Volume 5, Issue 4, April 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Proposed Approach for Intrusion Detection on KDD Dataset by Hybrid Classifiers

Maninder Singh, Sanjeev Rao
CSE Department, Chandigarh University, Gharuan
Chandigarh, India

Abstract-With the rapid increase of internet technology, the malicious activities on the network are also increasing. So the use of an efficient method is must to detect the intrusion. Security for all networks is becoming a big problem. In this paper we proposed hybrid approach of classifier with Adaptive boost with SVM RBF.

Keywords: IDS, NSL, CFS

I. INTRODUCTION

With the growing threat of attackers over the network the effective intrusion detection is needed to secure the information. Before applying the classifiers the dataset is needed to be collected. IDS aim to recognize unusual access or attacks to secure internal networks. More and more time is consumed as the database grows, and more and more false positives are generated, which affects the purpose of IDS. Intelligent IDS is a dynamic defensive system that is capable of adapting to dynamically changing traffic pattern and is present throughout the network rather than only at its boundaries, thus helping to catch all types of attacks. Different type of classifiers can be used to detect the intrusion on the network. In this paper, we used Weka tool for intrusion detection system (IDS) to compare the performance of classifiers.

One of the biggest challenges in network-based intrusion detection is the extensive amount of data collected from the network. Thus the existing approaches of intrusion detection have focused on the issues of feature selection or dimensionality reduction. Feature selection or reduction keeps the original features as such and select subset of features that predicts the target class variable with maximum classification accuracy. In this paper, the data mining algorithm different classifiers will be evaluated on the NSL KDD dataset.

Classifiers used in this paper are:

1. Nave Bayes- The naïve Bayes model is a heavily simplified Bayesian probability model. The naïve Bayes classifier operates on a strong independence assumption. This means that the probability of one attribute does not affect the probability of the other. Naive Bayes classifier is highly scalable, require a number of parameters linear in the number of variables in a learning problem.
2. Decision Stump- A decision stump is a machine learning model consisting of a one-level decision tree. That is, it is a decision tree with one internal node (the root) which is immediately connected to the terminal nodes.
3. Random Tree- Class for constructing a tree that considers K randomly chosen attributes at each node. Performs no pruning. Also has an option to allow estimation of class probabilities (or target mean in the regression case) based on a hold-out set.
4. OneR-Class for building and using a 1R classifier; in other words, uses the minimum-error attribute for prediction, discretizing numeric attributes.
5. MutiScheme-Class for selecting a classifier from among several using cross validation on the training data or the performance on the training data. Performance is measured based on percent correct (classification) or mean-squared error (regression).
6. AdaboostM1-Class for boosting a nominal class classifier using the Adaboost M1 method. Only nominal class problems can be tackled. Often dramatically improves performance, but sometimes overfits.

II. LITERATURE REVIEW

Mrutyunjaya Pandaa[1],In this paper, they investigated some novel hybrid intelligent decision technologies using data filtering by adding supervised or un-supervised methods along with a classifier to make intelligent decisions in order to detect network intrusions. They used a variant of KDDCup 1999 dataset, NSL-KDD to build the proposed IDS. In particular, they investigated the combination of Decision trees, principal component analysis. The performance comparison amongst different hybrid and combination of classifiers were made in order to understand their effectiveness in terms of various performance measures.

Richa Rawat[2],In this paper an overview is presented that deal specifically with ids using data mining techniques. Many data mining algorithm that has been proposed towards the enrichment of IDSs. Here we present decision tree and

svm(support vector machine) techniques that is proposed by researchers to detect intrusion in the network. But all of these data mining techniques are not satisfactory throughout. So here we are presenting boosting technique that detects better result than single classifier technique.

M. R. Yadav[3],In this paper, network intrusion detection is considered using supervised learning algorithm to classify attacks in the datasets. We consider both well-known KDD99 dataset. We evaluate our IDS in terms of detection speed, detection rate and false alarm rate. Fuzzy Genetic algorithm is able to classify both KDD99 dataset with high accuracy and low false alarm rate. The experiments illustrated detection rate of each attack. We can see that the fuzzy genetic algorithm can mostly distinguish behavior of each attack type in both KDD99 dataset and online dataset with low false negative rates.

