International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 04 | Apr 2021 www.irjet.net p-ISSN: 2395-0072

Network Intrusion Detection System Using Deep Learning

Tejas Kadam[1], Akansh Shetty[2], Adarsh Kanekar[3], K. S. Suresh Babu[4]

[1-3]Department of Information Technology, Pillai College of Engineering, New Panvel, Navi Mumbai, Maharashtra, India - 410 206

[4]Assistant Professor, Department of Computer Engineering, Pillai College of Engineering, New Panvel, Navi Mumbai, Maharashtra, India - 410 206

Abstract - *Intrusion detection is a major research topic in the fight against external attacks on business and personal networks. A Network Intrusion Detection System (NIDS) software that keeps track of network and system activity. A network intrusion detection system (NIDS) is responsible for detecting malicious activity and unauthorized access to computers. The aim of designing NIDS is to protect the integrity and confidentiality of data. Internet security is a critical concern, so the goal of designing NIDS is to protect data integrity and confidentiality. The high volume, variety, and speed of data produced in the network has made conventional data analysis techniques for detecting attacks extremely difficult. This project will use a variety of machine learning algorithms, such as Artificial Neural Networks, Multi-class Logistic Regression and Logistic Regression to solve the problems. The algorithms will be correlated and contrasted.*

Key Words: **NIDS, Machine Learning, Deep Learning, Artificial Neural Networks, Logistic Regression, Multi-class Logistic Regression, NSL-KDD Datasets.**

1. INTRODUCTION

Intrusion detection systems are typically integrated into other security systems or applications and are designed to safeguard information systems. Firewalls and anti-malware tools alone are insufficient to safeguard an entire network. They serve as a minor component of a larger security scheme. The use of a full-fledged IDS as part of your protection framework is critical, because it's designed to function through your entire network in a variety of ways. An IDS uses its intelligence with the help of trends acquired with the help of algorithms and thus decide when an attack is taking place. Knowing the reach of an attack is also important for deciding your response and obligations to stakeholders who rely on your systems' protection. A Network Intrusion Detection System (NIDS) is typically installed or positioned at strategic points in the network to protect traffic from attack. It's usually applied to whole subnets, and it tries to fit any traffic going through with a database of documented attacks. It observes network traffic passing through the points on the network where it is installed in a passive manner. They can be relatively simple to secure and make intruders difficult to detect. This means that an intruder cannot realize the potential attack that is being detected by the NIDS. Since network-based intrusion detection systems analyze a large amount of network traffic, their accuracy can be poor. This means they can miss an attack or fail to detect anything in encrypted traffic on occasion. They can need more manual intervention from an administrator in some cases to ensure they're configured correctly.

2. LITERATURE SURVEY

2.1 Clustering approach based on k means for Intrusion Detection System over Big data

On large datasets, the traditional K-means algorithm is inefficient. Peng et al. [1] proposed an improved K-means detection method with mini batch to increase detection efficiency. They began by preprocessing data from the KDD99 dataset. The nominal features were converted to numerical forms, and the max-min approach was used to normalize each dimension of the features. The principal components analysis (PCA) algorithm was then used to reduce the dimensions. Finally, they used the K-means algorithm to cluster the samples, but they made two improvements to K-means. (1) To avoid being trapped in a local optimum, they altered the initialization process. (2) They used the mini-batch technique to cut down on the time it took to complete a task. The proposed method outperformed the standard K-means in terms of accuracy and performance.

2.2 Intrusion detection in enterprise systems by combining and clustering diverse monitor data

In this paper, Bohara et al. [2] proposed an unsupervised learning detection system. They used the VAST 2011 Mini Challenge 2 dataset to perform experiments and extract features from the host and network logs. They chose features based on the Pearson correlation coefficient because each function has different influences. The logs were then clustered using the K-means and DBSCAN algorithms. Clusters were linked to irregular behaviors by testing the salient cluster characteristics. Finally, they manually examined the irregular clusters to assess the attack forms.

2.3 Using Artificial Neural Networks in intrusion detection systems to computer networks

According to L. P. Dias,[3], the ever-increasing usage of computer networks has prompted some questions about availability, vulnerability, and security. Intrusion Detection Systems (IDS) are widely used by network administrators because they are considered important in maintaining network security. One potential drawback is that such systems are typically based on signature systems, making them highly reliant on modified databases and therefore ineffective against novel attacks (unknown attacks). This paper proposes an IDS scheme based on an artificial neural network (ANN) and the KDDCUP'99 dataset. Experimental results clearly show that the proposed system can reach an overall accuracy of 99.9% regarding the classification of predefined classes of Intrusion attacks which is a very satisfactory result when compared to traditional methods.

