Network Based Multiple Intrusion Detection Using Machine Learning

Prof.Anish Kumar
Assistant Professor
Computer science and Engineering
MEA Engineering College
Perinthalmanna,India
anish@meaec.edu.in

Hunais. PM
Computer science and Engineering
MEA Engineering College
Perinthalmanna,India
19mcs49@meaec.edu.in

Mohammed Sanoob
Computer science and Engineering
MEA Engineering College
Perinthalmanna,India
19gcs38@meaec.edu.in

Mohamed Ashmal.M
Computer science and Engineering
MEA Engineering College
Perinthalmanna,India
19mcs21@meaec.edu.in

Mohamed Danis. AP
Computer science and Engineering
MEA Engineering College
Perinthalmanna,India
19mcs43@meaec.edu.in

***Abstract*—The rapid growth of the internet and communication industries has led to an expansion of both networks and data. This has also resulted in an increase of security threats and difficulties in detecting network intrusions. Intruders with malicious intents pose a major danger to the network's confidentiality, integrity, and availability. In order to decrease these risks, network traffic is monitored and any intrusions are stopped using a network intrusion detection system (NIDS). Machine learning (ML) is being used to detect network attacks through the usage of ML-based Network Intrusion Detection Systems (NIDS). These systems' main objective is to offer effective network-wide detection. The expansion of connected devices and network intrusion detection systems faces challenges at every stage of development and operation as attacker tactics and methods constantly change. As a result, ML methods are increasingly used in NIDS.**

Keywords: network traffic, machine learning and intrusion detection systems.

I. INTRODUCTION

In today's digital age, high-speed information and communication networks are becoming an essential part of life.They function as platforms for the swap of digitized information as well as service providers, offering a range of services to its consumers. People and companies have become lucrative targets for cyberattacks because of their reliance on computer networks. Through a number of network breaches, cybercriminals attempt to jeopardise the privacy, integrity, and accessibility of online data and services. Intrusion detection systems (IDSs) were created to detect such invasions. IDSs track and examine online traffic to separate legitimate from hazardous content. Intrusion Detection Systems (IDSs) are IDSs that are implemented within a network to detect network-based intrusions. These programmes monitor and assess network traffic online.

As the number of networked IT devices increases worldwide, intrusion detection systems (IDS) have grown in importance in recent years.IDS are used to detect hostile behaviour.Two categories of IDS approaches are anomaly-based methods and signature-based (misuse) methods. Unlike signature-based strategies, which only detect well-known malicious activity and are unaware of new activity,An anomaly-based approach can detect undetected attacks, including potential zero-day vulnerabilities. It works by detecting deviations from the normal flow of traffic. Signature-based IDS is particularly effective at identifying known threats by comparing predefined signatures to targeted traffic. However, when dealing with unknown targets, it is completely useless.

Automating the creation of analytical models is possible. It is an analytical method for data. One of the applications of artificial intelligence that relies less on human interaction is learning, decision-making, and pattern recognition performed by a computer. The two techniques to machine learning that are most frequently used are supervised learning and unsupervised learning. Algorithm training uses labelled samples that closely resemble input and yield desired outcomes. To learn about occurrences without prior classification, utilise unsupervised learning.Finding some order in data and learning about data are the two main goals of unsupervised learning. Reinforcement learning and semi-supervised learning are also used with these methods.The most used IDS technique is supervised machine learning. Decision trees, K-Nearest Neighbors (KNN), etc.

II. REVIEW OF RELEVANT LITERATURE

1.NNIDS: Neural Network based Intrusion Detection Technique.

Deep neural networks (DNN) are used in this live malware detection technique based on network activity, and it then searches for the best malware classification methods.It is capable of detecting novel and dynamically

changing intrusion instances.But it is susceptible to Protocol Based Attacks.They face the same protocol based attacks as network hosts

2.Web-Based Intrusion Detection System

An IDS that targets web-based intrusions operates in the application layer and is specially designed to detect external hackers and web-based attacks. However, it only concentrates on a limited set of intrusion methods as outlined in the central concepts chapter and may not detect other types of attacks.

3.Machine Learning based Intrusion Detection System for Web-Based Attacks

Applied machine learning techniques to the CSIC HTTP dataset 2010[1] which is a publicly accessible dataset of HTTP traffic.J48 is used in this situation to accurately classify a variety of applications.However, the pace of training and testing is slow.