Yiwu Zhejiang[4],This paper studies on methods taken in solving the existing network disorders in network intrusion detection. On the wing of the merits found in the Genetic Algorithm and Support Vector Machine, this paper tries to propose a new network intrusion detection model based on the combination of GA and SVM. Simulation results certified that, in comparison with the traditional network intrusion detection methods, the optimized Support Vector Machine model with a Genetic Algorithm proposed in this paper has better recognition of intrusion and a high accuracy, a low rate of losing or false alarms, and the results proved to be more efficient in execution of the algorithm for network intrusion detection.

Manju Khari[5],From the comparative analysis on the various machine learning techniques for the intrusion detection, it is concluded that the genetic algorithms are a reliable method for the detection of malicious intrusions. The comparison between various learning techniques will allow software professionals to find best machine learning technique to find clear, unambiguous knowledge about intrusion detection more effectively and efficiently.

Mohammad Sazzadul[6],In this progression, here they presented an Intrusion Detection System (IDS), by applying genetic algorithm (GA) to efficiently detect various types of network intrusions. To measure the fitness of a chromosome we used the standard deviation equation with distance. If we can use a better equation or heuristic in this detection process we believe the detection rate and process will improve a great extent, especially false positive rate will surely be much lower.

Amit arora[7],In this paper they proposed a model for intrusion detection, that suggest, for the detection of intrusion it is not necessary to perform the test on all the 41 features of NSL-KDD [2] data set. First by using feature selection the features are reduced to 33 features and further by removing the features, the biasing of learning algorithms towards the frequent and easily detectable records in the data set is reduced. And the suggested machine learning algorithm after selection process is Simple Cart for the intrusion detection that leads to improve the computer security alerts from computer security incidents using machine learning techniques.

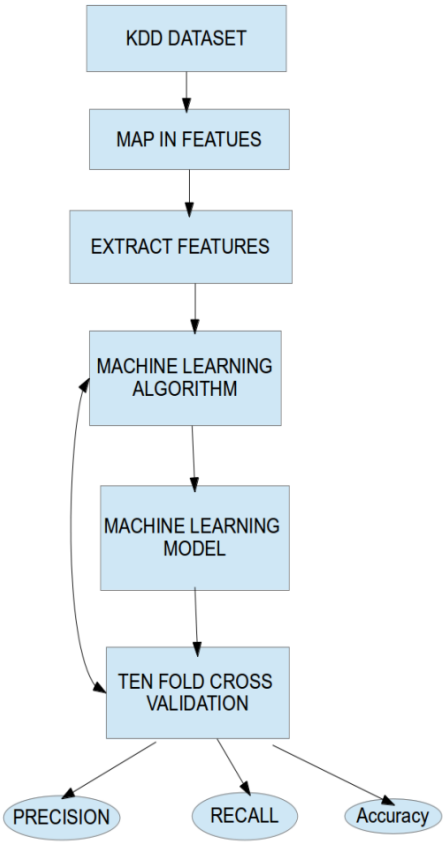
Dr. Saurabh Mukherjee[8],In this paper they proposed FVBRM model for feature selection and make its comparison with three feature selectors CFS, IG and GR. Experimental result illustrates feature subset identified by CFS has improved Naïve Bayes classification accuracy when compared to IG and GR. Although GR is an extended of IG , but in our analysis we have used both the techniques for feature selection and IG performs better than GR. FVBRM method shows much more improvement on classification accuracy with compared to CFS but takes more time. Future work will include customize of FVBRM feature selection method to improve the results for intrusion particularly for U2R attacks with reduced complexity and overheads.

