A Supervised Intrusion Detection System Using Ensemble Classifiers

[1]Tincy Thomas, [2]Vidhya Vijayan, [3]Aby Abahai T, [4]Eldo P Elias

Department of Computer Science and Engineering
MA College of Engineering, Kothamangalam

***Abstract*: Classifying network connections as normal or anomalous is an important problem in the area of intrusion detection. Current algorithms and methods result in high false alarm rate. A high false alarm rate makes IDS ineffective. We propose a novel IDS using Ensemble classifier which combine eight week classifiers (Naive Bayes, Random Tree, ZeroR, OneR, Bayesian network, J48, SVM, KNN) to enhance attack detection. Idea behind ensemble classification is to exploit the strength of weak learning algorithms to obtain a robust and efficient classifier. Kyoto data set is loaded for pre-processing. After pre-processing, the data set is used for training using ensemble classifier. Average of probability combination rule is applied to test the class of a sample. During the testing phase, instances of the Kyoto data set are introduced to proposed ensemble classifier by hiding their class to which they belong. We build the ensemble classifier using WEKA machine learning tool. This ensemble classifier predicts the networks traffic data as normal or malicious.**

***Keywords:* Intrusion Detection System (IDS), Ensemble classifiers, Supervised learning.**

I. INTRODUCTION

One of the biggest threats faced by the modern era of computing is obviously the attacks through the networks. This leads to undependable and inconsistent states of systems causing far reaching consequences on various domains, making the whole networking sphere worthless. Intrusions are different types of attacks on targeted devices in order to affect their integrity, confidentiality and availability. Intrusions constitute one of the main issues in computer network security. Intrusion detection is the act of detecting the intrusions through various techniques. Intrusions can be found by tracing the anomalous network activities. An intrusion detection system (IDS) monitors network or system activities for malicious actions and produces reports to an authority. If finding the intrusion is challenging, avoiding a normal activity misjudged as intrusion will be more challenging.

Through malicious actions, hackers can have unauthorized access that compromises the integrity, the confidentiality, and the availability of a resource. Any type of networks starting from LANs to cloud computing, suffers from several types of attacks that can exploit systems vulnerabilities to reach a specific purpose, such as IP spoofing, Denial of Service (DoS), flooding, User to Root (U2R), Remote to Local (R2L), port scanning, probes, etc. The main purpose of Network Anomaly Detection is to create the perfect system that can detect and stop all intrusions and attacks to guarantee the reliability and the efficiency of the system, but this seems hard to reach with the continuous evolution of the hacking tools. The existing tools such as firewalls, antiviruses, and encryption are not efficient for all types of malwares and attacks. DoS attacks are too complex for firewalls to distinguish from normal traffic.

We require effective and efficient intrusion detection system (IDS) to detect attacks and anomalies in the networks. Ensemble learning is a new trend in AI and data mining, in which several weak learning algorithms are combined. Ensemble based learning perform well when data is huge and very scarce. Supervised learning is used for building the IDS using ensemble classifiers.

II. RELATED WORKS

The philosophy behind the ensemble classifier is that one base classifier compensates the error made by another classifier. But simply training the base classifier may not solve the desired problem as the base classifiers are uncorrelated. Base classifiers are the individual classifiers mainly used to construct the ensemble of classifiers. We can consider various weak learners such as support vector machine, neural network, and k-nearest neighbor to construct the ensemble classifier. The base learners are basically generated sequentially such as boosting and in parallel such as bagging. In boosting the weak learners are iteratively added to form a strong learner to yield a good result in prediction accuracy.

Ensemble methods have two components, according to its operation such as empirical study and theoretical study. In the empirical study the predictions made by the combinations of a set of classifiers are often more accurate than the prediction made by a single best classifier. But in the theoretical study, it is proved that a set of weak learners can be boosted to strong learners. Although strong learners are desirable, but it is difficult to get, while weak learners are easy to obtain. This led to the generation of strong learners by ensemble methods. Ensemble methods have two fundamental steps, that is, generating the base learners and combining them. Rather than getting good prediction accuracy, also the computational cost is not much larger than generating a single classifier.

In general, several approaches can be used for improving intrusion detection performance and one of these is RFAODE, an IDS built using two classification algorithms RF and AODE. Random forest is a tree-based ensemble classifier which combines bagging and random selection of features to construct a number of decision trees. It is highly accurate with low classification

error. Randomization is applied to select the best node. Average one dependency estimator (AODE) is an accurate and multiclass classifier which enhances the accuracy. AODE is efficient in detecting network traffic as normal or attack. AODE is useful for large data sets. Average One-Dependence Estimator (AODE) resolved the attribute dependency issue in Naive bayes classifier. Random Forest (RF) improves accuracy and reduces the error rate [1].