2.4 Network Intrusion Detection System Using Attack Behavior Classification

Al-Jarrah, O.; Dept. of Computer. Eng., et al. [4] By embedding the temporal actions of the attacks into a TDNN neural network structure, they were able to increase the recognition rate of network attacks using an intelligent device. Packet capture engine, preprocessor, pattern recognition, classification, and monitoring and warning module are the five modules that make up the proposed framework. Using a packet capture engine, this device captures packets in real time and sends them to a preprocessing stage through two pipes. The preprocessing stage extracts relevant features for port scan and host sweep attacks, stores them in a TDNN's tapped rows, and generates outputs that reflect potential attack behaviors in a predetermined number of packets. The pattern recognition neural networks use these outputs to identify the 23 attacks classified by the classifier network and generate attack alerts. The systems are evaluated in terms of recognition capability and throughput using DARPA data sets. The device detects all forms of attacks much faster than rule-based systems like SNORT, according to test results.

3. SYSTEM METHODOLOGY

The aim of the proposed system is to provide a statistical analysis of the network packets. The representation of each packet becomes simpler as the models serves different purposes. The frontend receives the results of each model. In terms of representation and accuracy, this system delivers a better result.

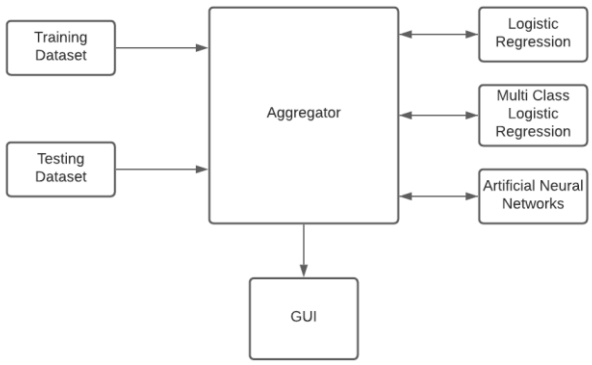

Figure - 1: System Architecture

3.1 NSL-KDD Dataset

NSL-KDD is an updated version of KDD cup99 data set. To overcome the shortcomings of the KDD99 dataset, the NSL-KDD was proposed. The records in the NSL-KDD Datasets were carefully selected based on the KDD99. The main advantage of NSL-KDD is that it removes duplicate and redundant records. Therefore, the experiments can be implemented on the whole dataset, and also from the different papers which are consistent and comparable. In the proposed system, the goal is to provide a statistical analysis of the packets in the network. The NSL-KDD training dataset is passed onto the aggregator. The function of the aggregator is to bridge the gap between dataset and prediction models. Each packet is passed into all 3 models. This dataset has 4 attack categories- U2R, R2L, DOS and Probe Attack. This dataset produces high accuracy because of its 41 features.

duration	num_failed_login	is_guest_login	dst_host_count
protocol_type	logged_in	count	dst_host_srv_count
Service	num_compromised	srv_count	dst_host_same_srv_rate
flag	root_shell	serror_rate	dst_host_diff_srv_rate
src_bytes	Su_attempted	srv_serror_rate	dst_host_same_src_port_rate
dst_bytes	num_root	rerror_rate	dst_host_srv_diff_host_rate
land	num_file_creations	srv_rerror_rate	dst_host_serror_rate

wrong_fragment	num_access_files	same_srv_rate	dst_host_srv_serror_rate
	num_shells		
urgent	num_outbound_cmds	diff_srv_rate	dst_host_rerror_rate
hot	is_host_login	srv_diff_host_rate	dst_host_srv_rerror_rate

Table - 1: Features of NSL-KDD Dataset

Dataset	Number of Records
NSL-KDD Training Dataset	125937
NSL-KDD Testing Dataset	22544

Table - 2: Training and Testing Dataset

3.2 Aggregator

The aggregator's job is to bridge the gap between the dataset and the prediction models. The aggregator receives the NSL-KDD training dataset or testing dataset. Each packet is sent to each of the three models. Aggregator is basically responsible to get the data, carry out the required process and then return it again to the output model for displaying the results. The data flow is simplified by Aggregator, making the project easier to debug.

