasonable to suppose that research in anomaly detection shares the aforementioned issue as the legitimacy of experimental research in academic computer science is often dubious. The lack of industrial adoption of anomaly-based intrusion-detection techniques may be partially explained by doubts regarding their veracity. In order to look into this matter, we assess 276 works that were published between 2000 and 2008 and examine the current experimental practice in the field of anomaly-based intrusion detection. We provide an overview of our findings and point out the typical mistakes made by the examined works.

e) Importance of Intrusion Detection System (IDS):

Computer intruders, who are dispersed over the Internet, have grown to be a serious concern in our world. The experts suggested a variety of methods, including encryption and firewalls, to stop this kind of intrusion and safeguard computer infrastructure; yet, the hackers were still able to get access to the systems. Researchers have focused a lot of emphasis on IDS, which keeps an eye on the computer resources and detects any anomalies or odd trends. This paper aims to explain the stages of the evolution of the concept of intrusion detection systems (IDS) and its significance for researchers, research centers, security, and the military. It also looks at the significance of intrusion detection systems and their

categories and classifications, as well as where they can be placed to lower network risk.

3. METHODOLOGY

i) Proposed Work:

The proposed system is designed to develop an efficient Network Intrusion Detection System (IDS) by integrating feature selection techniques with supervised machine learning algorithms. Initially, the NSL-KDD dataset is collected and preprocessed by removing irrelevant attributes, handling categorical values through encoding, and normalizing the data. Feature selection is then applied using both filter and wrapper methods to identify the most relevant features, reduce dimensionality, and eliminate redundant data. This step improves model efficiency and reduces computational complexity while maintaining important information required for accurate classification.

After feature selection, supervised learning models such as Support Vector Machine (SVM) and Artificial Neural Network (ANN) are trained using the selected features. The trained models are then evaluated on unseen test data to measure their performance in classifying network traffic as normal or malicious. A comparative analysis is conducted between SVM and ANN, where the ANN model with wrapper-based feature selection demonstrates superior performance in terms of accuracy and detection rate. The proposed system enhances intrusion detection capability, reduces false positives, and provides a reliable solution for real-time network security applications.

ii) System Architecture:

The system architecture of the proposed Network Intrusion Detection System (IDS) is designed to efficiently process network data and classify it into normal or malicious categories using supervised machine learning techniques. The process begins with the NSL-KDD dataset, which contains labeled network traffic data. This dataset is first passed through a data preprocessing stage, where normalization and cleaning operations are performed to remove noise, convert categorical values into numerical form, and prepare the data for further processing.

After preprocessing, the dataset is divided into training and testing sets. A feature selection module is then applied to extract the most relevant attributes using filter and wrapper methods, reducing dimensionality and improving model performance. The selected features are fed into the model training phase, where supervised learning algorithms such as SVM and ANN are used to build classification models. Finally, the trained model is applied to the testing data in the prediction phase, where network traffic is classified into different categories such as DoS, Probe, R2L, U2R, or Normal. This architecture ensures improved detection accuracy, reduced computational complexity, and effective identification of network intrusions.

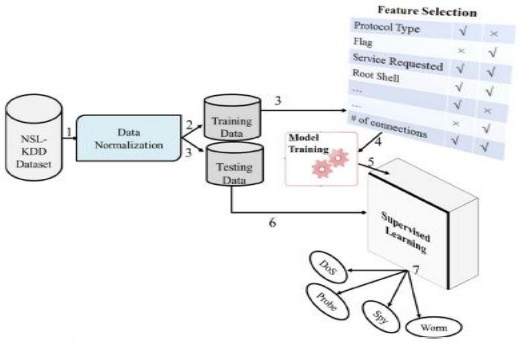


Fig 1: Proposed supervised machine learning classifier system

Fig1 proposed architecture

iii) Modules:

1. Data Collection and Preprocessing Module

This module is responsible for collecting the NSL-KDD dataset and preparing it for analysis. It performs data cleaning, removes irrelevant attributes, converts categorical values into numerical form, and applies normalization to ensure consistency in the dataset.

2. Feature Selection Module

This module selects the most relevant features from the dataset using filter and wrapper methods. It helps in reducing dimensionality, eliminating redundant data, and improving the efficiency and accuracy of the machine learning models.

3. Training Module

In this module, the processed and selected features are used to train supervised machine learning models such as Support Vector Machine (SVM) and Artificial Neural Network (ANN). The model learns patterns from labeled data to distinguish between normal and malicious traffic.

4. Testing and Evaluation Module

This module evaluates the trained model using test data. It measures performance metrics such as accuracy, detection rate, and error rate to analyze how well the model performs on unseen data.

5. Prediction Module

The prediction module applies the trained model to new incoming network traffic. It classifies the data into categories such as normal or different types of attacks (DoS, Probe, etc.), enabling real-time intrusion detection.

