Intrusion Detection System Using Euclidean Metrics and K-NN

Dr. Santosh Shivajirao. Lomte, Director, Radhai Mahavidyalaya College, Aurangabad (M.S.), India. drsantoshlomte@gmail.com

Saqr Mohammed H. Almansob, Department of Computer Science, BAM University, Aurangabad (M.S.), India. saqrmohammed2014@gmail.com

Abstract: As we observe these days, the Intrusion Detection System (IDS) is one of the main research problems in network security, and it is used to detect and parse the network traffic data. Intrusion Detection System (IDS) is, in fact, a software application which monitors the network or system activities and finds if any malicious operations occur. Tremendous growth and usage of the internet raise concerns about how to protect and communicate digital information in a safe manner. Nowadays, hackers use different types of attacks in order to retrieve valuable information. Many intrusion detection techniques, methods and algorithms help us to detect these attacks. The dataset of Intrusion Detection System can be classified into normal and abnormal traffic, which can be used to generate alert to detect threats. In this model, KDDcup99 dataset has been applied which consists of 494021 instances and 41 features. Furthermore, we see that there are four main types of attacks viz., Dos, Probe, R2L and 2LR. This article elucidates how the Intrusion Detection System uses Euclidean Metrics and K-NN and helps to derive the results of Detection Rate and False Positive Rate, along with feature selection and without feature selection. An Intrusion Detection System is an application used for monitoring the network and protecting it from the intruder. With the rapid progress in the internet-based technology new application areas for computer network have emerged. Hence, security is needed for the users to secure their system from the intruders. Firewall technique is one of the popular protection techniques, and it is used to protect the private network from the public network.

***Keywords* — Intrusion detection system (IDS), Euclidean Metrics, K-Nearest Neighbor, Feature Selection**

I. INTRODUCTION

Computer network security can be weak. The main reason for this is the weaknesses in making security policies, and other additional reasons are a weak computer system. Intrusion detection system has become one of the mechanisms of securing a computer network. Furthermore, detect the abuses of having the information from network. The performance is important but more significant is the environmental network. The detection of intruders, trespassers, insiders, man functionaries is made by, the hardware as well as software system in the intrusion detection system of the above traditional fire types [1]. Intrusion detection system is categorised on the characteristic parameter on the nature of their instructions these systems different on the ground of how they detected intruders they differ according to the function of their detection. The malfunction may be caused either by misuse or by anomalous use detection is essential to present such measurement [2].

II. RELATED WORK

Yihua liao et al [1] proposed K-Nearest neighbour (K-NN) as a classification for Intrusion Detection. The authors applied DARPA 1998 and BSM audit data in the experiment. They have noticed the K-NN algorithm classification is very effectively detecting intrusion attack. It archives a low false positive rate. Saeed M Algahtani et al. [2] proposed Decision Tree (J48), K-Nearest Neighbour (K-NN), Naïve Bayes and One R to present a comparative study of different classification techniques and cloud intrusion detection system. The authors used ISCX data set in the experiments with using 10-fold cross-validation. Andrey Ferriyan [3] proposed feature selection with a genetic algorithm to improve classification in a network intrusion detection system. The authors applied the NSL-KDD cup99 data set in their experiments. They have noticed the Random Forest gave the best results in classification rate. Mohammed Anber [4] proposed three classification algorithms which are, Diction Tree, K-Nearest neighbor (K-NN) and Random Forest to improve the

 DOI : 10.35291/2454-9150.2019.0152

accuracy in the intrusion detection system and to detection IPV6-based attack. Furthermore, comparative and analysis of three classifier algorithms. They have noticed that there is no single best algorithm that outperforms others in all measured metrics. The results obtained show the K-Nearest neighbour (K-NN) has the lowest false positive rate while Random Forest has the lowest false negative. Arief Rama Syerif [5] proposed a K-Nearest neighbor (K-NN) and Particle Swarm Optimization (PSO) Algorithm based intrusion detection mechanism to improving the accuracy. The authors used KDDCup99 data set in the experiment so, the results obtained show that increased up to 2% of accuracy by K-Nearest neighbor (K-NN). Yihua Liao, V rao et al [6] proposed a K-Nearest neighbor (K-NN) classification to intrusion detection. The authors applied 1998 DARPA BSM audit data. The results obtained show that the K-NN classifier can effectively detect an intrusion attack. Furthermore, it achieves a low false positive rate.Zhiyuqm et al [7] proposed a novel technique, Linear Discriminate Analysis (LDA) and difference distance map to reduce the heavy computational cost of anomaly intrusion detection system. Furthermore, used for the selection of significant features, the authors used DARPA1998 data set in their experimental. The results obtained show the techniques are able to transform high dimensional feature into low dimensional. M.A. Mithm Aravind et al [8] proposed an ensemble classifier to classify the data. Furthermore, classify five attack types namely, Dos, Normal, U2R, probe and R2L.Introduce Intrusion Detection System which detects the family of attack in the dataset. The data which is used in the experiments is UNSW.NB15data set.

