International Journal of Research in Advent Technology (IJRAT), Special Issue, March 2019
E-ISSN: 2321-9637
ICCCMIT 2019 organised by M.O.P. Vaishnav College for Women (Autonomous)
Chennai-34, India
Available online at www.ijrat.org

Classification Algorithm for Intrusion Detection in MANET using KDD Cup '99 Dataset

V.Asaithambi[*1], Dr.N.Rama[#2]
[*]*Assistant Professor, Department of Computer Science,*
Govt. Arts College for Men, Nandanam, Chennai-600035.
[#]*Principal, Govt. Arts and Science College for Women,*
Villupuram -605602.
[1]*vsasai@yahoo.com*
[2]nramabalu@gmail.com

Abstract: Mobile Ad hoc NETwork (MANET) is a network with the capability of self-configuring nature and the nodes in the network are changed their locations accordingly. The network is attacked by any person or software called intruder at any time. The nature of the intruder may be anything like attack or spoil the data or slowdown the performance of the network. This intrusion process has to be detected and prevented. The intrusion detection system is evolving for this purpose. There are many algorithms are developed for intrusion detection. Most of the algorithms are using the trace data i.e. the KDD Cup'99 dataset as universally accepted dataset. This paper will give an idea to use classification algorithm on the dataset and gives a model for intrusion detection.

Keywords: MANET, KDD Cup'99, Classification.

1. INTRODUCTION

A Mobile ad-hoc network (MANET) is a dynamic auto-configuring wireless network. No static network structure is used as its nodes and edges are dynamic in nature. The structure and communication of data is varying time to time. The data during communication may be attacked by any external factors. The attacks will be merged with network packets. The KDD Cup'99 dataset is giving the packet details during a particular duration of time. The data-mining algorithms like clustering and classification are used to categorize the packets of KDD Cup'99 dataset. The classification algorithm is widely used to analyse the performance of detection. Here the test dataset is used with classification algorithm. The data set is processed and manipulated to improve the detection of attacks.

This paper deals with classification algorithm with various form of KDD Cup'99 dataset to improve the intrusion detection and also to analyse the results from the classification algorithm with different form of KDD Cup'99 dataset.

2. LITERATURE REVIEW

In the domain of network security the Intrusion Detection Systems (IDS) are used to detect the type of intrusion in the form of attack made by the intruder. There are many research works that deal with IDS. An IDS was proposed with the distributed and cooperative manner in which a node detects a low confidence intrusion and initiate a global intrusion detection procedure with a cooperative detection engine. A new rule-based classification algorithm is developed for intrusion detection. In addition, a meta-learning approach is also used by them to enhance the behaviour of IDS.

An IDS was developed using a Multiclass SVM algorithm. This produced an efficient intrusion detection system. Later the SVM model is integrated with a decision tree model and this hybrid model provide better results than the individual models. A multiple-level hybrid classifier is proposed for developing an IDS. This system merges the tree classifiers and clustering algorithms to detect intrusions efficiently. Detection performance of 3-level tree classifiers are compared with these algorithms, and the above approaches have shown that considerable improvement in detection of intrusions.

Amini et al introduced an intelligent intrusion method for both detecting known and unknown attacks. The intrusions are detected by using unsupervised neural networks in real time. Without retraining the unsupervised neural networks can perform the analysis of new data. The evaluation of ART and self organizing map NNs uses offline data in their work. Koutsoutos et al presented a neural classifier ensemble using a combination of NNs which are capable of detecting network attacks on web servers. Their IDS was capable of identifying even unseen attacks and classify them. The success

rates of the performance of the NN used by them for detecting attacks from audit dataset was good and more than 78% in detecting novel attacks. However, the false alarms rates was high and hence it is required to do suitable enhancements on their work.