4.I-SiamIDS: an improved Siam-IDS for handling class imbalance in network-based intrusion detection Technique

I-SiamIDS tackles the class imbalance issue more effectively and recognises more attacks. By filtering the innocuous traffic numerous times, it lowers the chance that hostile traffic will be incorrectly classified and fewer attacks will go undetected.But it is expensive to implement and also difficult to optimize.

5.Intrusion Detection System Through Advanced Machine Learning for IoT Networks

ML algorithms could be advanced by a genetic algorithm (GA) that could choose the right characteristics. GA is employed in this case to select the best characteristics and improve accuracy. Genetic algorithms, however, are unable to guarantee consistent optimization response times.

6.Intelligent Cyber Attack Detection and Classification for Network-Based Intrusion Detection Systems

NIDS uses feature vectors that contain a summary of network traffic over a given period of time,HIDS, on the other hand, is installed on a specific host and tracks system information.Use anomaly-based intrusion detection. It can detect new or previously undetected types of attacks. However, it has a high false positive rate and has the potential to be fooled by well-placed attacks.

7.Recurrent Neural Network Based Intrusion Detection Technique.

It detects network traffic irregularities using an LSTM-based machine learning model, and it sends all malignant requests to a honeypot-based black hole server.It is accurate on both observed and unseeded data and is highly efficient, but it is also slow. Training can be challenging.

8.Machine Learning Based Intrusion Detection System

Comparative analysis done between SVM and Naïve Bayes for classification of dataset,to examine their accuracy and Misclassification Rate. Which has been enhanced to decrease false alarms, boost detection rates, and have the capacity to identify both known and unidentified assaults. However, as a result of the enormous volume of data, false alarm reports of network intrusion increase and detection accuracy declines.

9.Decision Tree-Based Intrusion Detection System for Internet-of-Things Networks

A technique that combines REP trees, JRIP algorithms, and Forest PA classifiers for hierarchical intrusion detection. Which of the following 7 attack types has the highest true negative ratio (TNR) and detection rate? (DR). However, unlike other selection predictors, decision trees are primarily unstable.

III. PROPOSED SYSTEM

A. Architecture Diagram

The architecture of a NIDS is designed to provide real-time monitoring and detection of suspicious activity on a network, as well as the ability to respond appropriately to any incidents that are detected.

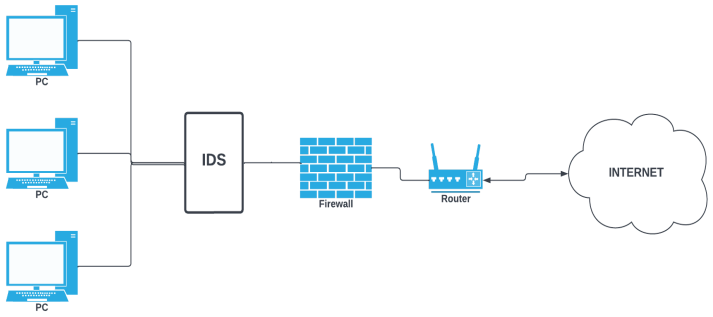

Fig 3.1 Architecture Diagram

IDS and firewalls are both crucial for network security. However, a firewall distinguishes itself from an IDS by proactively seeking and stopping external intrusions. Firewalls limit network access to prevent infiltration, yet an assault from within the network goes unnoticed.

After a suspected intrusion has happened, an intrusion detection system (IDS) analyses it to determine its characteristics before sounding an alert. Because a firewall is a proactive defence system, making it closer to an intrusion prevention technique (IPT) than an intrusion detection technique (IDT). An IPS, like an IDS, detects threats but goes further by actively countering known ones.

B. Flow Chart

The aim of a NIDS gadget is to stumble on and prevent any harmful activity on a community, which include attempts to harm the confidentiality, integrity, or availability of community assets. A series of steps to be followed :

