International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 06 | June-2018 www.irjet.net p-ISSN: 2395-0072

A HYBRID INTRUSION DETECTION TECHNIQUE BASED ON IRF & AODE FOR KDD-CUP 99 DATASET

Mr. Sandip Hingane[1], Dr. Umesh Kumar Lilhore[2]

M. Tech Scholar, Department of CSE, NIIST Bhopal
Head PG, Department of CSE, NIIST Bhopal

ABSTRACT-*Intrusion detection systems are widely used to detect intruders and malicious nodes from the network. An intrusion detection system is used to detection intruder based on activities such as normal and abnormal. Random Forest is an ensemble classifier and performs well compared to other existing classifiers for effective detection and classification of network attacks. In RF sometimes accuracy is a challenging factor. In this research work we are presenting and Hybrid intrusion detection system (HIDS) based on Improved Random forest (IRF) with bagging and Average One-Dependence Estimator (AODE). In proposed method a new sampling method is used to improve existing RF. Proposed HIDS method resolves issues of existing methods such as poor accuracy, precision and recall. For validation the performance of proposed HIDS method several performance measuring parameters such as accuracy, detection rate and false alarm rate are calculated and compared with existing random forest with SVM method for KDD cup-99 dataset. An experimental result analysis clearly shows that proposed HIDS method performs outstanding over existing method.*

Keywords- Hybrid intrusion detection system, Random forest, bagging, SVM, Improved RF

1. INTRODUCTION

The Internet has turned into the most fundamental device and one of the best wellsprings of data about the present world. Web can be considered as one of the significant segments of instruction and business reason. Consequently, the information over the Web must be secure. Web security is one of the major concerns now-a-days. As Internet is undermined by different assaults it is exceptionally basic to outline a framework to secure those information, and also the clients utilizing those information [7].

An ID is along these lines an innovation to satisfy that prerequisite. System heads adjust interruption location framework keeping in mind the end goal to counteract vindictive assaults. Along these lines, interruption discovery framework turned into a basic piece of the security administration. Interruption recognition framework recognizes and reports any interruption endeavors or abuse on the system. IDS can distinguish and piece vindictive assaults on the system, hold the execution typical amid any malignant episode, play out an accomplished security examination [1, 2].

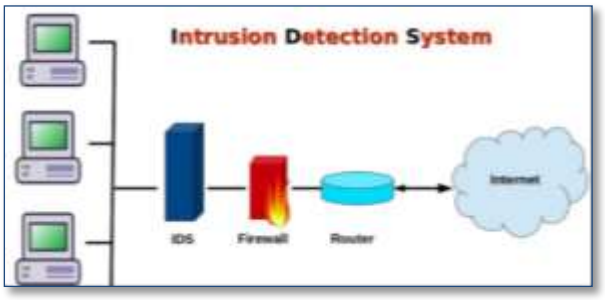


Figure 1.1 IDS

2. DATA MINING & MACHINE LEARNING TECHNIQUES

Data mining technique is a method of finding all the unknown information from a given data set or patterns. Just in case of intrusion detection system, we have a tendency to use the construct of knowledge mining we'll ascertain the data patterns which are able to determine and track all the users' info associate degreed activities to observe and trace an trespasser [7]. In existing system we've a bent to square measure specializing in knowledge engineering processes inside that the choices taken on the thought of some mounted rule. In the main intrusion detection system is split in to 2 broad classes' i.e. intrusion detection system victimization association rule mining and intrusion detection system victimization event correlation data mining [8].

3. EXISTING METHODS & CHALLENGES

Following methods are widely used in IDS-

3.1 AODE-Averaged one dependence estimators (AODE) could be a probabilistic classification learning technique. it absolutely was developed to handle the attribute-independence downside of the favored naive Bayes classifier. It oft develops considerably a lot of correct classifiers than naive Bayes at the value of a modest increase in the quantity of computation.