3. PROPOSED METHODS

The KDD Cup'99 dataset is used to form different datasets. Each dataset is processed using classification algorithm. The result for each dataset will be analysed. The following figure Fig.1 gives an idea for overall system model. It has three steps to complete the process. The input module has the data which is taken from standard KDD Cup'99

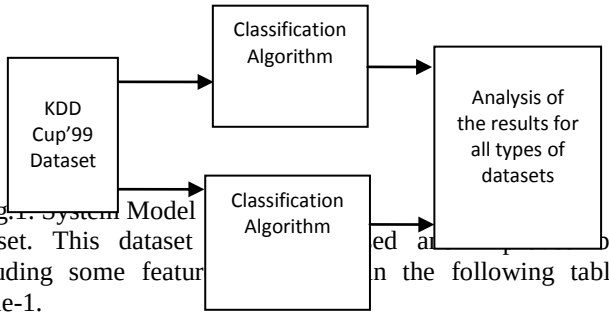

Fig.1. System Model

dataset. This dataset ed a by including some featur n the following table Table-1.

Sl. No in KDD Cup'99 Data Set	Feature Name
NA	Src_host
NA	Dst_host
2	Protocol_type
3	Service
4	Src_bytes
5	Dst_bytes
7	Land
8	Wrong_fragment
23	Count
30.	Srv_rerror_rate
36.	Dst_host_same_ src_port_rate
NA – Not Applicable	

Table-1: Per-processed Dataset

The classification algorithm used in this paper help to improve the performance of intrusion detection on the various form of dataset.

Classification Algorithm Performance: The performance of a classification algorithm can be summarized using Classification accuracy, Accuracy by class and Confusion matrix. Classification accuracy is the ratio of the number of correct predictions out of all predictions made. The true-positive and false-positive rates for the predictions for each class are considered for the Accuracy by class. Confusion matrix is a table showing the number of predictions for each class compared to the number of instances that actually belong to each class.

Confusion matrix: The confusion matrix is a table with two rows and two columns. The four outcomes of this matrix are used to measures like error-rate, accuracy, specificity, sensitivity, and precision. The most important measures ROC and precision-recall are also derived from the confusion matrix. The following equations are used to calculate above said measures.

The confusion matrix and four outcomes from confusion matrix are shown as follows

Confusion Matrix		Predicted	
		Positive	**Positive**
Observed	**Positive**	TP	FN
	Negative	FP	TN

TP(True Positive) : correct positive prediction
FP(False Positive) : incorrect positive prediction
TN(True Negative) :correct negative prediction
FN(False Negative):incorrect negative prediction

The measures derived from confusion matrix are

Error rate : Error rate (ERR) is calculated as the number of all incorrect predictions divided by the total number of the dataset. The error rate is ranges from 0.0 to 1.0 respectively from best to the worst error rate.

ERR = (FP+FN) / (TP+TN+FN+FP)

Accuracy: Accuracy (ACC) is calculated as the number of all correct predictions divided by the total number of the dataset. The accuracy is ranges from 1.0 to 0.0 respectively from best to the worst in accuracy. It can also be calculated by 1 – ERR.

ACC = (TP+TN) / (TP+TN+FN+FP)

Sensitivity: Sensitivity (SN) is calculated as the number of correct positive predictions divided by the total number of positives. It is also called recall (REC) or true positive rate (TPR). The sensitivity is ranges from 1.0 to 0.0 respectively from best to the worst in sensitivity.

SN = TP / (TP+FN)

Specificity: Specificity (SP) is calculated as the number of correct negative predictions divided by the total number of negatives. It is also called true negative rate (TNR). The specificity is ranges from 1.0 to 0.0 respectively from best to the worst in specificity.

SP = TN / (TN + FP)

Precision : Precision (PREC) is calculated as the number of correct positive predictions divided by the total number of positive predictions. It is also called positive predictive value (PPV). The precision is ranges from 1.0 to 0.0 respectively from best to the worst in precision.

PREC = TP / (TP + FP)

False positive rate: False positive rate (FPR) is calculated as the number of incorrect positive predictions divided by the total number of negatives. The false positive rate is ranges from 0.0 to 1.0 respectively from best to the worst in false positive rate. It can also be calculated as 1 – specificity.
FPR = FP / (TN + FP)

All the above measures are calculated for the test dataset and the reformed dataset using the J48 decision tree classification algorithm.

