Adaptive Approaches for IDS : A Review

Bharti Harode[1], Dr Anurag Jain[2]
Research Scholar[1], Professor[2]
Computer Science and Engineering Department
Radharaman Institute Of Science And Technology, Bhopal, Madhya Pradesh, India

Abstract : Intrusion is set of actions that attempt to compromise the integrity, confidentiality, or availability of a network and its resource and an intrusion detection system (IDS) is a system for the detection of such intrusions. Intrusion Detection System consists of three components Data collection, normalization and classification. This paper reviewed some classification approaches and also reviewed ensemble classifier on basis of adaptive and scalable nature. We also discuss pros and cons of all approaches. This paper compared results of all approaches in terms of True positive rate and false positive rate. We also discuss timelines and average classification.

IndexTerms **- Intrusion Detection, Types of attack, Classification Techniques.**

I. INTRODUCTION

Intrusion Detection Systems

Now a days, internet play a wide role in technological development of society, but intrusion is big challenge to maintain security of data over the network. For improving the network security by the help of IDS. IDS consists of three components: Data Collection being first component looks after the task of data collection and its preprocessing. Data is collected from System/applications log, network packets or network cable. Preprocessing involves converting data to a common format and transferring data to the next component which is Analysis Component or Detection component. In Detection component data is parsed and filtered, detection algorithms are executed to detect any intrusion attempts and if intrusion is detected an appropriate response action is taken against intrusion like drop packets, send alerts, update routing tables, kill processes etc. The response can be passive, in which simply alarm is raised and authority is informed and active, which tries to quench the effects of intrusion by taking necessary actions. Classification is a Data analysis task which involves identifying a set of categories, to which a new observation belongs, on the basis of a training dataset containing observations with known categories membership.

Fig.: IDS Components

Depending on the detection techniques used, IDS can be classified into following categories (1) signature or misuse based IDS (2) anomaly based IDS [07][08]. The signature based IDS is based on signatures and uses attack scenarios which are preknown and compare them with incoming packets traffic. A number of approaches are there for signature detection, which differs in representation and matching algorithm used for detection of intrusion patterns. Anomaly detection bases looks for behavior that deviates from normal behavior. This technique is based on the detection of anomalies present in the traffic. The anomaly detection systems have adaptive nature, they deals with new attack but the specific type of attack cannot be identified.

II. TYPES OF ATTACK

An attack can be classified among one of four categories [04][05][10][11]:

1. Denial-of-Service (DoS): Attackers tries to suspend legitimate users from using network resources by overloading the server. Smurf, teardrop, neptune, back and syn flooding comes under this category.
2. Probe: Attackers tries to gain information about the network in order to make use of a known permeability. Port Scans or sweeping of a given IP-address range comes under this category.
3. User-to-Root (U2R): Attacker uses their local access to the system in order to gain unauthorized super user privileges(root access). Perl, buffer overflow attacks comes under this category.
4. Remote-to-Local(R2L): Attackers do not have local access to the system, thus tries to gain access from a remote system through sending packets. Guess password, spy, ftp write comes under this category.

III. DATASET

All the researchers are widely using KDDcup 99 dataset and several research work have been done with the use of KDDCup99 [06]. But many disadvantages are there in this dataset one of the most important shortcoming in the KDD data set is the presence of large number of redundant record, consequently learning algorithms become biased towards the frequent records, and thus stops them from learning infrequent records which are generally more harmful to networks such as U2R and R2L attacks[09]. So many other datasets are also being used by researchers for IDS such as ISCX 2012, CAIDA, MAWI. The NSL-KDD data set has the following advantages over the original KDD data set: Redundant records are not present in the train set, thus the classifiers will not produce any biased result. There is no duplicate records in the test sets; which results in better reduction rates. The number of records being selected from each difficulty level group is inversely proportional to the percentage of records present in the original KDD data set. As a result, the classification rates of machine learning methods vary in a vast range, which makes it more effective to have a precise evaluation of different learning techniques. The number of records in the train and test sets are reasonable, which

makes it efficient to run the experiments on the complete set without the need to randomly pick a small portion which makes evaluation results of different research works to be consistent and comparable.