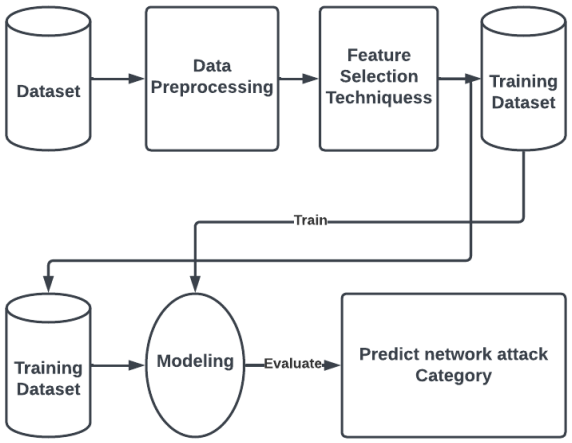

Fig 3.2 Dataflow Diagram

1. Data Collection

For trespass detection systems, the KDD CUP 99 Dataset is a frequently used and well-known dataset. There are two million records for testing and five million records for training. The KDD CUP 99 Dataset records are classified as either attacks or normal based on their 41 distinctive features. DoS, Probe, R2L, and U2R are the four subcategories under which the assaults fall. The Intrusion Detection System programme at MIT's Lincoln Laboratory was examined in 1998 and 1999, and this led to the creation of the KDD CUP 99 Dataset. DARPA provided funding for the Lincoln Laboratory at MIT's Intrusion Detection System programme, which produced the well-known DARPA98 dataset.The well-known KDD was created using the DARPA98 dataset and later refined for an international competition for knowledge retrieval and data mining tools.

2. Data Preprocessing

The first technique on this project is to preprocess the data. In this phase, dataset reading and attack type feature adding is done where the attack type has five different values such as DoS, normal, probe, R2L,U2R. The labels are assigned to it's attack categories. Next, find the missing values of all the features, if not found move to the next process.

3. Feature mapping

Apply feature mapping to all the features and remove all the unimportant features before modelling. It is done using an algorithm called RFE(Recursive Feature Elimination), which is a feature selection algorithm that works by recursively removing features that are least important to the model. This is done by iteratively training a model on a subset of the features, and then removing the features that have the least impact on the model's performance. This RFE uses decision tree algorithm CART as the facilitator parameter which is essential in feature elimination.

4. Training and Testing

The dataset is divided into both training and testing data. The 90 percent of data from the dataset is taken for training and the other 10 percent is used for testing. Then repeat the process ten times.

During the training phase, the machine learning model learns from the labeled training data to identify patterns. The testing phase evaluates the trained model's performance on unseen data to measure its effectiveness in making accurate predictions.

5. Modelling

In modelling, the first dataset splitting has been done. Training set and testing set are the two dataset components.. The process of data cleaning and feature extraction is performed to collect the training dataset and eliminate any unnecessary information. Dimensionality reduction is applied during this process, transforming the initial raw data into manageable groups. These large datasets typically have a vast number of variables which require significant computing resources to process.Based on this, machine learning algorithms such as decision trees are used to classify and test patterns.The dataset is divided into training and test sets so that you can create custom models for attack detection.

IV. RESULT

The proposed algorithm was compared to other techniques for detecting attacks on the KDD CUP 99 test dataset. The comparison was based on the accuracy of each technique. The proposed algorithm was found to be more accurate than the other techniques in detecting attacks. This is likely due to the fact that the proposed algorithm takes into account a wider range of features when making its predictions.

Fig 4.1 Probe Output

Initially, there are four models made exclusively for 4 attack categories DoS, Probe, R2L and U2R. Feed the four models with input from the user device, and the respective models detect their attacks. Fig 4.1 shows the output of probe attack detection of probe machine learning model.

Fig 4.2 DoS Output

Fig. 4.2 shows the result of detection of DoS attack from DoS Model. Respectively Fig 4.3 shows the result of detection of R2L attack from the R2L model.

Fig 4.3 R2L Output

The below figure Fig 4.4 shows the output of U2R attack detection, The U2R attack involves an unauthorized user escalating their privileges from a regular user to a root or administrative level user.

Fig 4.4 U2R Output

The below Fig 4.5 shows the output “No attack detected”, If there is no attack detected which means it is a normal connection. The input packet details are passed through all four models and the four models cannot detect it, which result in normal connection.

Fig 4.5 No attack Output

Fig 4.6 is an RFECV plot for DoS attacks in NIDS, that shows the relationship between the number of features used in the model and the performance of the model in detecting DoS attacks. The plot typically shows the cross-validation score as a function .of the number of features used in the model.