A Novel Ensemble Classifier for IDS using NB AND ADTREE [2] deals with a novel ensemble classifier based on naive Bayes and ADTree for intrusion detection. ADTree is a well-known supervised boosting decision tree algorithm. Naive bayes is a linear classifier and assumes that all features are independent. Naive Bayes will not perform well, where complex attribute dependencies are present. The ensemble combines ADTree and Naïve Bayes to improve classification accuracy of the detection system. NSL-KDD data set is used in the experiments. Outputs of four base classifiers ANN, SVM, kNN and decision trees are fused using three combination strategies: majority voting, Bayesian averaging and a belief measure [3].

The "majority voting rule "assigns a given input pattern to the majority class among the K outputs of the classifiers combined. The "average rule" assigns a given input pattern to the class with the maximum average posterior probability, the average being computed among the K classifiers (this rule can be applied if classifiers provide estimates of posterior probabilities, like multi-layer perceptron neural networks). The third fusion rule is based on the computation of a "belief" value for each data class given the set of outputs of the K classifiers. Belief values are based on estimates of the probabilities that a pattern assigned to a given data class actually belongs to that class or to other classes. These probabilities can be easily computed from the confusion matrix on the training set. The classification is then performed by assigning the input pattern to the data class with the maximum "belief" value.

Detection of distributed Denial of Service attacks using an ensemble of adaptive and hybrid Neuro-Fuzzy Systems [4] is achieved by combining ensemble of classifier outputs and Neyman Pearson cost minimization strategy, for final classification decision.

III. PROPOSED SYSTEM

An Intrusion Detection System based on ensemble classifier is proposed. An ensemble of classifiers is a set of classifiers where individual decisions are combined to classify new samples. In our approach, three base models of classification are used. The model learns the dataset and individually classifies the data. Ensemble methods or ensemble learning train multiple learners instead of a single learner and combine the result of different learners to yield a better result than individual weak learners. Hence, it is also called multiple classifier system. Ensemble methods always combines multiple hypotheses to form a better hypothesis. In other words, an ensemble is a technique to combine a large number of weak learners in an attempt to produce a strong learner. The term ensemble is usually reserved for methods that create multiple hypotheses by using the same base learner. The broader term of explicit multiple classifier systems also covers hybridization of hypotheses that are not induced by the same base learner.

A. Classification Model

Nine algorithms are taken for the classification using Kyoto dataset. The test option which used in all techniques is cross-validation with 10 folds.

1) Random Tree

Random Tree is a supervised Classifier; it is an ensemble learning algorithm that generates many individual learners. It employs a bagging idea to produce a random set of data for constructing a decision tree. Random trees have been introduced by Leo Breiman and Adele Cutler. The algorithm can deal with both classification and regression problems. Random trees are a collection of tree predictors that is called forest.

2) J48

J48 is a version of an earlier ID3 algorithm developed by J. Ross Quinlan. J48 is an open source java implementation of the C4.5algorithm in the weka data mining tool. The J48 Decision tree classifier follows the following simple algorithm. In order to classify a new item, it first needs to create a decision tree taken six different data sets based on the attribute values of the available training data. So, whenever it encounters a set of items (training set) it identifies the attribute that discriminates the various instances most clearly.

3) Naive bayes

Naive Bayes methods are a set of supervised learning algorithms based on applying Bayes‟ theorem with the "naive" assumption of independence between every pair of features.

4) Support Vector Machine(SVM)

Support Vector Machine (SVM) was originally designed for binary classification, in which the margin is large and that try to separate the instances of various classes. The margin is the minimum distance instances of different classes to the classification hyper plane. The larger margin minimizes the generalization error of the classifier. The support vectors are the data points that are closest to the separating hyperplane. These points are on the boundary of the slab. SVM is associated with analyzing the data and recognizing the patterns used both for classification and regression model.

5) k-Nearest Neighbour (k-NN)

k-NN is another supervised learning model similar to SVM, also extensively used in intrusion detection, with the difference that it will make its decisions based on a number of k neighbours. This means that the main difference with SVM is that it uses a

Euclidean distance rather than a hyper-plane, so if it cannot distinguish two categories using one plane, it will attempt again on a different one.

6) ZeroR

ZeroR is the simplest classification method which relies on the target and ignores all predictors. ZeroR classifier simply predicts the majority category (class). Although there is no predictability power in ZeroR, it is useful for determining a baseline performance as a benchmark for other classification methods.

