Proposed Machine Learning Based Classification for Network Attacks

Sivaprasad Abirami [*] ***, S.Palanikumar,***
Department of Information Technology, Noorul Islam Centre for Higher Education,Thuckalay, Kumaracoil, Tamilnadu, India,
****Corresponding Author***
Email Id: E-mail:abirami.sivaprasad@sakec.ac.in

ABSTRACT

The proposed approach explained in this paper helps in capturing network packets and determines which attack was launched in a given network. Machine learning algorithms are used to distinguish between harmful and harmless packets. The machine learning based classifier is used to classify the incoming packets and identify the malicious and non malicious network packet based on the guideline available in the KDD dataset. The main aim of this project is to provide a proactive network attack detection system with the help of a machine learning based classifier. The system is trained under supervised learning with the attributes available in the KDD dataset.

Keywords*—Machine Learning, packet classification, Packet analysis, Data Mining, Network attacks, Security, KDD data set.*

INTRODUCTION

Machine Learning Classifier

The topic itself represents that the machine learns to classify, its true. The machine is trained with the help of KDD dataset to learn in a supervised way and then after learning the classifier classify the incoming packets based on the class label.

In this case, the classifier classify whether the incoming network packet is malicious or not with the help of machine learning classification algorithms.

Python

It's an open source tool and has build in data science package which helps to use the algorithms directly. Python also helps to do the comparative analysis of various classification algorithms with various parameters and can also visualize the same using different plots.

The visualization give a clear view of which classification algorithms perform in a better way and chose the algorithm for the implementation of the proposed system.

Intrusion Detection system [IDS]

The system detect the intrusion happens in the network by packets send by the attacker. In the proposed system, the live packets are captured and indentify the malicious packets.

KDD

KDD dataset is a well set platform or benchmark in the research of IDS techniques. It had 42 attributes initially, after preprocessing and converting every attributes to upper class 41 attributes were the result as refined. There are 4 classes U2R, R2L, Probe and DOS. There

exists a lists of known attacks in these 4 classes from which Denial of Service, Man in the Middle, Teardrop Attacks are selected.

Denial of Service

As name states it is meant to deny the service to the legitimate users on the network by repetitive pinging from perpetrator. Thereby, exhausting network resource available for user. In most of cases it shall be achieved by flooding the network to an extent.

Man in the Middle

Interestingly to achieve this attack there are two prominent ways:

1) **Via 2** devices only where server will act as receiver and when sender sends the information packet. Attacker can intercept and have access to their private exchanges.
2) **Via 3** Simple yet complex to achieve where equation is crystal clear ;We have 2 users exchanging data between them and 3rd device would be used to intercept between them.

Teardrop Attack

It is technically a type of DOS attack in the sense that it restricts users from having control access over the data. To achieve that what happens is that it will stop reassembly of packets causing them to overlap and hence corrupting the data.

SCOPE OF THE PROJECT

As per the previous research, the machine learning algorithms give a better result in classifying the network packets. The project implements various machine learning algorithm and the comparative analysis of the same. We are highly concerned about how we will use classification algorithms with highest accuracies to help us in determining how, where and which attacks are been applied on the network. Basically there is comparison of data packets captured via network with given dataset and definitions in KDD dataset. With help of random forest and decision tree we could quickly identify accurate results regarding where the packet captured is malicious or not. The system is tested with various attacks as mentioned in the KDD dataset. In this project tools like Eclipse, Weka, NetBeans, SQLYog, IPmessenger were used and entire project is developed in JAVA language. The IDS is considered to be an infrastructure which alerts the network user when there is unusual behavior. It is considered to be an alert based security system. The IDS is used to identify or detect the malicious packets entering the network. The basic functions of IDS are: Read the incoming packets and find malicious packets, Send alarm to user, protect the system from attacks and Monitor the network [1-2].The IDS can be implemented in two ways: Anomaly based IDS and Signature based IDS [3-6]. The signature based works by a concept of pattern matching, means the system will be pre configured with lots of patterns. When the pattern matches, the system produces an alarm when the malicious pattern is identified. Anomaly based on the other hand work with some parameters like protocols, devices, ports used etc to find the malicious packets by comparing the current activity with the defined normal behavior. In this type of IDS the possibility of false positive is more .The machine learning algorithms based on artificial intelligence gives a better results regarding anomaly based IDS and reduce false positive ratio.