IV. CLASSIFICATION TECHNIQUES

Classification is a data mining process of classifying data to different classes based on some model. Many classification algorithms [05] has been developed based on different strategies. All these algorithms uses one or the other mathematical technique like decision tree, linear programming and neural networks. All these algorithms analyze the datasets based on some conditions or strategies to make its prediction. Few are discussed below:

1. Naive Bayes Classifier: Particularly based on Bayesian theorem, the Naïve Bayes Classifiers are a collection of classification algorithms, where a group of algorithms share a common principle, i.e. each pair of features being classified is independent of each other and the possibility of one attribute does not affect the other.
2. Decision Tree Classifier: Based on Divide and conquer strategy, decision tree is a simple and widely used classification technique. It creates a tree structure with the help of conditions and test questions to test data. In the decision tree, the root and internal nodes includes attribute test conditions to separate records on the basis of different characteristics.
3. Support Vector Machine (SVM): It is a perfect example of linear classifier. Classification tasks is carried out by constructing hyper planes in a multidimensional space which performs separation of cases of different class labels.
4. Random Forest: It uses Ensemble algorithms, which combines two or more algorithms of same or different kind for classification of objects. Random forest classifier creates a set of decision trees from randomly picked subset of training set. Then aggregation of the votes from different decision trees is done to decide the final class of the test object.
5. J48: It makes use of divide-and-conquer technique. A decision tree is created recursively based on the greedy algorithm. Decision tree so formed is consists of the root node, branches, parent nodes, child nodes and leaf nodes. A node in a tree denotes attributes present in dataset.
6. Bayesian Network (BN): Bayesian network uses probabilistic approach, using directed acyclic graph it represents random variable sets with their conditional dependencies.
7. OneR: OneR stands for "One Rule" or " One –Attribute-Rule", is a simple and accurate, classification algorithm. The idea behind this algorithm is to find one attribute to use to build a model that makes fewest prediction errors.
8. Sequential Minimal Optimization (SMO): SMO is a support vector machines, which uses optimize training method. It is more sensitive towards noise as compared to other algorithm.

Naive Bayes classifier is a low variance classifier works efficiently with small dataset. Easy to implement, fast, scalable and converges quickly. No problem of dimensionality, but can't be used with large dataset and also when features are correlated. Support Vector Machine has higher accuracy and good theoretical background than fitting. Kernels are available for non-linear data. SVM are hard to interpret. Not recommended to use when support vectors are very large as predicting new instances can be time consuming. Decision tree/ Random Forest are fast, easy to interpret, scalable and non-parametric algorithms. One need not to worry about parameters and separability of data. Accuracy rate and performance is better than other algorithms. Good to use with dataset having missing value attributes. Neural Network algorithms has high degree of non-linearity, but parameters are hard to tune. Building model is time taking.

V. LITERATURE SURVEY

A lot of work has been done in the area of intrusion detection so far. Some of the related previous researches are discussed in this paper. In [01] author proposed a novel multi-classifier layered approach, by combining naive bayes classifier with NBTree to improve detection rate and precision of minority class without hurting the performance of majority class. In [02] author conduct an investigation and evaluation on various data mining algorithms to study the performance of each classifier against noise-free dataset and noisy (10 percent and 20 percent) dataset. In [03] author proposed Feature- Vitality Based Reduction method (FVBRM) in which one input feature is deleted from the dataset at a time, the resultant dataset is then used for the training and testing of the classifier, this process keeps until it performs better than the original dataset. In [05] author performed their analysis on the NSL-KDD dataset with the help of figures and tables and found that NSL-KDD dataset is a best candidate data set to simulate and test the performance of IDS, as mostly the attacks are launched through the inherent drawbacks of the TCP protocol. In [12], author proposed a model "Intelligent Intrusion Detection System" which employed AI approach for detection of intrusion. The technique employed neural network, fuzzy logic, and network, with simple data mining techniques to process the data. In [13] author shows that by selecting right features and good training parameters for designing Artificial Neural Network, the performance and preciseness of an IDS can be improved. In [14] author used feature ranking algorithm, based on Support Vector Machine(SVM), Multivariate Adaptive Regression Splines (MARS) and Linear Genetic Programs (LPGs), to reduce the feature space. In [15] author proposed "Enhanced Support Vector Decision Function "for feature selection, which mainly focuses on feature's rank and the correlation between the features.

