Intelligent Network Intrusion Detection System using Data Mining Techniques

Amreen Sultana
M.Tech Scholar
MJCET
Hyderabad, Telangana
syedaamreen128@gmail.com

M.A.Jabbar
Associate Professor
MJCET
Hyderabad, Telangana
jabbar.meerja@gmail.com

***Abstract*—With the tremendous growth of usage of internet and development in web applications running on various platforms are becoming the major targets of attack. New threats are create everyday by individuals and organizations that attack network systems. Intrusion is a malicious, externally induced operational fault. Intrusion is used as a key to compromise the integrity, availability and confidentiality of a computer resource. Hence intrusion detection systems (IDS) are becoming a key part of system defence, to detect anomalies and attacks in the network. Data mining based IDS can effectively identify intrusions. Average one dependence estimators (AODE) is one of the recent enhancements of naïve bayes algorithm. AODE solves the problem of independence by averaging all models generated by traditional one dependence estimator and is well suited for incremental learning. In this paper, we propose intelligent network intrusion detection system using AODE algorithm for the detection of different types of attacks. In order to evaluate the performance of our proposed system, we conducted experiments on NSL-KDD data set. Empirical results show that proposed model based on AODE is efficient with low FAR and high DR.**

Keywords—Intrusion detection,data mining,AODE algorithm,NSL-KDD data set.

I. INTRODUCTION

The intrusion detection system (IDS) deals with large amount of data, and plays an important role in detecting various kinds of attacks. Intrusion detection system (IDS) can be considered as classification problem. Intrusion can be defined as malicious, externally induced and operational fault [1]. Intrusion is a key to compromise availability, integrity and confidentiality of a computer system [2]. Hence intrusion detection systems are becoming a key part of system defence, to detect anomalies and attacks in the network. Data mining is the process of extracting useful information from huge repositories [3]. Intrusion detection system (IDS) effectively identify these information and predicts the results that can be used in future. In data mining, classification is one of the most important techniques applied to intrusion detection. In this paper ,we proposed intelligent network IDS based on AODE classifier. AODE is an extension of naïve bayes classifier,which relies on conditional independence assumption. AODE enhances the attack detection accuracy. AODE estimate whether network traffic is normal or abnormal. We performed our experiments on the NSL KDD Data set [4]. NSL KDD Data set is mostly used for testing intrusion detection system (IDS). This paper is organized as follows. In section 2, we review various techniques that exist to classify network traffic. Our proposed approach is discussed in section 3. In section 4, we present our experimental results and analyze our findings. Finally, we conclude in section 5.

II. RELATED WORK

In the literature, several data mining techniques are applied for intrusion detection system (IDS). Each technique has its advantages and shortcomings. Performance of each model varies in terms of DR, FAR and accuracy.

Arif Jamal Malik et.al, proposed "Network intrusion detection using hybrid binary PSO and random forest (RF)". Binary PSO is used to select more appropriate features for classification. Random forest (RF) algorithm is used as a classifier [5]. Authors compared their approach with seven classification approaches. PSO-RF approach achieved good accuracy compared to other classification approaches.

In 2012, P. Natesan and Barasubramanie have proposed "Multi stage filter using enhanced ada boost for network intrusion detection". Authors proposed a multistage filter for intrusion detection. Their method uses three stages to detect attacks. Performance of their approach is tested with KDD99 dataset. They used adaboost, Decision tree and Naïve Bayes classifiers [6].

Preeti Aggarwal and Sudhir Kumar Sharma proposed "Analysis of KDD dataset attributes for class wise intrusion detection" in 2015 [7]. Authors in their paper analyzed KDD data set with respect to four classes 1) Basic 2) Traffic 3) Content 4) Host. Analysis was done in weka tool on random tree algorithm.

In 2011, "A hybrid intelligent approach for network intrusion detection" was proposed by Mrutyun Jaya Panda et.al [8]. Authors investigated hybrid intelligent techniques using data filtering by supervised and unsupervised methods.

978-1-5090-2399-8/16/$31.00 ©2016 IEEE

III. RESEARCH METHOD

In this section, we first describe intrusion detection system (IDS) AODE algorithm and then discuss our proposed method for classification of intrusion detection system (IDS).