4. SAMPLE RESULTS AND REPORTS

The following reports shows the performance of the classification algorithm using the test dataset before reforming the dataset and after reforming the dataset using random subset algorithm. The fig.2. gives the statistics of correctly classified instances. The following results shown the details of various measures for the test data set and processed dataset

Time taken to build model: 23.48 seconds for test dataset

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	49355	99.9049 %
Incorrectly Classified Instances	47	0.0951 %
Kappa statistic	0.9974	
Mean absolute error	0.0001	
Root mean squared error	0.0086	
Relative absolute error	0.342 %	
Root relative squared error	6.7772 %	
Coverage of cases (0.95 level)	99.9211 %	
Mean rel. region size (0.95 level)	4.3537 %	
Total Number of Instances	49402	

Time taken to build model: 13.31 seconds for processed dataset

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	49359	99.913 %
Incorrectly Classified Instances	43	0.087 %
Kappa statistic	0.9976	
Mean absolute error	0.0001	
Root mean squared error	0.0084	
Relative absolute error	0.3386 %	
Root relative squared error	6.6411 %	
Coverage of cases (0.95 level)	99.9231 %	
Mean rel. region size (0.95 level)	4.3562 %	
Total Number of Instances	49402	

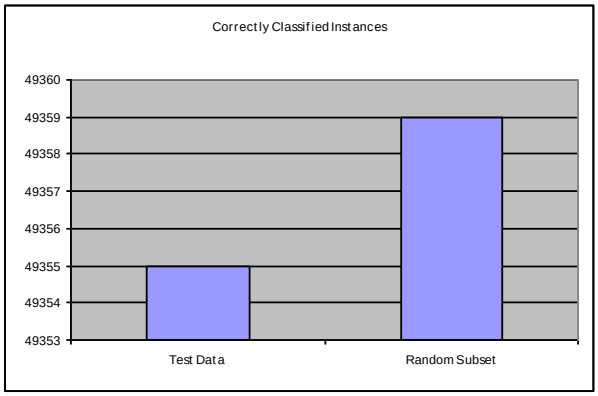

Fig.2. Correctly classified instances

The fig.3. shows the various measures using confusion matrix outcomes for both test dataset and reformed dataset.

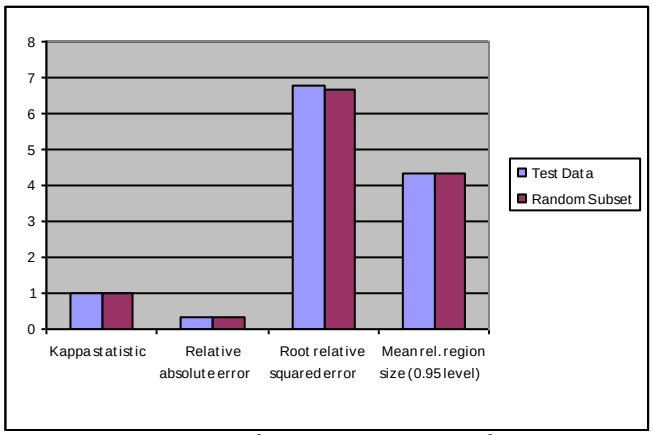

Fig.3. Comparison of measures using confusion matrix

5. CONCLUSION

This paper will give an idea to develop algorithms to calculate and compare the results of classification algorithm using the KDD Cup'99 dataset. The test data set is taken from the original dataset. The random subset algorithm is used to extract the subset of features which forms a dataset. Both the datasets are used for the decision tree classification algorithm and various measures are calculated and compared. This process can be extended for other feature extraction algorithms in future and the best feature selection algorithm can be identified for classifying the attacks on the network dataset.

REFERENCES

[1] T.Miranda Lakshmi, A.Martin, R.Mumtaj Begum, and Dr. V.Prasanna Venkatesnan,"An Analysis on Performance of decision Tree Algorithms using Student's Qualitative Data", I.J. Modern Education and Computer Science, June 2013.

[2] Hemlata Chahal, "ID3 Modification and Implementation in Data Mining", International Journal of Computer Applications (0975-8887), Volume 80-No7, October 2013.

[3] Daniel K. Blandford Guy E. Blelloch Ian A. Kash, Computer Science Department Carnegie Mellon University, Pittsburgh, PA 15213 fblandford,blelloch,iakg@cs.cmu.edu, *An Experimental Analysis of a Compact Graph Representation*.