3.3 Artificial Neural Network

Artificial Neural Network (ANN) mimics the way a human brain works. ANN undergoes a training phase where it learns to detect patterns in data visually, aurally or textually. During this supervised phase, the network compares its actual output produced with what it was meant to produce the desired output. Using the values gathered from the training phase, testing phase is carried out.

Artificial Neural Network gives out the exact attack type of the packet if it is abnormal. ANN succeeds to produce an accuracy of above 99% due to its thousands of neurons working together. ANN is used as the primary classification of attacks for the program.

3.4 Logistic Regression

Logistic Regression is statistical and its basic form uses a logistic function to model a binary dependent variable. In regression analysis Logistic Regression is estimating the parameters of a logistic model that is a form of binary regression and gives the output as 0 or 1(pass/fail). In the logistic model, the logarithm of odds for the value 1 is a linear combination of one or more independent variables. The independent variables can each be a binary variable or a continuous variable. Logistic regression model informs whether a packet is normal or abnormal with an accuracy of approximately 75%.

3.5 Multi-class Logistic Regression

Multi-class or Multinomial classification is the problem of classifying more than two classes on the given instance. Logistic regression gives output only in 0 and 1, that means it carries out binary classification. In multi-class logistic regression, we have more than two classes.
Multi-class Logistic Regression classifies if the packet is normal or gives out the specific attack type if abnormal. This model gives an accuracy of approximately 63%.

4. REQUIREMENT ANALYSIS

The implementation details are given in this section.

4.1 Hardware

Processor	Intel i7-8750H
HDD	1 TB
RAM	8 GB

Table - 3: Hardware specification

4.2 Software

Operating System	Kali Linux
Programming Language	NodeJS, JavaScript, HTML

Table - 4: Software specification

4.3 Dataset

Datasets are taken from NSL-KDD Training Dataset and NSL-KDD Testing Dataset. There is no duplication of records in NSL-KDD hence it contains definite set of records.

5. RESULT

In this project paper, we have worked upon different algorithms and obtained the statistical data for displaying the final result. Our project successfully displayed the type of attack and distinguished between a normal packet and that of the intrusion packet. We have used different technologies like NodeJS, Html and CSS to get the desired output.