6. Result Visualization Module

This module presents the output in a user-friendly format using graphs and accuracy comparisons. It helps in understanding model performance and comparing results between SVM and ANN algorithms.

iv) Algorithms:

1. Support Vector Machine (SVM)

Support Vector Machine is a supervised machine learning algorithm used for classification tasks. It works by finding an optimal hyperplane that separates data points into different classes with maximum margin. In the proposed system, SVM is used to classify network traffic into normal and attack categories based on selected features. It performs well for linearly separable data and provides good generalization, but its performance may decrease with large and complex datasets.

2. Artificial Neural Network (ANN)

Artificial Neural Network is a powerful supervised learning algorithm inspired by the human brain. It consists of input, hidden, and output layers, where neurons are interconnected with weights. In this system, ANN is used to learn complex patterns in network traffic data and classify it accurately. The model uses multiple hidden layers and a learning rate

to improve performance. ANN shows better accuracy compared to SVM, especially when combined with wrapper-based feature selection.

4. EXPERIMENTAL RESULTS

The experimental evaluation of the proposed Network Intrusion Detection System was carried out using the NSL-KDD dataset, which consists of labeled network traffic records representing both normal and various attack types. Initially, the dataset was preprocessed by removing irrelevant features, converting categorical attributes into numerical values, and applying normalization techniques to ensure uniformity. The processed dataset was then divided into training and testing sets, where approximately 80% of the data was used for training and 20% for testing. Feature selection techniques, including filter and wrapper methods, were applied to reduce dimensionality and improve model efficiency.

The performance of the system was evaluated using two supervised machine learning algorithms: Support Vector Machine (SVM) and Artificial Neural Network (ANN). The SVM model achieved an accuracy of around 96.88%, while the ANN model demonstrated superior performance with higher detection accuracy and better classification of attack types. The results clearly indicate that ANN combined with wrapper-based feature selection provides improved intrusion detection capability compared to SVM. Additionally, graphical analysis shows that ANN consistently outperforms SVM in terms of accuracy, making it a more suitable choice for real-time intrusion detection systems.

Accuracy: A test's accuracy is its capacity to distinguish healthy from ill cases. Find the percentage of instances with genuine positives and negatives to assess test accuracy.

Accuracy = TP + TN /(TP + TN + FP + FN)

$$Accuracy = \frac{(TN + TP)}{T}$$

Precision: Classification accuracy or positive cases constitute precision. The formula for accuracy is:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\Pr e\,cision = \frac{TP}{(TP + FP)}$$

Recall: A model's recall measures its ability to recognize all appropriate machine learning class instances. The ratio of accurately predicted positive observations to total positives indicates a model's class instance detection skill.

$$Recall = \frac{TP}{(FN + TP)}$$

mAP: Mean Average Precision ranks quality. It considers the number and order of relevant ideas. Calculating MAP at K uses the arithmetic mean of each user or query's Average Precision (AP).

$$\boldsymbol{mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k}$$

$$\boldsymbol{AP_k = \text{the AP of class } k}$$

$$\boldsymbol{n = \text{the number of classes}}$$

F1-Score: A high F1 score suggests an accurate machine learning model. Integrating recall and precision improves model correctness. Accuracy

measures how often a model predicts a dataset correctly.

