IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Intrusion Detection In Networks And Server

[1]Mohd Muzammil, [2]Jakka Arun, [3]Sidharth Nookala, [4]Raj Mohan, [5]Dr. K.Butchi Raju

[1]Student, [2]Student, [3]Student, [4]Student, [5]Professor

[1]Department of Computer Science and Engineeering,

[1]Gokaraju Rangaraju Institute Of Engineering and Technology, Hyderabad, India

Abstract: The most sensitive and important data are stored on servers; strong security is required to avoid data theft and misuse. When an intrusion occurs in a system, an intrusion detection system (IDS) is used to identify it and alert the admin. A network and devices are inspected by an IDS for malicious activity or policy breaches. Any unlawful behaviour or a violation is often captured continuously using a security information and event management system and notified to an admin. In order to differentiate between hostile behaviour and spurious reports, a SIEM system aggregates outputs from many sources and use event filtering algorithms. In order to track traffic to and from all networked devices, intrusion detection systems (IDS) are installed at one or more strategically located locations inside the network. Our study is on UNSW_NB15 dataset which comprises of different attacks.

Index Terms **– Security, Computer Network, System**

I. INTRODUCTION:

Security is very essential and important need in today's digital world . Cyber security is the technique of protecting networks, electronic devices and data against malicious intrusions. Information security is the protection of internet-connected systems from online attacks. In today's society, the majority of data is in digital form and kept on internet-connected digital devices,cloud servers.In India there has been continuous increase in online payments.Customers have trust on the web applications and apps only when their data is secured and privacy is maintained. The field of computer science completely depends on cyber security , If there happens to be no security then no body will use those applications. The big gaint companies have made a positive trust and continuously improve their products from the threats.The attackers are using new techiques to get the information from server,cloud. Login information,encryption keys and banking data will be the majority of the data that is always at risk. Therefore, one must have an intrinsic security plan to safeguard the privacy of information. IDS using machine learning is a software that evaluates the packet data of incoming traffic and indicates whether or not it is is a genuine packet. Whenever a malicious packet is received, a notification type signal is sent to admin.

An intrusion detection (IDS) checks the network traffic for unusual behaviour and issues notifications when it detects. It is software that scans a network or system for potentially dangerous behaviour and rule breaches. Generally, intrusions are off two varities : Signature based and Anomaly based. The Siganture based method can quickly identify threats for whom the patterns has been present in the system. The amount of bytes, number of ones, or number of zeros in the network traffic are only a few examples of the specific patterns. The Anomaly based method detects malware depending on deviations from normal behaviour. False negatives from IDS are an issue since they allow threats to easily pass through the system and network because they are mistaken as for genuine traffic. As a result of this issue, nobody will be aware of any intrusions that have occurred, which can occasionally pose major hazards,loss to the company. Our technique makes easier and reduces the cost of loss in the company.

In the project we have employed machine learning (ML) models, and the forecast was made using the model with the best accuracy. In this project, two different datasets were employed. One for training and the other dataset for testing .Prediction is made using the additional dataset. Several models, including the Decision tree classifier,Naïve Bayes logistic regression, gradient boosting classifier, and support vector machine (SVM) for supervised learning, were used in this research. The incoming packet information is categorised into 10 distinct categories of attack type. Our machine learning-based IDS built a method to distinguish 9 various forms of cyberattacks and malware with a detection rate of more than 89%.

II. LITRATURE SURVEY:

[1] "A critical review of intrusion detection systems in the internet of things".The paper has given info on attackers that are targeting IoT devices.The IoT technology is evolving continuously in the digital world.It gives a comprehensive review of IDS and overview of techniqes,validation strategy,deployment strategy.It provides up to date taxonomy with machine learning techniques to make IoT IDS.

[2] "Network Intrusion Detection System using Deep Learning". This paper gave us insight and knowledge of deep learning techniques that are employed.We have learned different attack categories. It tells about the strengths and limitation of the security system that were developed without using the intelligence

[3] "Network Intrusion Detection System using Neural Network". The paper gave us an insight on how the neural network learns from the raw data provided. The network adjusts the weights according to the target. The paper also tells that the attacks are evolving but the neural network solution provides a generalized way of identifying the IDS .

[4] "Intrusion Detection : A comprehensive review". This paper gave us understanding on IDS and intrusion prevention, It is about the classification of IDS into wireless based.The network data is used to identify IDS.