3.2 RF-In recent times Random forest (RF) is used wide for IDS downside. RF combines textile and random choice of options. RF consists of the many classification trees. RF improves the accuracy and reduces error rate for big information sets. RF generates out of bag error throughout training part. In random forest mainly three turning parameters area utilized which are, No of trees, minimum node size, numbers of descriptors area unit used for cacophonous every node [1].

3.3 SVM- In machine learning process, support vector machines are supervised learning models with associated learning methods that analyze data mainly used for the classification process and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier [1,7].

3.4 NAIVE BYES-In machine learning (ML) process, naive classifiers are a family of straightforward "probabilistic classifiers" supported applying theorem with sturdy (naive) independence assumptions between the options [7]. A naive mathematician has been studied extensively since the Fifties. It had been introduced beneath a unique name into the text retrieval community within the early Nineteen Sixties and remains a preferred baseline methodology for text categorization, the matter of judgment documents as happiness to at least one class or the opposite such as spam or legitimate, sports or politics, etc. with word frequencies because the options.

3.5 DECISION TREES- A decision tree could be a call support tool that uses a tree-like graph or model of choices and their potential consequences, together with natural event outcomes, resource prices, and utility. It's a method to show associate degree algorithmic rule that solely contains conditional management statements. Decision trees area unit unremarkably utilized in research, specifically in call analysis, to assist determine a method possibly to succeed in a goal, however, are a well-liked tool in machine learning.

3.6 SAMPLING METHOD-Sampling information could be a methodology used for anomaly detection and change detection, to Illustrate; DoS attack detection [15]. Cisco Net Flow [6] could be a sampling technique to decrease the heavy load on router central processing unit in high-speed networks. However, sampling has negative impacts on the applied mathematics characteristics of traffic and thence on the performance of intrusion detection. Sampling method play a important role in efficient selection of accurate data.

3.7 CHALLENGES -

An Intrusion detection system is widely used in several data mining and machine learning techniques such as Random forest, SVM, Naive Bayes and AODE. These algorithms still encounter with several issues such as higher error rate %, time and poor accuracy %. These challenges attract researchers to work in the field of data security. Existing SVM and Naive Bayes method encounter with attribute dependency problem. Other challenges are poor results in following parameters, which are as follows-

I **Precision-** Precision refers to the closeness of a measured value to a standard or known value. It shows positive predictive values. A Precision value is a percentage of total correctly classified data or instances over the total returned instances.

II **Accuracy-** Accuracy is a percentage of total correctly classified data or instances from the complete or whole dataset.

III **Sensitivity (True positive rate)** -Sensitivity (also called the true positive rate, the recall, or probability of detection in some fields) measures the proportion of positives that are correctly identified as such (e.g. the percentage of sick people who are correctly identified as having the condition).

IV **Recall**-In information retrieval, recall is the fraction of the relevant documents that are successfully retrieved.

V **F-measure-**The F-score (or F-measure) considers both the precision and the recall of the test to compute the score. The precision p is the number of correct positive results.

4. PROPOSED WORK

An intrusion detection system is used to detection intruder based on activities such as normal and abnormal. A Random Forest also called RF is mainly an ensemble based classifier method. RF classifier performs much better as compared to other existing classifiers for effective detection and classification of network attacks. In RF sometimes accuracy is a challenging factor. In this research work, we are presenting and Hybrid intrusion detection system (HIDS) based on the Random forest (RF) with bagging and Average One-Dependence Estimator (AODE). In proposed method a new sampling method is used to improve existing RF. Proposed HIDS method resolves issues of existing methods such as poor accuracy, precision and recall.

In proposed method ensemble techniques such as RF (Improved RF) and Bagging are used. The term ensemble

is usually reserved for methods that generate multiple hypotheses using the same base learner. Random forest tree is used in HIDS because it improves the overall accuracy and also reduces the error rate %. During training phase Random forests generates and check out of bag errors. In random forest mainly three parameters play vital roles: total number of trees, total number of descriptors (Basically used in splitting) and node size (minimum size).

An AODE is a classifier, which classify data in various multi-classes, which improves the total accuracy of the work. AODE classifier can also able to analyze network traffic as normal or attack more efficiently. AODE classifiers can manage small size data as well as large size data. In proposed HIDS bagging technique is also used with RF (Improved) and AODE. Bagging is an ensemble technique which reduces the problem mainly related to over fitting of the training dataset.