3.1 Intrusion detection system (IDS)

An intrusion detection system (IDS) continuously monitors the network for any suspicious or malicious activity and alarms the network administrator, if it detects any such kind of malicious activity. Intrusion detection system (IDS) are classified into host based and network based. Basic architecture of intrusion detection system (IDS) is shown in figure 1. The intrusion detection system (IDS) architecture contains five components [3]

1. Data gathering device (sensor)
2. Detector (Intrusion detection (ID) analysis engine)
3. Knowledge base (database)
4. Configuration device
5. Response component

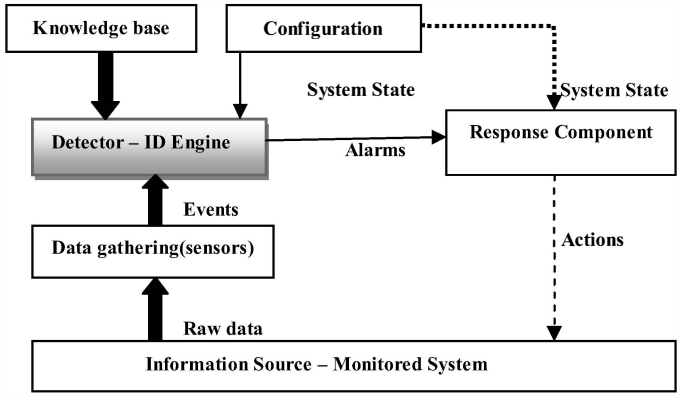

Fig 1.Architecture of intrusion detection system (IDS)[9]

Intrusion detection system (IDS) should have the following characteristics:1. Predictive performance 2. Time performance and 3. Fault tolerance.

Attacks in intrusion detection system (IDS) are classified into four types [10]:

1) DOS attack:

2) Probe attacks (Surveillance and scanning):

3) R2L and U2R attacks

Table 1 shows attack types.

TABLE 1:ATTACK TYPES

Attack Category	Examples
Dos	Back, neptune, smurf, teardrop
Probe	Satan, portsweep, ipsweep, nmap
R2L	guesspassword, imap, mutihop,spy
U2R	rootkit, bufferoverflow, loadmodule, perl.

3.2 AODE algorithm

Average one dependence estimators (AODE) is one of the recent enhancements of naïve bayes algorithm. AODE solves the problem of independence by averaging all models generated by traditional one dependence estimator and is well suited for incremental learning. AODE produces favourable results compared to traditional models. AODE classifier is widely applied to several problems like bio medical, intrusion detection, spam filtering. AODE algorithm is proposed to resolve the issues that were identified in ODE and SPODE.

Advantages of AODE algorithm are listed as follows [12]

- Probabilistic classification learning technique.
- Preferable for data sets where there is dependency among attributes.
- Low variance.
- Predicts class probabilities.
- Accurate and multiclass classifier.
- Useful for large data set.

3.3 Proposed method

Our proposed approach is shown in figure 2. Our network intrusion detection model applies on the AODE classifier.

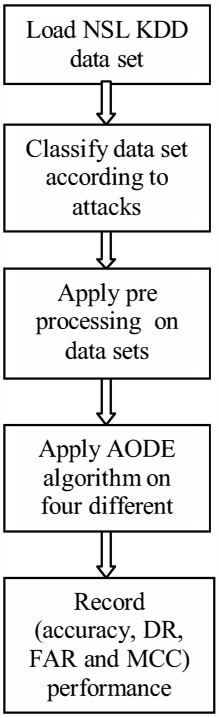

Fig 2. Our proposed approach

Our proposed algorithm is described below

Algorithm: Intelligent network intrusion detection system using data mining techniques.

Input: NSL-KDD Data set

Output: Classification of different types of attacks.

Step 1: Load NSL KDD data set.

Step 2: Apply preprocessing technique – discretization.

Step3: Clustered the datasets into four types.

Step 4: Partition each cluster into training and test sets.

Step 5: Data set is given to AODE algorithm for training.

Step 6: Test dataset is then fed to AODE for classification of attacks.

Step 7: Record the accuracy, detection rate (DR), false alarm rate (FAR), Matthews correlation coefficient (MCC).

In step 1 and 2, data set is loaded into the weka tool and preprocessing is done. NSL-KDD Data set is in ARF Format. In step 5 and 6, AODE algorithm is applied on data sets.10 cross validation is applied for classification. In step 7, accuracy and other metrics are calculated using confusion matrix.

IV. EXPERIMENTAL RESULTS

For our experimental analysis, we downloaded NSL KDD dataset from [13].We adopted the following pre processing techniques to run the experiment.

1) Replace missing values:

In weka, we used replace missing values filter to replace all missing feature values of NSL KDD data set. This filter replaces all missing values with the mean and mode of the training data.

2) Discretization:

Numeric attributes were discretized by discretization filter using unsupervised 10bin discretization.