III. METHODOLOGY:

Proposed Approach:

We developed a machine learning-based intrusion detection with a very much reduced coding effort. Introducing a new anomaly or attack type into the system is very easy. Just training the system with the new attack category one time is enough. If any intrusions of that type happen the system detects it a good accuracy of approximately 90%. The system can perfectly determine which packet is normal. So whenever there happens any intrusion, the system may not detect the type of intrusion perfectly but it can perfectly differentiate the packet as an anomalous packet and raises an alert. True positivity rate of predicting a normal packet is high.

Advantages

The technique is based on machine learning is more generic. Cost of maintenance is low. It is effective in identifying unknown threats or even known attack variations. Provides an additional layer of protection with more security. Developing such a system requires a less of coding effort.

Architecture:

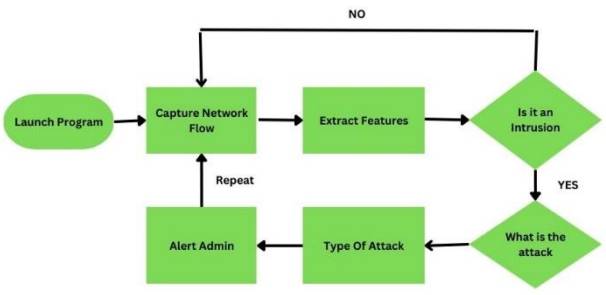

Fig.1: The above architecture show the flow of our intrusion detection system.

Components

DataSet:

The dataset used in this project is UNSW_NB_15 for both the training and testing. The UNSWNB_15 dataset contains packet data pertaining to ten distinct types of attacks. A total of 45 attributes, including the attack category, are present in the dataset, which has about 82332 records that are separated for training and testing purposes.

PreProcessing:

The dataset contains numerical and categorical attributes , the numerical values are standardized using standard scaler and categorical values are encoded using label encoder so that the machine leaning algorithm process it easily.

Feature Selection:

All attributes don't contribute to the target variable so we have performed feature selection to enhance the accurateness.

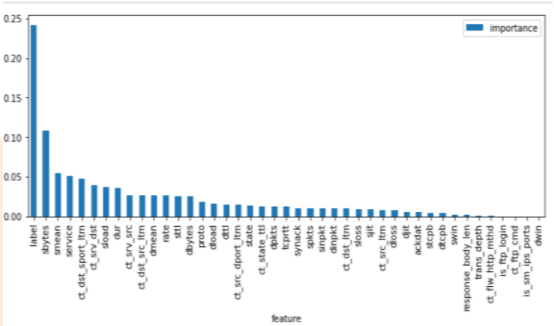

Fig.2: The graph shows the features that has contribution to target

Training the model :

The dataset is partitioned into two portions: 80% for training and 20% for testing.The algorithms used to build the model are

1) Decision Tree : A Decision tree has a structure similar to flowchart like tree used for classification. In this technique each leaf node has a class label, each internal node is a test on a feature, and the branches reflect the conjunctions of features that determine those class labels,top most node is root Classification rules are represented by the routes from root to leaf. To choose the attributes with the highest degree of accuracy, metrics like information gain (entropy), Gain ratio, and Gini index approach are utilised.

2) Logistic Regression : The Logistic Regression algorithm is used to estimate the likelihood of a specific result based on the parametric values supplied. An S - shaped sigmoid function that predicts maximum values is produced through logistic regression. It is used to predict categorical attribute using some of the independent attributes values. Therefore, the outcome or target attribute must be either a discrete value or a categorical value.

3) Naïve Bayes : The naive Bayes classifier makes use of the bayes theorem. In this algorithm it calculates the probabilities of the event based on the given event probability is true. It assumes that features are independent and contributing equally to the outcome. The posterior probability for each class is then computed using the Naive Bayesian equation. The result or outcome of the forecast is the category with the maximum posterior probability.

4) Gradient Booster Classsification : In this technique the objective is to minimise a loss function, the functional gradient method known as Gradient Boosting continually chooses a function that points in the direction of a weak hypothesis or a negative gradient. A powerful predictive model is created using the gradient boosting classifier by combining many weak learning models.

5) SVM : The SVM is one of the simplest supervised learning models for classification. The input is subsequently indirectly mapped to a higher dimensional feature space. A hyperplane is a subspace that has 1 less dimension than its surrounding space. When categorising complicated data, as is the situation with the majority of real-world statistics, it is particularly effective in situations with a large number of

characteristics. Finding a hyperplane of data that is appropriate for the dataset is the SVM's ultimate objective.