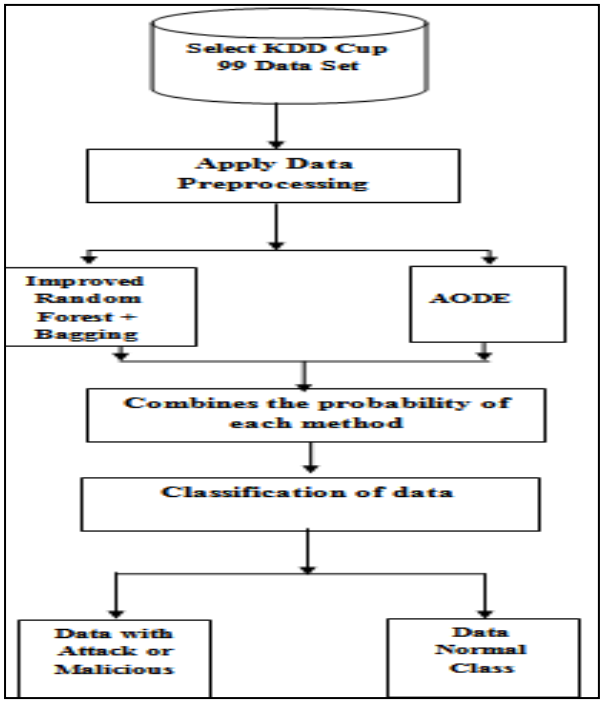


Figure 4.1 Flow chart for proposed method

5. SIMULATION & RESULT

In this research work, we are presenting and Hybrid intrusion detection system (HIDS) based on the Improved Random forest (IRF) with bagging and Average One-Dependence Estimator (AODE). Proposed HIDS method resolves issues of existing methods such as poor accuracy, precision and recall. Proposed HIDS method and existing (RF + SVM) method both are implemented over MATLAB simulator R-20013a and WEKA 3.6 tool.

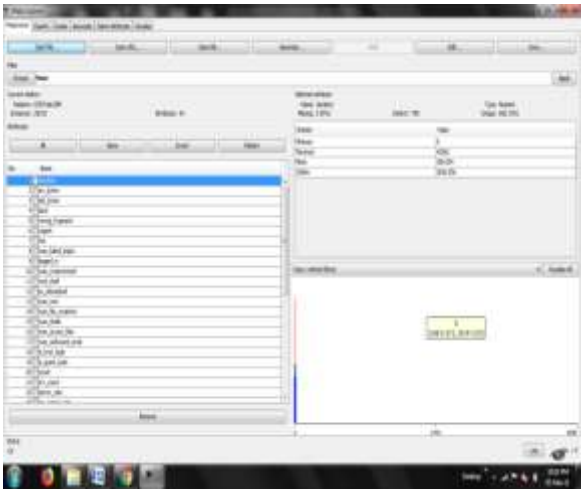


Figure 5.1.1 MATLAB results

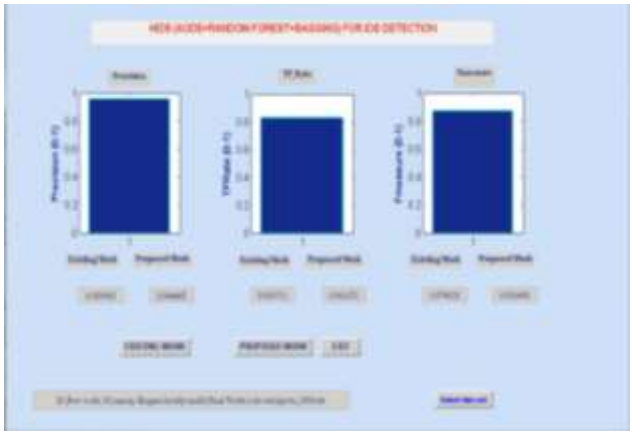
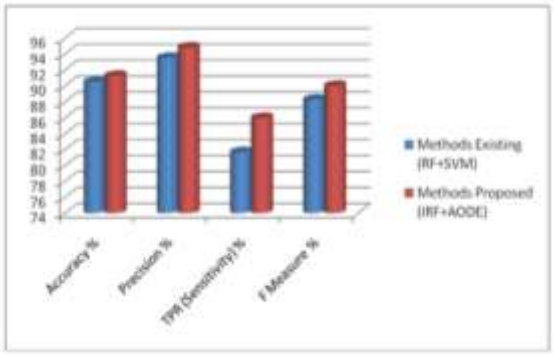