III. PROPOSED WORK

The proposed model consists of five steps which are: the first step involves collecting dataset, the second step is the pre-processing technique which normalization of the dataset with certain data. Step number three includes feature selection with Euclidean distance for ranking the features; the instance of data which is applied in the model is 494021 instances. Step four is the classification technique and the last step evaluates the results.

The proposed by K-NN algorithm and Euclidean distance, in order to detect the different types of attack in a dataset. In this model, KDDcup99 dataset has been applied which consists of 494021 instances and 41 features.

A.Dataset:

In this model KDDcup99, the dataset has been applied which contains 494021 instance and 41 features. Furthermore, there are four main attacks type which are Dos, Probe, R2L and 2LR.

B. Pre-processing:

The dataset is uncompleted and contains noise in it. For this reason, the Pre-processing step is very important to enhance the quality of data. In this work some steps have been taken to Pre-processing KDDCup99 as follows

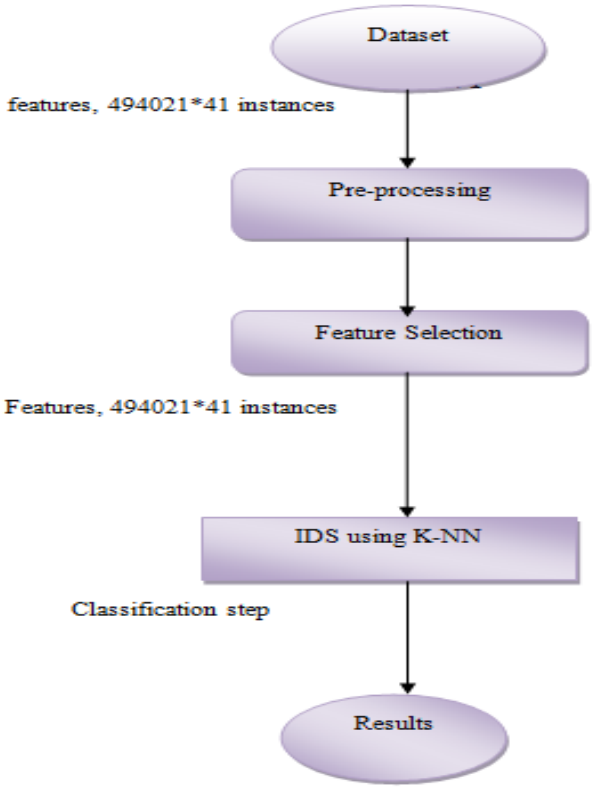

Figure 1 proposed model of Feature selection and K-NN algorithm.

i- Pre-process the input data into numerical data using some calculation

ii- Normalization of data finally to be numeric data whose number is between 0 and 1

C. FEATURE SELECTION:

There are 41 features in KDD cup99 dataset. Therefore, these are calculated according to Euclidean distance and put a minimum value as 570. Train data must be represented by a number of observations 41 futures and class label) numeric matrix. The number of selected feature finally to classify is eight from 41 which has the highest information, create sub-datasets with sub-selected features for training and testing, so all the values are treated as scores given to the features. The selected features are in the order Feat 5, Feat 24, Feat 30, Feat 6, Feat 23, Feat 25, Feat 34, and Feat 26. F = {F1, F2. F3, F4…….F41} and let any feature Fj = {X1, X2, X3, X4……Xni}, where j is the total value of features and N, is the number of instances.

IV. RESULTS AND DISCUSSION

The performance of the proposed approach is evaluated by using the KDD Cup99 dataset. The KDD Cup 99 is an audited set of the standard dataset which includes training

 DOI : 10.35291/2454-9150.2019.0152

and testing set. In the experiments, original KDD Cup99 dataset has been applied without redundancies in it. The Euclidean distance used to a dataset containing 494021 instances. Euclidean Metrics is the normal distance between two points that can be measured using a ruler. The Euclidean distance between the point's p and q is the length of the straight segment between them. In the karate coordinates, if p = (p1, p2, ..., Pn) and q = (q1, q2, ..., qn) are two points

Performance Measures

The proposed model is evaluated by used the following evolution detection Rate, False positive Rate and Accuracy Rate.

DR = DR/attack; FAR = FAR/normal; AR = AR/normal; .Detection Rate (DR) = Number of attacks correctly classified as attack/Total number of attacks in the dataset. False Positive Rate (FAR) = number of normal events classified as attack/Total number of normal events in the data.