7) OneR

OneR, short for "One Rule", is a simple, yet accurate, classification algorithm that generates one rule for each predictor in the data, then selects the rule with the smallest total error as its "one rule". To create a rule for a predictor, we construct a frequency table for each predictor against the target. It has been shown that OneR produces rules only slightly less accurate than state-of-the-art classification algorithms while producing rules that are simple for humans to interpret.

8) Bayes Net

A Bayesian network, Bayes network, belief network, Bayes(ian) model or probabilistic directed acyclic graphical model is a probabilistic graphical model (a type of statistical model) that represents a set of variables and their conditional dependencies via a directed acyclic graph (DAG). For example, a Bayesian network could represent the probabilistic relationships between diseases and symptoms. Given symptoms, the network can be used to compute the probabilities of the presence of various diseases.

Formally, Bayesian networks are DAGs whose nodes represent variables in the Bayesian sense: they may be observable quantities, latent variables, unknown parameters or hypotheses. Edges represent conditional dependencies; nodes that are not connected (there is no path from one of the variables to the other in the Bayesian network) represent variables that are conditionally independent of each other. Each node is associated with a probability function that takes, as input, a particular set of values for the node's parent variables, and gives (as output) the probability (or probability distribution, if applicable) of the variable represented by the node.

B. Ensemble Classifier

The major goal of ensemble methods is to combine the prediction of several models that is built with a learning algorithm to improve the robustness over a single model.

Basically, there are two types of ensemble learners: homogeneous learners and heterogeneous learners. Ensemble methods use a single base learning algorithm produces homogeneous learners, i.e., learners of the same type, leading to homogeneous ensembles. If learners are of different types that lead to heterogeneous ensembles, and use multiple learning algorithms. The generalization ability of an ensemble is often much stronger than that of base learners. Actually, ensemble methods are able to draw the interest mainly because they can boost weak learners to strong learners. Weak learners are even just slightly better than a random prediction but, strong learners can make very accurate predictions. Ensemble architecture combines n number of weak learners to form a strong learner. The weak learners are also called the base learners which are usually generated from base learning algorithms that can be decision tree, neural network or any kind of learning algorithms.

Ensemble methods are of two types such as averaging method and boosting method. Averaging method builds several models independently and then takes the average prediction to select the best one. Examples of these methods are bagging, random forest etc. In boosting method, the second model depends on the first one. Models are built sequentially. Its primary goal is to combine weak model to produce a powerful one, and it reduces the bias of combined model. Examples of these methods are AdaBoost, LogitBoost, etc.

Boosting algorithms may be considered stronger than bagging on noise-free data. However, there are some strong empirical indications that bagging algorithm is much more robust than boosting in the case of noisy settings. Boosting involves incrementally building an ensemble iteratively by training each new model instances to emphasize the training instances that previous models misclassified. In some cases, boosting has been shown to yield better accuracy than bagging, but it also tends to be more likely to over-fit the training instances.

By definition, boosting belongs to a family of algorithms that are able to convert the weak learners to strong learners. A weak learner is just slightly better than random guess, while a strong learner is very close to perfect performance. In boosting algorithm, models are built in a sequential manner by exploiting a classification algorithm to the reweighted versions of the training instances. It combines the performance of many weak learners to produce a powerful strong learner.

C. Proposed Ensemble Classifier

The individual week classifiers are combined to model an ensemble classifier which uses meta learning algorithm called Voting. Predictions made by the combinations of a set of classifiers are often more accurate than predictions made by the best single classifier. In theoretical work, it was proved that weak learners can be boosted to strong learners. Since strong learners are desirable but difficult to get, while weak learners are easy to obtain in real practice, this consequence forwards an actual direction of generating strong learners by ensemble methods. Generally, an ensemble is constructed in two steps; i.e. generating the base learners and then combining them. To get a good ensemble the base learners should be as accurate as possible. In general, the computational cost of constructing an ensemble is not much larger than creating a single learner. This is because when we want to use a single learner, we usually need to generate multiple versions of the learners for model selection. But the computational cost for combining base learners is small since most combination strategies are simple.