Figure 5.1.2 MATLAB results

COMPARISONS OF SIMULATION RESULTS

Parameters	Methods	
	Existing (RF+SVM)	Proposed (IRF+AODE)
Accuracy %	90.47	91.25
Precision %	93.5	94.8
TPR (Sensitivity) %	81.66	85.96
F Measure %	88.25	89.99

Table 5.4.1 Comparisons of simulation results



Graph 5.4.1 Comparisons of simulation results

From the experimental results we can say table 5.4.1 and graph 5.4.1 is showing that proposed method (IRF + AODE) performing outstandingly in terms of accuracy %, precision %, TPR % and F measure % over existing (RF + SVM) method.

6. CONCLUSIONS & FUTURE WORKS

In this research work, we have presented Hybrid intrusion detection system (HIDS) based on the Improved Random forest (IRF) with bagging and Average One-Dependence Estimator (AODE). The proposed method and existing method (RF + SVM) are implemented over MATLAB and WEKA machine learning tools and various performance measuring parameters are calculated such as precision, recall, true positive rate and accuracy. Experimental results are clearly showing that proposed method performing outstanding over existing methods.

6.1 FUTURE WORKS-Proposed method and existing method both are implemented over static data. In future work we can implements these methods with real time data and more machine learning methods with stacking can be use for performance comparison.

REFERENCES

[1]. Yaping Chang, WeiLi, "Network Intrusion Detection System based on Random Forest and SVM ", 2017 IEEE International Conference on CSE & EUC, PP 635-639.

[2]. Amreen Sultana, M.A.Jabbar, "Intelligent Network Intrusion Detection System using Data Mining Techniques", IEEE 2nd International Conference on Applied and Theoretical Computing and Communication Technology (ICAC), July 2016, pp 329-334.

[3]. M A Jabbar a, Rajanikanth Aluvalub, Sai Satyanarayana Reddy S," RFAODE: A Novel Ensemble Intrusion Detection System", ELSEVIER 7th International Conference on Advances in Computing & Communications, ICACC-2017, 22-24 August 2017, Cochin, India, pp 226-234.

[4]. Levent Koc and Alan D. Carswell," Application of an AODE Based Classifier to Detect DOS Attacks", IJCSNS International Journal of Computer Science and Network Security, VOL.15 No.2, February 2015, pp 24-29.

[5]. Adel Ammar," Adel Ammar A Decision Tree Classifier for Intrusion Detection Priority Tagging", Journal of Computer and Communications, 3, March 2015, pp 52-58.

[6]. Sean T Miller, Curtis Busby-Earle, "Multi-Perspective Machine Learning a Classifier Ensemble Method for Intrusion Detection", ICMLSC '17 Proceedings of the 2017 International Conference on Machine Learning and Soft Computing, Jan 2017, pp 7-12.

[7]. Bhupendra Ingre, Anamika Yadav," Performance Analysis of NSL-KDD dataset using ANN", Conference SPACES-2015, pp 92-97.

[8]. Roshani Gaidhane, Prof. C. Vaidya, Dr. M. Raghuwanshi," A Survey: Learning Techniques for Intrusion Detection System (IDS)", International Journal of Advance Foundation and Research in Computer (IJAFRC) Volume 1, Issue 2, Feb 2014, pp 21-29.