PREVIOUS RESEARCH

Testing was done using 10-fold cross validation. The experimental result showed SVM having higher accuracy of 92.40 as compared to KNN which had accuracy of 71.28% in classifying quality of water [2]. The analysis is done with 2 important metrics Detection Rate

and False Alarm Rate. Random Tree Algorithm was used to test dataset in python. This study helped to understand to obtain higher accuracy DR should be high and FAR should be low [6]. An IDS system is constructed with the help of the Random forest classifier and identifies the normal and malicious packets. The author used four classes of attacks and used NSL_KDD dataset for these purpose.10cross validations were applied and the eexperimental results proved RF had higher accuracy then J48 [4]. An experiment is conducted by testing KDD dataset of IDS with various Machine Learning algorithms such as Randon Forest, Naïve Bayes, MLP, Decision Tree etc. None of the algorithm can handle all the attacks efficiently. Naïve Bayes execution was the best as compared to other algorithms [8]. Anurag Jain, Bhupendra Verma and J. L. Rana Selecting right algorithm with accordance to IDS. With the help off KDD-99 Dataset which is found to be improved during research was used and feature elimination and feature selections were used to reduce and get more relevant features directly. After going through and understanding these many algorithms for classifier selection. Model evaluation and discussion is done to get best TP & Worst TP rates [12]. Malwan Bahjat Abdulrazaq, Azarabidsalih used KDD dataset for comparative study of different algorithms. They used 60% for training and 40% for testing. Experimental results were based on following 3 algorithms J48,KNN,Naïve Bayes .The accuracy of Classification gets best result when Pob class is used with DT while R2l and U2R used with KNN and DOS used with NB. The Naïve Bayes classifier underperform and gives less accuracy but it is faster as compared to other algorithms, KNN is slower as it takes more time to built and test data [9]. A survey of various anomaly detection using machine learning algorithms are conducted by Shikha Agrawal. In the survey the anomaly detection is done in 3 different ways such as classification, clustering and Hybrid method. Although, there are many methods to store and transfer data safe even still there are loopholes in various methods. The machine leaning based system provide a better results and the pros and cons of all the methods are discussed [13]. The DOS attack was implemented and packets were captured during the attacks and accuracy was displayed with help of Support Vector Machine and Naïve Bayesian algorithms [15]. The test result done using python for comparing three algorithms with various parameters are given below.