Table 1: The DR and FAR of both the models

	K-NN classifier with proposed model 8 features		K-NN classifier without feature Selection41 features	
	Detection Rate	False Positive Rate	Detection Rate	False Positive Rate
DOS	95.06	8.02	82.16	9.87
Normal	97.1	5.04	94.05	16.03
U2R	32.12	3.11	10.03	0.8
R2L	28.14	1.02	13.02	0.36
Probe	89.08	0.6	80.17	0.88

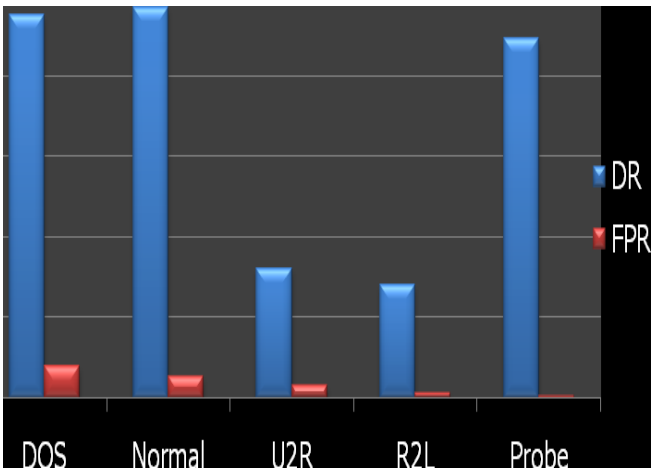

Figure 2 Performance results of Detection Rate and False Positive Rate with feature Selection 8 features.

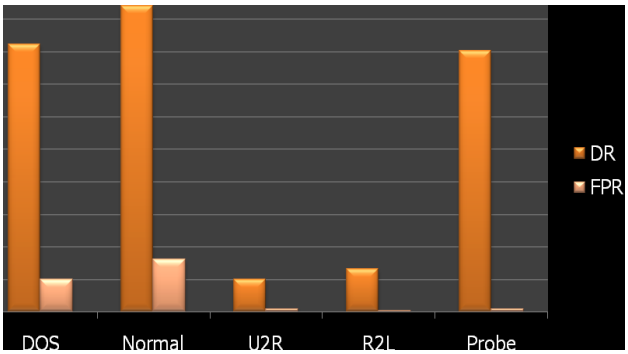

Figure 3 Performance results of Detection Rate and False Positive Rate without feature Selection 41 feature

V. CONCLUSION

We conclude the research paper, which is proposed by K-NN algorithm and Euclidean distance, in order to detect the different types of attack in a dataset. In the experiments conducted, the original KDD Cup99 dataset has been applied. Also, the Euclidean distance is used to a dataset containing 494021 instances. In order to classify, the number of selected features being used is 8 (which is the minimum value to obtain a score of 570), and the model is constructed using K-NN algorithm for classification purpose. The performance result of Detection Rate and False Positive Rate having a feature Selection of 8 is better to the result of Detection Rate and False Positive Rate without the feature selection of 41 features.

REFERENCES

[1] Arief Rama, S. Windu Gata" INTRUSION DETECTION SYSTEM USING HYBRID BINARY PSO AND K-NEAREST NEIGHBORHOOD ALGORITHM"2017 International Conference on Information & Communication Technology and System (ICTS) 978-1-5386-2827-0/17/$31.00 ©2017 IEEE

[2] M.Govindarajan, Rlvl.Chandrasekaran"Intrusion Detection Using k-nearest neighbor"978-1-4244-4787-9/09/$25.00 ©2009 IEEE

[3] Yihua liao, V. Rao Vemuri" Use of K-Nearest Neighbour classifier for intrusion detection" Computers and Security Volume 21 Issue 5, October, 2002

[4] Saeed M Algahtani"A comparative analysis of different classification techniques for cloud intrusion detection systems' alerts and fuzzy classifiers" Computing Conference, 2017, DOI: 10.1109/SAI.2017.8252132, IEEE 11 January 2018.G. R. Faulhaber, "Design of service systems with priority reservation," in *Conf. Rec. 1995 IJREAM Int. Conf. Communications*, pp. 3–8.

[5] Andrey Ferriyan"Feature selection using genetic algorithm to improve classification in network intrusion detection system" Creation and Intelligent Computing

 DOI : 10.35291/2454-9150.2019.0152

(IES-KCIC), 2017 International Electronics Symposium, DOI: 10.1109/KCIC.2017.8228458, Surabaya, Indonesia IEEE, 21 December 2017

[6] Mohammed Anber"Comparative performance analysis of classification algorithms for intrusion detection system"Privacy, Security and Trust (PST), 2016 14th Annual, DOI: 10.1109/PST.2016 IEEE.7906975, 24 April 2017.

[7] Arief Rama Syarif "Intrusion detection system using hybrid binary PSO and K-nearest neighbourhood algorithm" Information & Communication Technology and System (ICTS), 2017 11th International Conference. DOI: 10.1109/ICTS.2017.8265667IEEE 23 January 2018.

[8] J. Williams, "Narrow-band analyzer (Thesis or Dissertation style),"PhD dissertation, Dept. Elect. Eng., Harvard Univ., Cambridge, MA, 1993.

[9] Zhiyuan Tan" Network Intrusion Detection based on LDA for payload feature selection" GLOBECOM Workshops (GC Wkshps), 2010 IEEE, Miami, FL, USA.

[10] M.A. Mithun Aravind" **Design of an intrusion detection system based on distance feature using ensemble classifier**" Signal Processing, Communication and Networking (ICSCN), 2017 Fourth International Conference IEEE. 30 October 2017.

 DOI : 10.35291/2454-9150.2019.0152