[9]. Sejal K. Patel, Umang H. Mehta, Urmi M. Patel, Dhruv H. Bhagat,"A Technical Review on Intrusion Detection System", International Journal of Computer Science & Engineering Technology (IJCSET), Vol. 6 No. 01 Jan 2015, pp 17-25.

[10]. AbdullaAminAburomman,"Surveyoflearningmethodsi nintrusiondetection system", 2016 International Conference on Advances in Electrical, Electronic and System Engineering, 14-16 Nov 2016, pp 362-366.

[11]. Vipin Das, Vijaya Pathak, Sattvik Sharma, Sreevathsan, MVVM.Srikanth, Gireesh Kumar T", Network intrusion detection system based on machine learning algorithms", International Journal of Computer Science & Information Technology (IJCSIT), Vol. 2, No 6, December 2010, pp 138-152.

[12]. Nutan Farah Haq, Musharrat Rafni, "Application of Machine Learning Approaches in Intrusion Detection System: A Survey", (IJARAI) International Journal of Advanced Research in Artificial Intelligence, Vol. 4, 2015, pp 9-19.

[13]. Trupti Phutane, Apashabi Pathan,"A Survey of Intrusion Detection System Using Different Data Mining Techniques", International Journal of Innovative Research in Computer and Communication Engineering Vol. 2, Issue 11, November 2014, pp 6801-6808.

[14]. Chirag Modi, Dhiren Patel a, Bhavesh Borisaniya a, Hiren Patel b, "A survey of intrusion detection techniques in Cloud", ELSEVIER Journal of Network and Computer Applications, June 2012, pp 42-57.

[15]. Yanjie Zhao, "Network Intrusion Detection System Model Based on Data Mining", IEEE SNPD 2016, May 30-June 1, 2016, Shanghai, China, pp 206-212.

[16]. M.A. Jabbar, B.L. Deekshatulu, Priti Chandra, "Computational intelligence techniques for early diagnosis of heart disease", ICETECH, IEEE 2015, pp 127-133.

[17]. C.Ellcan, "Results of the KDD CUP 99 classifier learning "ACM SIG-KDD, Explorations newsletter, Feb-2000, 119-128.

[18]. Theodoros Lappas and Konstantinos Pelechrinis, "Data Mining Techniques for (Network) Intrusion Detection Systems", International Journal of Computer and Telecommunications Networking, Vol. 64, Issue 7, June 2015, pp: 206-219.

[19]. Krishna Kant Tiwari, Susheel Tiwari, Sriram Yadav, Intrusion Detection Using Data Mining Techniques, IJACT, Nov 2016, pp 21-27.

[20]. Kapil Wankhade, Sadia Patka," An Efficient Approach for Intrusion Detection Using Data Mining Methods", IEEE 2013 International Conference on Advances in Computing, Communications, and Informatics (ICACCI), June 2013, pp 1615-1618.

[21]. Sharmila Kishor Wagh, Vinod K. Pachghare, Satish R. Kolhe, "Survey on Intrusion Detection System using Machine Learning Techniques", International Journal of Computer Applications (0975- 8887) Volume 78, No. 16, September 2013, pp 30-38.

[22]. Sandip Hingane, Dr. Umesh Kumar Lilhore, "Intrusion Detection Techniques: A Review", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN: 2456-3307, Volume 3, Issue 1, pp.129-135, January-February.2018

[23]. Nilofer Shoaib Khan, Prof. Umesh Lilhore, "Review of various intrusion detection methods for training data sets", International Journal of Modern Trends in Engineering and Research (IJMTER) Volume 03, Issue 12, 197-202.

[24]. Soumya Tiwari, Umesh Lilhore and Ankita Singh. Artificial Neural Network and Genetic Clustering based Robust Intrusion Detection System. International Journal of Computer Applications 179(36):36-40, April 2018.

[25]. Mrs. Nilofer Shoaib Khan , Prof. Umesh Lilhore, "An Efficient NIDS by using Hybrid Classifiers Decision Tree & Decision Rules", International Journal of Science & Engineering Development Research (www.ijsdr.org), ISSN:2455-2631, Vol.2, Issue 1, page no.76 - 79, January 2017.