4.1. Intrusion detection data set

We use NSL KDD data set, an enhanced version of KDD CUP 99 data set. NSL KDD data set consists of 42 attributes. The attribute labeled 42 in the NSL KDD data set is the class attributes which indicates whether a given instance is a normal or an attack. In this data set the attack can fall on any one of the following categories. 1. DOS attack 2. Probe attack 3. U2R attack 4. R2L attack. Refer [2] for NSL KDD data set.

Features classified for various attacks are listed in table 2.

TABLE 2: FEATURES SELECTED FOR VARIOUS ATTACK

Attack Type	Feature Number
Dos	1,2,3,4,5,6,23,25,27,30,32,33,34,35,39
Probe	1,2,3,5,6,9,20,23,25,27,30,32
R2L and U2R	1,2,3,4,5,6,10 to19,21 to 23,25,27,30,32,35,39

4.2. Performance measure

We used accuracy, detection rate DR), false alarm rate (FAR) and matthews correlation coefficient (MCC) which are derived using confusion matrix.

TABLE 3: CONFUSION MATRIX

	Classified as Normal	Classified as Attack
Normal	TP	FP
Attack	FN	TN

Where

TN –Instances correctly predicted as non-attacks.

FN – Instances wrongly predicted as non-attacks.

FP –Instances wrongly predicted as attacks.

TP –Instances correctly predicted as attacks.

Accuracy = Number of samples correctly classified in test data / Total number of samples in test data

$$\text{Detection Rate (DR)} = \frac{TP}{(TP+FN)}$$

$$\text{False Alarm Rate (FAR)} = \frac{FP}{(FP+TN)}$$

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

4.3. Simulation results and discussion

We conducted all our experiments using weka tool [14]. The performance of our proposed model is shown in table 4 and for naive bayes shown in table 5.

TABLE 4: PERFORMANCE OF OUR MODEL

S.No	Attack Type	Accuracy	Detection Rate (DR)	False Alarm Rate (FAR)	Matthews Correlation Coefficient (MCC)
1	DOS	97.19	98.63	4.44	0.943
2	Probe	96.48	98.19	5.45	0.927
3	U2R and R2L	96.25	98.65	6.48	0.925

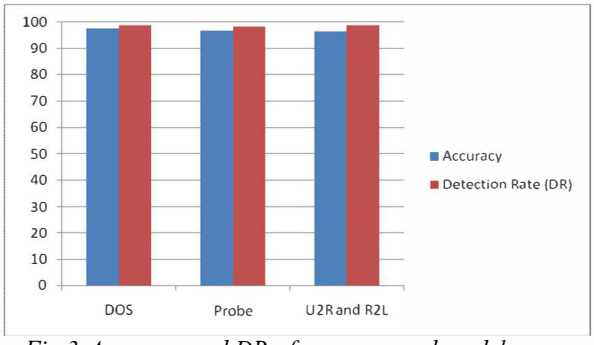

Fig 3:Accuracy and DR of our proposed model

TABLE 5: PERFORMANCE OF NAÏVE BAYES

S.No	Attack Type	Accuracy	Detection Rate (DR)	False Alarm Rate (FAR)	Matthews Correlation Coefficient (MCC)
1	DOS	89.90	94.85	15.72	0.72
2	Probe	90.48	96.07	15.87	0.812
3	U2R and R2L	90.47	95.60	15.37	0.811

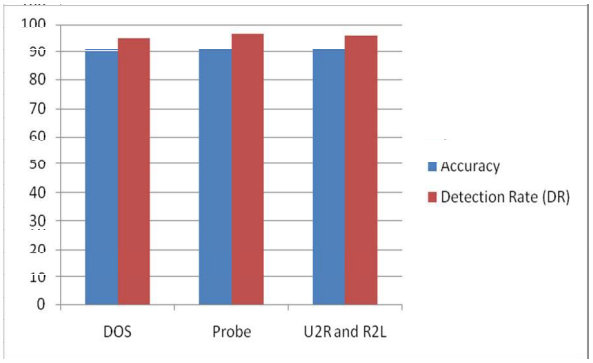

Fig 4: Accuracy and DR of naïve bayes model

It is evident from tables 4 and 5, figures 3 and 4 that our proposed model yielded high DR and low FAR to classify the attacks.