$$F1 = 2 \cdot \frac{(Recall \cdot \Pr e\,cision)}{(Recall + \Pr e\,cision)}$$

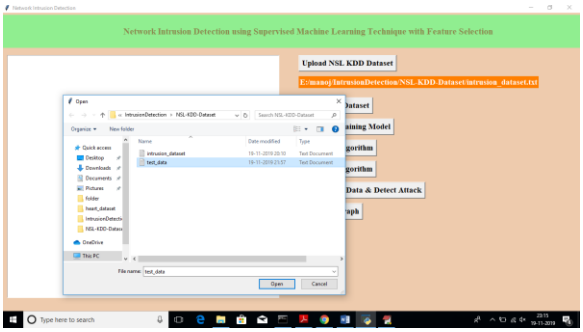


Fig2 upload dataset

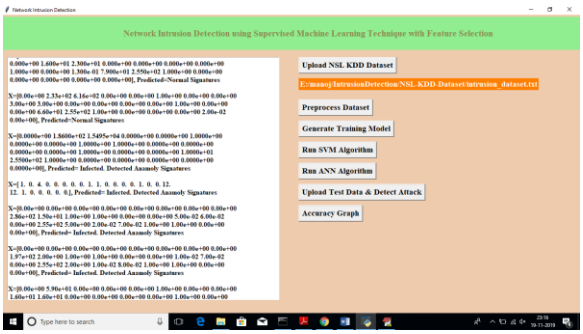


Fig3 predicted results

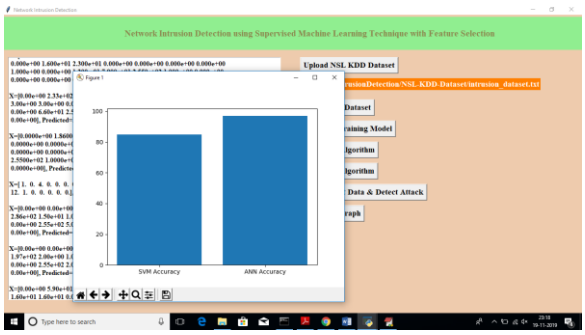


Fig4 accuracy compare graph

5. CONCLUSION

This paper presents an efficient Network Intrusion Detection System (IDS) using supervised machine learning techniques combined with feature selection methods. The system utilizes SVM and ANN algorithms to classify network traffic as normal or malicious, with preprocessing and feature selection improving model performance. Experimental evaluation using the NSL-KDD dataset demonstrates that the ANN model with wrapper-based feature selection achieves higher accuracy and better detection rates compared to the SVM model.

The proposed approach significantly enhances intrusion detection capability by reducing false positives and improving classification efficiency. It provides a reliable solution for identifying various types of network attacks in real-time environments. However, the system still faces challenges in detecting zero-day or unknown attacks, indicating the need for further research using advanced deep learning and hybrid techniques to improve overall IDS performance.

6. FUTURE SCOPE

The proposed Network Intrusion Detection System can be further enhanced by integrating advanced deep learning techniques such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks to improve the detection of complex and zero-day attacks. Future work can also focus on developing a real-time intrusion detection system capable of handling high-speed network traffic with minimal latency. Additionally, incorporating hybrid models and ensemble learning methods may further improve classification accuracy and reduce false positives.

Moreover, the system can be extended to work in cloud and IoT environments, where security challenges are more dynamic and critical. Implementing automated feature engineering and adaptive learning mechanisms can help the model continuously update itself based on new attack

patterns. Enhancing scalability and deploying the system in distributed environments will make it more practical for real-world applications in large-scale networks.

REFERENCES

[1] H. Song, M. J. Lynch, and J. K. Cochran, "A macro-social exploratory analysis of the rate of interstate cyber-victimization," American Journal of Criminal Justice, vol. 41, no. 3, pp. 583–601, 2016.

[2] P. Alaei and F. Noorbehbahani, "Incremental anomaly-based intrusion detection system using limited labeled data," in Web Research (ICWR), 2017 3th International Conference on, 2017, pp. 178–184.

[3] M. Saber, S. Chadli, M. Emharraf, and I. El Farissi, "Modeling and implementation approach to evaluate the intrusion detection system," in International Conference on Networked Systems, 2015, pp. 513–517.

[4] M. Tavallaee, N. Stakhanova, and A. A. Ghorbani, "Toward credible evaluation of anomaly-based intrusion-detection methods," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40, no. 5, pp. 516–524, 2010.

[5] A. S. Ashoor and S. Gore, "Importance of intrusion detection system (IDS)," International Journal of Scientific and Engineering Research, vol. 2, no. 1, pp. 1–4, 2011.

[6] M. Zamani and M. Movahedi, "Machine learning techniques for intrusion detection," arXiv preprint arXiv:1312.2177, 2013.

[7] N. Chakraborty, "Intrusion detection system and intrusion prevention system: A comparative study," International Journal of Computing and Business Research (IJCBR) ISSN (Online), pp. 2229–6166, 2013.

Author Profiles



Mr. Y. Nagamalleswararao completed his Master of Technology (M.Tech) from JNTUK, M.Sc (IS) from ANU, and BCA from ANU. He has expertise in System Administration, Network Administration, and Oracle Administration. He is also a Web Developer and Python Developer. Currently, he is working as an Assistant Professor in the Department of MCA at SRK Institute of Technology, Enikepadu, NTR District. His areas of interest include Artificial Intelligence and Machine Learning.



Ms. K. Pavani is working as an Assistant Professor and Head of the Department of MCA at SRK Institute of Technology, Vijayawada. She completed her M.Tech and MCA in Computer Science. She has 10 years of teaching experience at SRK Institute of Technology, Enikepadu,

Vijayawada, NTR District. Her areas of interest include Machine Learning with Python and DBMS.



Mrs. ,S.Dharani Sri Lakshmi Bashitha[3] is an MCA student in the Department of Computer Applications (MCA) at SRK Institute of Technology, Enikepadu, Vijayawada, NTR District. He completed his degree in B.Sc. (Computers) from Sri CRR Degree College For women, Vijayawada. His areas of interest include DBMS and Machine Learning with Python.