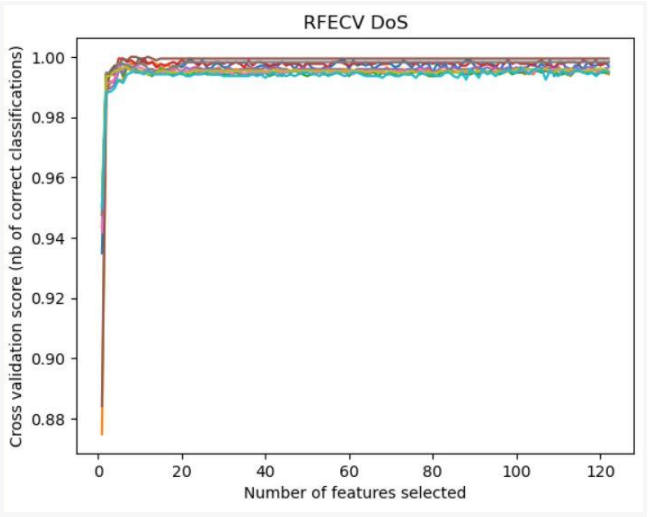

Fig 4.6 DoS Accuracy x No.of features graph

The RFECV DoS plot helps in selecting the optimal number of features for the NIDS model that can improve its performance in detecting DoS attacks. By selecting the optimal number of features, the model effectively identifies the patterns and characteristics of DoS attacks, which enhance the accuracy and efficiency of the NIDS system in detecting and preventing DoS attacks.

An RFECV plot for R2L attempts in NIDS, shown in Fig 4.7, illustrates the correlation between the model's performance in identifying R2L attacks and the number of features it uses. The cross-validation score is often plotted as a function of the model's feature count.

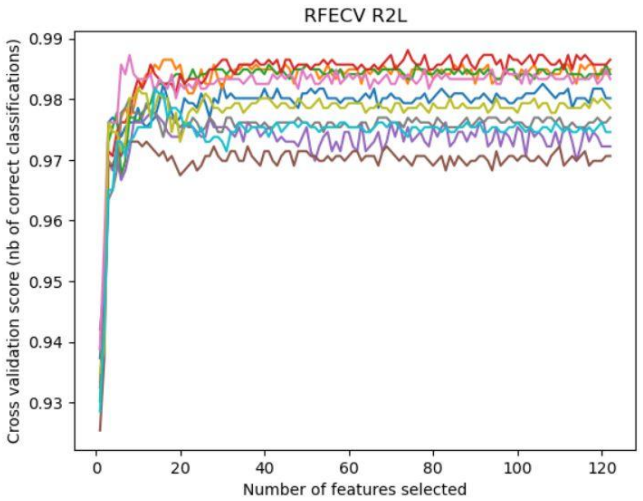

Fig 4.7 R2L Accuracy x No.of features graph

The NIDS model's performance in identifying R2L attacks is improved by choosing the right number of features, which is assisted by the RFECV R2L plot. The model effectively recognises the patterns and characteristics of R2L attempts by choosing the ideal number of features, which improves the NIDS system's accuracy and efficiency in identifying and stopping R2L attacks.

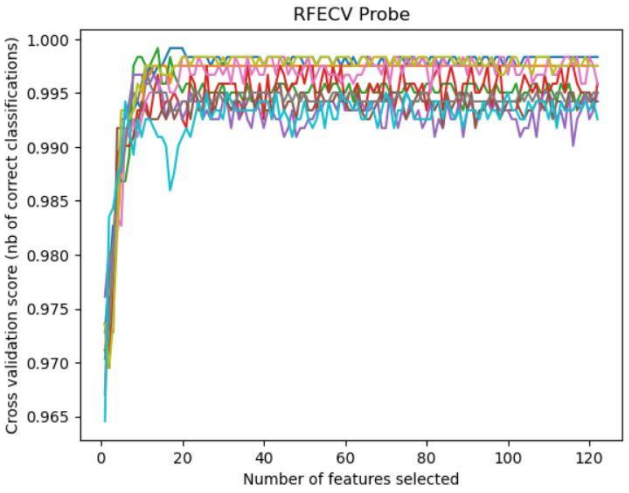

Fig 4.8 Probe Accuracy x No.of features graph

Fig 4.8, an RFECV plot for Probe attacks in NIDS, illustrates the correlation between the model's performance in identifying Probe attacks and the number of features it uses. In most cases, the cross-validation score is plotted as a function of the model's feature count.

The RFECV Probe plot aids in choosing the NIDS model's ideal feature count to enhance its performance in identifying Probe attacks. By choosing the right number of features, the model successfully recognises the patterns and features of Probe attacks, improving the NIDS system's effectiveness and accuracy in recognising and preventing Probe attacks.