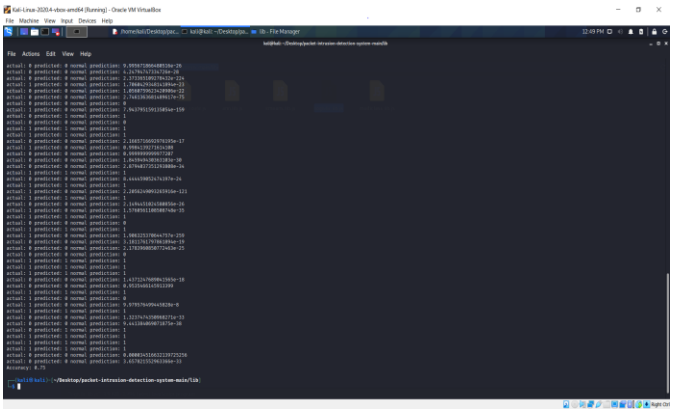

Figure - 2: Logistic Regression Output

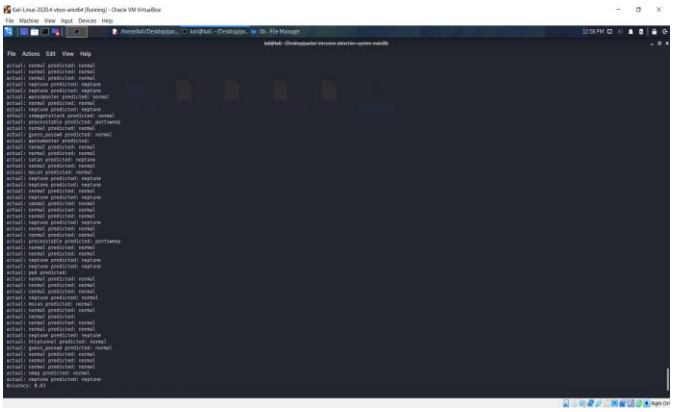

Figure - 3: Multi-class Logistic Regression Output

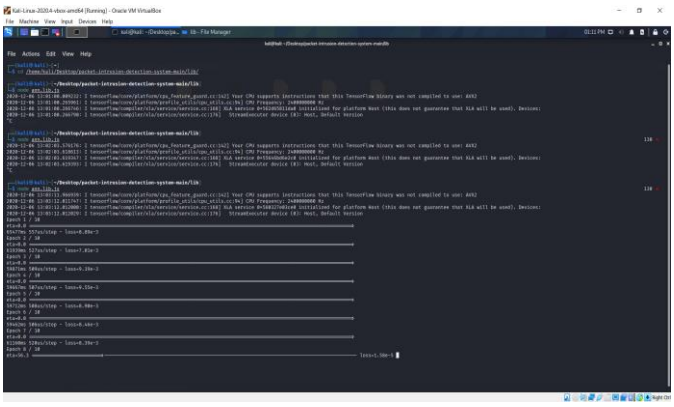

Figure - 4: Artificial Neural Network Output

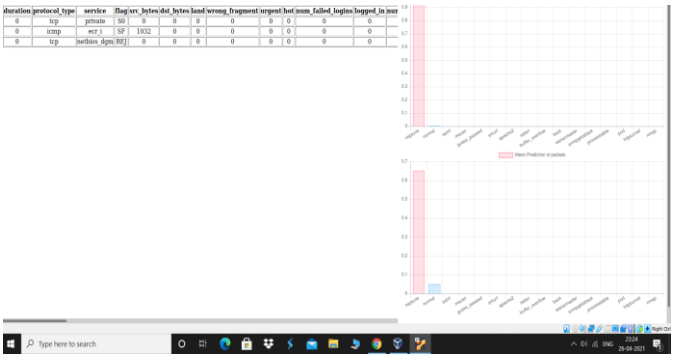

Figure - 5: GUI

6. CONCLUSION

The analysis of various recommendation techniques is presented in this paper. We may conclude that using a hybrid approach can have better accuracy and output outcomes than using different strategies such as Artificial Neural Network, Logistic Regression and Multi-class Logistic Regression. This paper presents a comparison of the different techniques listed above. Assigning different models to classify different terms helps to simplify the output. The functionality of each model is decided according to the output it provides and the accuracy it has. This project helps to understand the behavior of the packets and simulates a live network. Considering the future scope of this project, it can be made live by capturing some real-time packets in the network and detecting new attacks with some advanced algorithm techniques.

ACKNOWLEDGEMENT

We would like to take this opportunity to express our sincere gratitude to our project guide Prof. K. S. Suresh Babu for his constant guidance and support. The vision towards this project would not have been clear without inputs and guidance. The knowledge and experience of our guide has helped us a lot. Special thanks to our HOD Dr. Satishkumar Varma for his valuable feedback and providing us the lab facilities to work on this project. We would also like to thank our principal Dr. Sandeep Joshi and Pillai College of Engineering for giving us an opportunity to learn through this project.

REFERENCES

[1] Peng, K.; Leung, V.C.; Huang, Q. Clustering approach based on mini batch k means for intrusion detection system over big data. IEEE Access 2018, 6, 11897–11906.

[2] Bohara, A.; Thakore, U; Sanders, W.H. Intrusion detection in enterprise systems by combining and clustering diverse monitor data. In Proceedings of the Symposium and Bootcamp on the Science of Security, Pittsburgh, PA, USA, 19–21 April 2016; pp. 7–16.

[3] L. P. Dias, I. I. F. Cerqueira, K. D. R. Assis and R. C. Almeida, "Using artificial neural networks in intrusion detection systems to computer networks," 2017 9th Computer Science and Electronic Engineering (CEEC), Colchester, 2017, pp. 145-150, doi: 10.1109/CEEC.2017.8101615.

[4] Al-Jarrah, O.; Dept. of Computer. Eng., Jordan Univ. of Sci. & Technol., Irbid, Jordan; Arafat, A., Network Intrusion Detection System Using Attack Behaviour Classification, 5th International Conference on Information and Communication Systems (ICICS), 2014.

[5] Using Jpcap API to Monitor, Analyze, and Report Network Traffic for DDoS Attacks June 2014, Conference: 2014 14th International Conference on Computational Science and Its Applications (ICCSA).

[6] Network Intrusion Detection System Based on Machine Learning Algorithms, International Journal of Computer Science & Information Technology (IJCSIT), Vol 2, No 6, December 2010.

[7] Machine Learning and Deep Learning Methods for Intrusion Detection Systems: A Survey, Hongyu Liu and Bo Lang, State Key Laboratory of Software Development Environment, Beihang University, Beijing 100191, China, 2019.