VI. EXPECTED OUTCOME/ CONCLUSION

1. If solution is optimized rate of True Positive (TP) will be improved.
2. False positive will be minimized.
3. Decision Rate will be improve.
4. Minimized cost of misclassification will be achieved.
5. Intrusion detection rate will be improved.

REFERENCES

[01] Neelam Sharma, Saurabh Mukherjeeb "A Novel Multi-Classifier Layered Approach to Improve Minority Attack Detection in IDS " 2nd International Conference on Communication, Computing & Security (ICCCS-2012)
[02] Jamal Hussain and Samuel Lalmuanawma "Feature analysis, evaluation and comparisons of classification algorithms based on noisy intrusion dataset" 2nd International Conference on Intelligent Computing, Communication & Convergence(ICCC-2016)
[03] Dr. Saurabh Mukherjeea, Neelam Sharmaa "Intrusion Detection using Naive Bayes Classifier with Feature Reduction" © 2011 Published by Elsevier
[04] Yasmen Wahba1, Ehab ElSalamouny 2 and Ghada ElTaweel "Improving the Performance of Multi-class Intrusion Detection Systems using Feature Reduction" IJCSI International Journal of Computer Science Issues, Volume 12, Issue 3, May 2015
[05] L.Dhanabal1, Dr. S.P. Shantharajah" A Study on NSL-KDD Dataset for Intrusion Detection System Based on Classification Algorithms" International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 6, June 2015
[06] Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu, and Ali A. Ghorbani "A Detailed Analysis of the KDD CUP 99 Data Set" Symposium on Computational Intelligence in Security and Defense Applications (CISDA 2009) IEEE2009
[07] Jian Pei Shambhu J. Upadhyaya Faisal Farooq Venugopal Govindaraju. Proceedings of the 20th International Conference on Data Engineering (ICDE'04) 1063-06382/04 $ 20.00 © 2004 IEEE
[08] Debar, H., Dacier, M., and Wespi, A., A Revised taxonomy for intrusion detection systems, Annales des Telecommunications, Vol. 55, No. 7–8, 361–378, 2000.
[09] KDD Cup 1999. Available on: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html, October 2007.
[10] Sapna S. Kaushik, Dr. Prof.P.R.Deshmukh," Detection of Attacks in an Intrusion Detection System", International Journal of Computer Science and Information Technologies, Vol. 2 (3), 2011, 982-986
[11] V. Bolón-Canedo, N. Sánchez-Maroño, and a. Alonso- Betanzos, "Feature selection and classification in multiple class datasets: An application to KDD Cup 99 dataset," Expert Syst. Appl., vol. 38, no. 5, pp. 5947–5957, 2011.
[12] Norbik Bashah, Idris Bharanidharan Shanmugam, and Abdul Manan Ahmed," Hybrid Intelligent Intrusion Detection System" World Academy of Science, Engineering and Technology, 2005
[13] Saman M. Abdulla, Najla B. Al-Dabagh, Omar Zakaria, Identify Features and Parameters to Devise an Accurate Intrusion Detection System Using Artificial Neural Network, World Academy of Science, Engineering and Technology 2010.
[14] A. H. Sung, S. Mukkamala. (2004) The Feature Selection and Intrusion Detection Problems. In Proceedings of the 9th Asian Computing Science Conference, Lecture Notes in Computer Science 3029 Springer 2004, pp.
[15]S Zaman, F Karray Features selection for intrusion detection systems based on support vector machinesCCNC'09 Proceedings of the 6th IEEE Conference on Consumer Communications and Networking Conference 2009