User Interface for Testing:

A Graphical user interface is developed with the help of the python Django Web Framework which consists of various inputs that have to be provided by the user. The model saved in the developing phase is deployed onto the GUI. The inputs given by the user are used by the model in predicting intrusions.

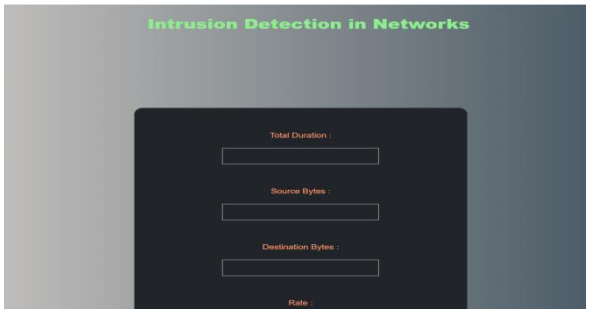

Fig.3: The Frontend interface for user testing

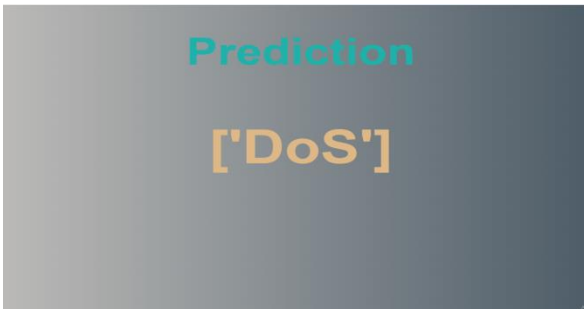

Fig.4: The predicted output

Results:

Table-1

Model	Accuracy
Decision Tree	88.66
Logistic Regression	84.63
Naïve Bayes	82.64
Gradient Booster Classification	89.88
SVM	87.62

IV. CONCLUSION:

We have used machine learning based approach, and were able to identify the intrusions and find the type of attack.The models will work with good accuracy on generalized data and can identify normal connections also. We have worked on the UNSW_NB15 dataset taken from kaggle, which comprises of 10 different types of attacks.We have build the 5 ML models using different algorithms and deployed the best on user interface . The project can be deployed at strategic points on network.The users can also test using the GUI.

V. REFERENCES:

[1] Lirim Ashiku,Cihan Dagli Network Intrusion Detection System using Deep Learning (10 June 2021.)
[2] Wang Z: "Deep learning-based intrusion detection with adversaries". (IEEE Access. 2018;6:38367–384).
[3] Dong G, Liu; H. Feature engineering for machine learning and data analytics; Boca Raton: CRC Press; 2018.
[4] Satish Kumar; Sunanda Gupta; Sakshi Arora.Research; Trends in Network-Based Intrusion Detection Systems: A Review(ISSN: 2169-3536).
[5] Ahmed A Mohammed Younis Thanoun,Ahmed Aleesa.Deep-Intrusion detection system with enhanced unsw-nb15 dataset(February 2021)
[6] Kasongo SM,Sun Y. "A deep gated recurrent unit based model for wireless intrusion detection system." (Cakovec: ICT Express; 2020).
[7] Patrick Vanin Eoin O'ConnellBrian Lee. A Study of Network Intrusion Detection Systems Using Artificial Intelligence/Machine Learning(18 November 2022)
[8] Towardsdatascience.com/building-an-intrusion-detectionsystem-using-deep-learning-b94883 32b321.

[9] KDD Cup 1999 dataset https://www.geeksforgeeks.org/intrusion-detection-systemusing-machine-learning-algorithms/.
[10] Axelsson, S (2000). "Intrusion Detection Systems: A Survey and Taxonom" (retrieved 21 May 2018)
[11] Zeeshan Ahmad, Cheah Wai Shiang, Johari Abdullah, Farhan Ahmad.Network intrusion detection system: A systematic
study of machine learning and deep learning approaches(16 October 2020)
[12] 'Newman, Robert' (2009-06-23). Computer Security: Protecting Digital Resources. Jones & Bartlett Learning. ISBN9780763759940.
[13] Sydney M. Kasongo & Yanxia Sun .Performance Analysis of Intrusion Detection Systems Using a Feature Selection
Method.(25 November 2020)
[14] Iqbal Gondal, Peter Vamplew & Joarder Kamruzzaman .Survey of intrusion detection systems: techniques, datasets and challenges
(17 July 2019)
[15] Harrington P. "Machine learning in action". (New York: Manning Publications Co.; 2012.)