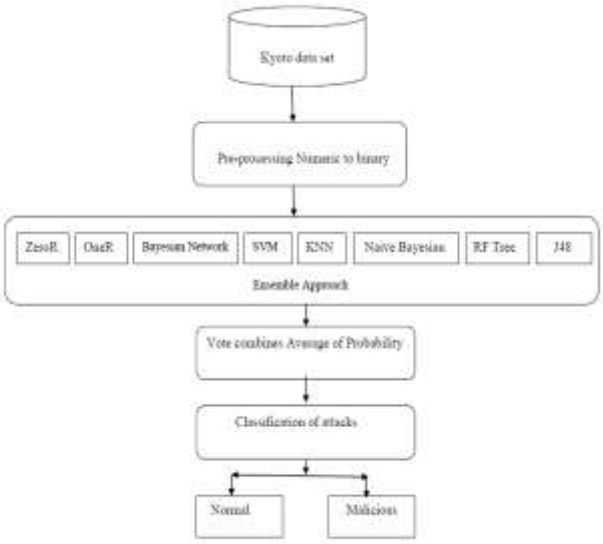

Fig.1. Simple Ensemble Architecture

Combined output is obtained by averaging the outputs of individual learners directly. The simple averaging gives the combined output H(x) as

$$H(x) = \frac{1}{T} \sum_{i=1}^{T} h_i(x) \quad (1)$$

In weighted averaging the combined output is obtained by averaging the output of individual learners with different weights. The weighted averaging gives the combined output H(x) as

$$H(x) = \sum_{i=1}^{T} w_i \, h_i(x) \quad (2)$$

IV. PERFORMANCE ANALYSIS

The performance evaluation can be carried out by comparing the performance metrics of nine classifiers OneR, Random Tree, J48, SVM, KNN, Bayesian Network, Naive Bayes Classifier and ZeroR. Ensemble classifier increases the performance and achieves 100% accuracy with reduced False Alarm Rate. The following are the metrics used for performance evaluation.

True Positive Rate (TPR) = TP / (TP + FN)
True Negative Rate (TNR) = TN / (FP + TN)
False Negative Rate (FNR) = FN / (FN + TP)
False Positive Rate (FPR) = FP / (FP + TN)
Accuracy = (TP + TN) / (TP + TN + FN + FP)
Positive Predictive Value (PPV) = TP / (TP + FP)
Negative Predictive Value (NPV) = TN / (TN + FN)
False Discovery Rate (FDR) = FP / (FP + TP) = 1 - PPV
F1 Score = 2TP / (2TP + FP + FN)

Table 1. Ensemble Classifier Detailed Result By Class

TPR	FPR	Prec.	Recall	F-Measure	MCC	ROC	PRC	Class
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Attack
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Normal
Weighted AVG.								
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	

Table 2. Ensemble Classifier Result

Description	Result	%
Time taken to build model	0.04s	
Total Time of Completion	0.002s	
Correctly Classified Instances	14	100%
Incorrectly Classified Instances	0	0%
Kappa statistic	1	
Mean absolute error	0.2496	
Root mean squared error	0.2733	
Relative absolute error		49.2969%
Root relative squared error		54.6517%
Total Number of Instances	14	

V. CONCLUSION

To discriminate the normal traffic and the attack traffic by using the Ensemble approach improves the detection accuracy with a less computational time and minimum cost compared to a single classifier. AdaBoost is an efficient false positive detection technique to minimize the false alarms. For the proposed model when we are using same dataset for training and testing, the accuracy percentage is high and error is less. Kyoto dataset is loaded for training and testing the classifier. But with different dataset for training and testing, the accuracy rate is comparatively less. Eight weak classifiers such as Support vector machine, ZeroR, OneR, Naive bayes, KNN, J48, Bayes net, Random Tree are combined and their performance accuracy is better than the individuals. We also concluded that with adding more number of learners, to the combination model, the detection accuracy increases and probability of misclassified instances reduces in each iteration.

ACKNOWLEDGMENT

We would like to thank our faculty, Mar Athanasius College of Engineering (MACE), APJ Abdul Kalam Technological University (KTU) for their support in doing our project.

REFERENCES

[1] MA Jabbar, Rajanikanth Aluvalub, Sai Satyanarayana Reddy, ―RFAODE: A Novel Ensemble Intrusion Detection System‖, Procedia Computer Science 115 (2017), 226–234.

[2] R. Singh, H. Kumar, R. Singla, An intrusion detection system using network traffic profiling and online sequential extreme learning machine, Expert Systems with Applications.

[3] N. Moustafa, J. Slay, The evaluation of network anomaly detection systems: Statistical analysis of the unsw-nb15 data set and the comparison with the kdd99 data set, Information Security Journal: A Global Perspective 25 (1-3) (2016) 18–31.

[4] M.A.Jabbar et.al,‖Cluster based Ensemble classification for Intrusion Detection System‖, ACM, ICMLC 2017, February 24-26, 2017, Singapore, Singapore, pp253-257.

[5] M.A.Jabbar et.al,‖A novel Intelligent Ensemble Classifier for Network Intrusion Detection System‖,SOCPAR Springer (2016).