For DOS attack, our proposed model achieved an accuracy of 97.19%, which is 7% more than naive bayes algorithm. FAR recorded for naive bayes is 15.72 which is almost 11% more than our proposed model. For a good classifier to detect attacks it should have high DR and low FAR. For a probe attack FAR is recorded as 15.87% for naive bayes algorithm which is almost 10% more than our proposed model. For R2L and U2R ,FAR has been recorded as 15.37% which is almost 8% more than our proposed model. Table 6 and figure 5 shows accuracy of naive bayes and our proposed model for attack detection.

TABLE 6: ACCURACY OF NAIVE BAYES AND OUR PROPOSED ALGORITHM

Attack Type	Our Model	Naïve Bayes
DOS	97.19	89.90
Probe	96.48	90.48
U2R and R2L	96.25	95.47

Mathews correlation coefficient recorded by our model is high compared with naive bayes classifier.

Average accuracy recorded by our proposed approach is 96.64%, where as for naive bayes it is only 90.28%. Average value of MCC obtained by our approach is 0.93 and for naive bayes 0.80 only.

The experimental result shows that our proposed approach can achieve good accuracy, high DR with low FAR.

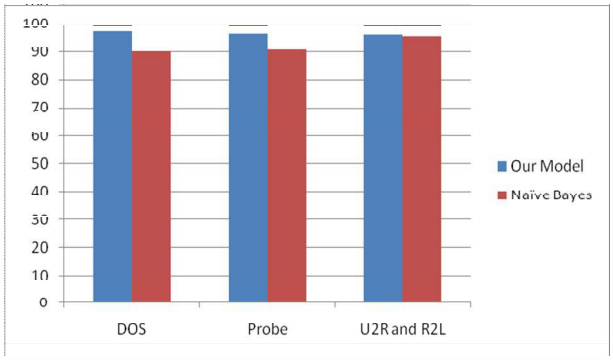

Figure 5:comparison of our approach with naïve bayes

V . CONCLUSION

In this paper, we applied the AODE algorithm to detect four types of attack like DOS, probe, U2R and R2L. 10 cross validation is applied for classification. The proposed approach is compared and evaluated using NSL KDD data set. Experimental result prove that accuracy, DR and MCC for four types of attacks are increased by our proposed method. Empirical results show that proposed model compared with naive bayes generates low false alarm rate and high detection rate. For future work, we will apply feature selection measure to further improve accuracy of the classifier.

References

[1] D. Powell and R. Stroud, "Conceptual model and architecture", Deliverable D2, project MAFTIA IST-19993-11583,IBM Zurich research laboratory research report R23377, NOV(2011).

[2] G V Nadiammai, "Effective approach towards intrusion detection sytem using data mining techniques", Egyptian Informatics Journal, 15, pp 37-50(2014).

[3] M.A. Jabbar, B.L. Deekshatulu, Priti Chandra, "Computational intelligence techniques for early diagnosis of heart disease", ICETECH, IEEE (2015).

[4] C.Ellcan,"Results of the KDD CUP 99 classifier learning".ACMSIGKDD, Explorations newsletter, 1(2):64(2000).

[5] Arif Jamal Malik, Waseem Shahzad, Farrukh Aslam Khan, "Netword ID using hybrid binary PSO and RF algorithm", Security and Communication Network, (2012).

[6] P.Natesan, P.Balasubramanie, "Multi Stage Filter Using Enhanced Adaboost for Network Intrusion Detection", International Journal of Network Security & Its Applications (IJNSA), Vol.4, No.3, May 2012.

[7] Preeti Aggarwal, Sudhir Kumar Sharma, "Analysis of KDD Dataset Attributes - Class wise For Intrusion Detection". 3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015).

[8] Mrutyunjaya panda et.al, "A hybrid intelligent approach for network intrusion detection", Procedia Engineering,45,pp 1-9(2012).

[9] Aleksander Lazarevic, Vipin Kumar, Jaideep Srivastava Intrusion Detection: A survey" pp 19-78 (2005).

[10] D. Marchette, Computer Intrusion Detection and Network Monitoring, A Statistical Viewpoint, New York, Springer, 2001.

[11] Zubair A. Baig, Abdulrhman S. Shaheen, and Radwan AbdelAal, "An AODE-based Intrusion Detection System for Computer Networks". Pp (28-35) IEE(2011).

[12]"Averaged one dependence estimator". Available on https://en. wikipedi a.org/wiki/, March 2016.

[13] "Nsl-kdd data set for network-based intrusion detection systems". Available on: http://nsl.cs.unb.ca/NSL-KDD/, November 2014

[14] I.H.Witten, Frank E, Hall M.A, "Data mining: practical machine learning tools and techniques, 3rd edition Burlington, M.A Morgan Kaufmann, 2011