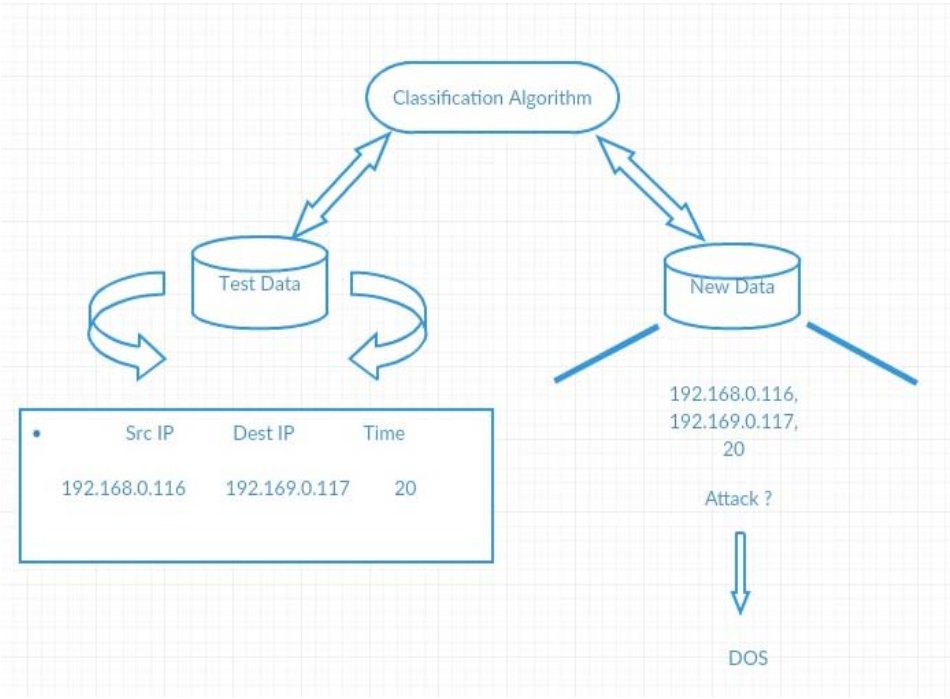

Fig 1. Flow of Classification

Algorithms	TP Rate	Precision	Accuracy	ROC	Root Mean Square Error
J48	0.931	0.989	93.10%	.969	.0763
Random Forest	0.938	0.991	93.77%	.996	.06782
Random tree	0.906	0.992	90.57%	.953	.0763
Decision table	0.924	0.944	92.44%	.984	.0903
MLP	0.919	0.978	91.90%	.990	.0813
Naïve Bayes	0.912	0.988	91.23%	.	.
Bayes Network	0.907	0.992	90.73%	.	.

Fig 2. Survey on Classification Algorithm

PROPOSED SYSTEM

The proposed system classify the packets with respect to existing IDS how we could have done better with classifying what attacks are happening in our system. How over the time we could train our system such that it could withstand such anomalies at first hand situations. Since we are into detecting which attack has occurred. We need to have a robust system that will help us to detect the following attacks from our KDDCUP99 dataset. We will use it as in training & testing dataset.

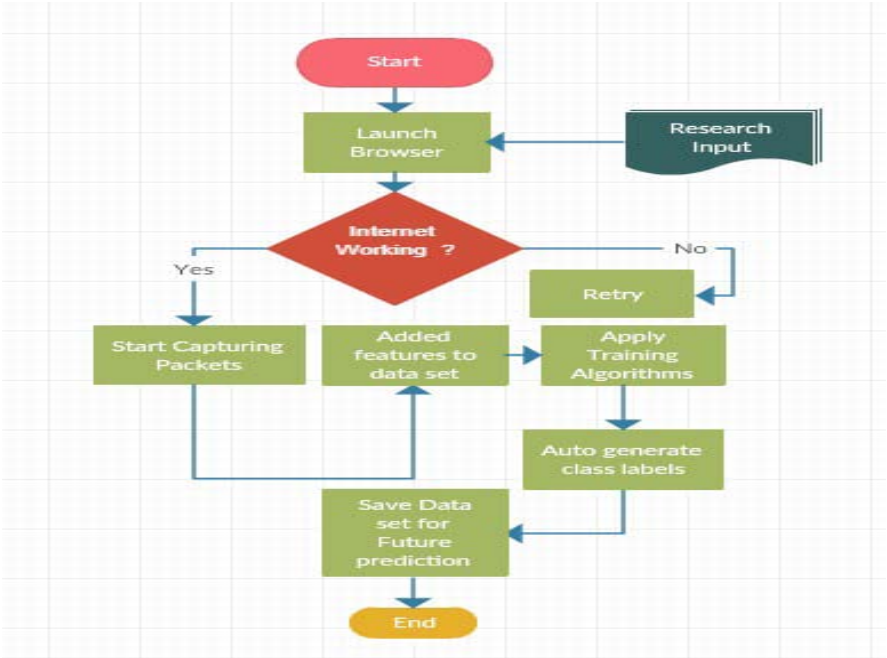

Fig. 2. Proposed System Flow Diagram

CONCLUSION

This paper talks about the study on KDD data set with various classification algorithms. The various attacks with the related attributes are discussed. The comparative study on various machine learning based classification algorithms is studied and the results are analyzed. Based on the analysis, Random Forest and J48 is providing a better result with the negligible

amount of incorrect classification and with a high accuracy and with an acceptable time to build.