Levent Koc[9],In this paper, they explained the need to apply data mining methods to network events to classify network attack events. We summarized the results of earlier studies and explored the earlier models on the performance improvement of the Naïve Bayes model in data mining and introduced the HNB model as a solution to the intrusion detection problem. We compared the performance of the Naïve Bayes and leading extended Naïve Bayes approaches with the new HNB approach as an intrusion detection system. The results of our experimental study, which uses the KDD'99 dataset, show that the Hidden Naïve Bayes multiclass classification model augmented with various discretization and feature selection methods exhibits better overall results in terms of detection accuracy, error rate and misclassification cost than the traditional Naïve Bayes model, the leading extended Naïve Bayes models and the KDD'99 winner. The results also indicate that our model significantly improves the detection of denial-of-service attacks compared with the other models. Considering its simplicity and its advantage over the Naïve Bayes model's conditional independence assumption, hidden Naïve Bayes is a promising model for datasets with dependent attributes, such as the KDD'99 intrusion detection dataset.

Amuthan Prabakar[10],Most of the Network Anomaly detection systems are designed based on availability of data instances. Many anomaly detection techniques have been specifically developed fir certain application domains, while others are more generic. In this paper, we present a cascaded algorithm using K–Means and C4.5algorithms for Supervised Anomaly Detection. The proposed algorithm is used for detect the anomalies presented in the supervised data set. We use KDD99 data set for conducting the experiments. Performance analysis is measured by using five measures, 1)

detection accuracy (or) True Positive Rate (TTR), 2)False Positive Rate (FPR), 3) Precision, 4) Total Accuracy (TA), and 5) F–Measures (FM). The proposed algorithm gives impressive detection accuracy in the experiment results.

III. PROPOSED METHOLOGY



• Firstly we select the Benchmark Dataset for Intrusion Detection System. After analysis the dataset we mapped the features.
• Experimental analysis produced the mapped features and we will select the main selection from the features. After gathering the selection features we introduce Machine Learning Algorithm which have many datasets for WEKA. Then we will apply Adaptive boost with SVM RBF Kernel.
• The process has been done by Intrusion Detection model. We analyzed the tenfold cross validation. From the analysis we get precision, Recall and Accuracy. Precision come out from valid data set. Recall comes from feedback system. Accuracy comes from Intrusion Detection Model.
• A feature selection based machine learning model. First feature selection is performed on NSL-KDD data set. The feature selection algorithms determine the important and essential features of the test data for intrusion detection, examining those features help in reliable detection of abnormal behavior.
• Feature selection reduces the number of members from the selected features that are 41,without affecting the effective indicators of potential behavior of attacks.

IV. CONCLUSION

So from all the analysis we have concluded that the different classifiers show different results on the KDD dataset. But all of these data mining techniques are not satisfactory enough. Many data mining algorithm that has been proposed towards the enrichment of IDSs. As from the comparison analysis of the difference classifiers, we can say that the random tree classifier is giving more accurate results as compared to other classifiers.
If we reduce the KDD dataset features by applying some technique, then the complexity might get reduced and the results can be more accurate. Also if we combine these classifiers with some other technique like the adaptive boost algorithm, the result can be more accurate than the results of a single classifier.

REFERENCES

[1] Mrutyunjaya Pandaa, Ajith Abrahamb "International Conference on Communication Technology and System Design" 2011.
[2] Richa Rawat , Anurag Jain "International Journal of Scientific & Engineering Research" Volume 4, Issue 7, July-2013.
[3] Miss. M. R. Yadav Prof. P. B. Kumbharkar "International Journal of Advanced Research in Computer Science and Software Engineering" Volume 4, Issue 4, April 2014.

[4] Yiwu Zhejiang, Study on Genetic Algorithm Optimization for Support Vector Machine in Network Intrusion Detection Xiaoqiang WANG.
[5] Manju Khari, Anjali Karar "International Journal of Advanced Research in Computer Science and Software Engineering" Volume 3, Issue 4, April 2013.
[6] Mohammad Sazzadul Hoque, Md. Abdul Mukit "International Journal of Network Security & Its Applications" Vol.4, No.2, March 2012.
[7] Karan Bajaj ,Amit arora, International Journal of Computer Applications, Volume 76– No.1, August 2013.
[8] Dr. Saurabh Mukherjee, Neelam Sharma,SciVerse ScienceDirect,2013.
[9] Levent Koc, Expert Systems with Applications,2012.
[10] Amuthan Prabakar, International Conference on Communication Technology and System Design 2011.