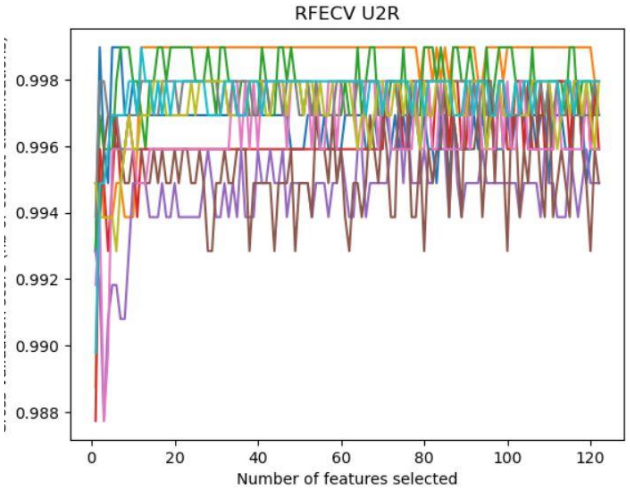

Fig 4.9 U2R Accuracy x No.of features graph

The RFECV plot in Fig 4.9 for U2R attacks in NIDS demonstrates the correlation between the model's ability to detect U2R attacks and the number of features it uses. The cross-validation score as a function of the model's feature count appears frequently on the plot.

In order to increase the NIDS model's ability to detect U2R attacks, the RFECV U2R plot assists in choosing the ideal number of features for the model. The model successfully identifies the patterns and features of U2R attacks by choosing the ideal number of features, which improve the NIDS system's accuracy and effectiveness in identifying and stopping U2R attacks.

V. CONCLUSION

This paper includes current information on the most recent trends and achievements in the industry and presents a thorough evaluation of ML related Network Intruder Detection techniques. Using the KDD Cup99 data sets, which are freely accessible and include a wealth of findings, the usefulness of the suggested approaches was assessed. These datasets, however, are obsolete and might not correctly reflect contemporary network threats, which hinders the effectiveness of the suggested solutions in real-world settings. A possible future direction for research in this area is enhancing detection accuracy for incursions by suggesting a more effective NID framework employing straightforward machine learning methods and efficient detection mechanisms.Effective implementation of a machine learning-based NID system enables accurate identification of network intrusions.

REFERENCES

[1] Zeeshan Ahmad,Adnan Shahid Khan,Cheah Wai Shiang and Farhan Ahmad,"Network intrusion detection system:A systematic study of machine learning and deep learning approaches (Ahmad #)"in 2020.

[2] S Latif, FF Dola, MD Afsar and IJ Esha,"Investigation of Machine Learning Algorithms for Network Intrusion Detection",in 2022.

[3] T Wisanwanichthan and M Thammawichai,"A double-layered hybrid approach for network Intrusion detection system using combined naive bayes and SVM" in 2021 IEEE.2021,pp.138442-138450

[4] GDC Bertoli,LAP junior and O Saotome,"An end-to-end framework for machine learning-based network intrusion detection",in 2021 IEEE

[5] Samson Ho,Saleh AL Jufout,Khalil Dajani and Mohammad Mozumdar,"A Novel Intrusion Detection Model for Detecting Known and Innovative Cyberattacks Using Conventional Neural Network",IEE Open Journal of the Computer Society in 2021.

[6] [6]A Javid,Q Niyaz,W Sun and M Alam,"A deep learning approach for network intrusion detection system"in 2016.

[7] [7]MB Pranto,MHA Ratul,MM Rahman,IJ Diya and ZB zahir,"Performance of machine learning techniques in anomaly detection with basic feature selection strategy-a network intrusion detection system"in 2022.

[8] T Saba, T Sadad, A Rehman and Z Mehmood,"Intrusion detection system through advanced machine learning for the internet of thing networks",IEEE IT Professionals in 2022.

[9] N Oliveira, I Praca, E Maia and O Sousa,"Intelligent cyber attack detection and classification for network-based intrusion detection systems",in 2021

[10] Raisa Abedin Disha and Sajjad Waheed,"Performance

analysis of machine learning models for intrusion detection systems using Gini Impurity-based Weighted Random Forest(GIWRF)feature selection technique",in 2022.

[11] S Sharma, P Zavarsky and S Butakov,"Machine learning based intrusion detection system for web-based attacks",2020 IEEE Intl Conference on Intelligent Data and Security (IDS)

[12] P Bedi, N Gupta and V Jindal,"I-SiamIDS:an improved Siam-IDS for handling class imbalance in network-based intrusion detection systems",Applied Intelligence 51,1133-1151(2021)

[13] S Nayyar, S Arora and M Singh,"Recurrent neural network based intrusion detection system",2020 International Conference on Communication and Signal Processing (ICCSP)