REFERENCES

1) Yihua Liao, V.RaoVemuri "Using K-Nearest Neighbor Classifier for Intrusion Detection",Oct 2002.
2) AmriDanades,DeviePratama,DianAnggraini,DinyAnggriani "Comparison of Accuracy Level K-Nearest Neighbor Algorithm and Support Vector Machine Algorithm in Classification Water Quality Status" in 2016 IEEE 6th Conference on System Engineering and Technology October 3-4,2016 Bandung-Indonesia.
3) GaganjotKaur,AMit Chhabra "Improved J48 Classification Algorithm for Prediction of Diabetes" International Journal of Computer Application(0975-8887) Volume 98-No.22,July 2014.
4) Nabila Farnaaz and M.A.Jabbar "Random Forest Modeling for Network Intrusion Detection System" in Twelfth International Multi-Conference on Information Processing -2016(IMCIP-2016).
5) G.V.Nadiammai , M.Hemalatha "Effective approach toward Intrusion Detection System using data mining techniques" 22 May 2013.
6) PreetiAggarwal,Sudhir Kumar Sharma "Analysis of KDD Dataset Attribute-Class wise for Intrusion Detection" in 3rd International Conference on Recent Trends in Computing 2015(ICRTC-2015).
7) Okfalisa,Mustakim,IkbalGazalba,NurulGayatri Indah Reza "Comparative Analysis of K-Nearest Neighbor and Modified K-nearest Algorithm for Data Classification" in 2017 2nd International Conferences on Information Technology,Information Systems and Electrical Engineering(ICITISEE).
8) Mohammad Almseidin, MaenAlzubi, Szilvester Kovacs, Mouhammd Alkasassbeh "Evaluation of Machine Learning Algorithms for Intrusion Detection System.
9) Dr.Malwan Bahjat Abdulrazaq, Azarabidsalih "Combination of Multi Classification Algorithm for Intrusion Detection System" in International Journal of Scientific & Engineering Research,Volume 6,Issue 1, January-2015.
10) RajeshWankhede,VikrantChole,Shruti Kolte "A Review on Intrusion Detection System using Classification Technique" in International Journal of Advanced Computational Engineering and Networking ,ISSN:2320-2106,Voulme-3,Issue-12,Dec-2015.
11) David Ahmad Effendy,Kusrini Kusrini,Sudarmawan Sudarmawan "Classification of Intrusion Detection System (IDS) Based on Computer Network" in 2017 2nd International Conferences on Information Technology
12) AnuragJain, BhupendraVerma and J.L.Rana "Classifier Selection Models for Intrusion Detection System" in Informatics Engineering,an International Journal(IEIJ),Vol.4,No.1,March 2016.
13) ShikhaAgrawal, Jitendra Agrawal "Survey on Anomaly Detection using Data Mining Techniques " In 19th International Conference on Knowledge Based and Intelligent Information and Engineering Systems.
14) I.Dhanabal, Dr.S.P.Shantharajh "A Study on NSL-KDD Dataset for Intrusion Detection System Based on Classification Algorithm" in International Journal of Advanced Research in Computer and Communication Engineering Vol.4,Issue 6,June 2015.
15) Abirami Sivaprasad, Neha Ghawalkar, Srushti Hodge, Maitri Sanghavi, Vidhya Shinde "Machine Learning based Traffic Classification using Statistical Analysis".