[4] G.Kesavaraj, Dr. S.Sukumaran, "A Study On Classification Techniques in Data Mining", IEEE-31661, July 4-6, 2013.

[5] Asaithambi V, Zackariah N, Nirmala K, IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012) March 30, 31, 2012 285Decision Equations for Efficient Search Algorithms for Network Security

[6] Preeti Aggarwala, Sudhir Kumar Sharmab, 3rd International Conference on Recent Trends in Computing 2015, Analysis of KDD Dataset Attributes - Class wise For Intrusion Detection (*ICRTC*-2015).

[7] Senthilnayaki Balakrishnan, Venkatalakshmi K, Kannan A , *International Journal of Computer Science and Application (IJCSA)* Volume 3 Issue 4, Intrusion Detection System Using Feature Selection and Classification Technique (November 2014).

[8] (2017, March 4), Random Forest[Online]. Available: http://en.wikipedia.org/wiki/Ra ndom_Forest.

[9] Shailesh Singh Panwar, Dr. Y. P. Raiwani, International Journal of Computer Engineering and Technology (IJCET), ISSN 0976-6367(Print),ISSN 0976 - 6375(Online), Volume 5, Issue 10, pp. 21-31, *Data Reduction Techniques To Analyze NSL-KDD Dataset* October (2014).

[10] Swasti Singhal, Monika Jena, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-2, Issue-6, May 2013,*A Study on WEKA Tool for Data Preprocessing, Classification and Clustering*

[11] Ellis Horowitz and Sartaj Sahni, (1999) Tata Mc Hill, *Fundamentals of Computer Algorithms*.

[12] Matt Curtin (1997) Kent Information Services, *Introduction to Network Security*.

[13] Feng Li and Ju Wu, "Uncertainty Modeling and Reduction in MANETs", IEEE Transactions on Mobile Computing, Vol.9, No.7, pp. 1035-1048, 2010.

[14] Bellovin, S.M. "Network firewalls", IEEE Communications Magazine, Vol.32, pp. 50-57, 1994.

[15] Naccache, D. "Secure and Practical Identity-based Encryption", IET Transactions on Information Security, Vol.1, No.2, pp.59-64, 2007.

[16] Liu, Y., Tian, D. and Wangannids, A. "Intrusion Detection System based on Artificial Neural Network," In Proceedings of the Second International Conference on Machine Learning and Cybernetics, Wan, pp.2-5, 2003.

[17] Ryutov, T., Neuman, C., Kim, D. and Zhou, L. "Integrated Access Control and Intrusion Detection for Web Servers", IEEE Transactions on Parallel and Distributed Systems, Vol.14, No.9, pp 841-850, 2003.

[18] Lunt, T. "Detecting intruders in computer systems," in Proceedings of Conference on Auditing and Computer Technology, pp. 1-17, 1993.

[19] Denning, D.E. "An Intrusion Detection Model", IEEE Transaction on Software Engineering, Vol.13, pp. 222-232, 1987.

[20] Pikoulas, J., Buchanan, W.J, Manion, M. and Triantafyllopoulos, K. "An Intelligent Agent Intrusion System," in Proceedings of the 9th IEEE International Conference and Workshop on the Engineering of Computer Based Systems - ECBS, IEEE Computer Society, Luden, Sweden, pp. 94–102, 2002.

[21] Anderson J.P, "Computer Security Threat Monitoring and Surveillance," 1980.

[22] Xiang, C. and Lim, S.M. "Design of Multiple-Level Hybrid Classifier for Intrusion Detection System", IEEE Transactions on System, Man, Cybernetics, Part A, Cybernetics, Vol. 2, No. 28, pp. 117–122, 2002.

[23] Zhang Y., Lee W. and Huang Y. A. "Intrusion Detection Techniques for Mobile Wireless Networks", ACM Journal of Wireless Networks, Vol. 9, No. 5, pp. 545-556, 2003.

[24] Cohen, W.W., "Fast Effective Rule Induction", In Proceedings of the 12th International Conference on Machine Learning, pp. 115–123, 1995.

[25] Bace R, Mell P, "Intrusion Detection Systems, Computer Security Division, Information Technology Laboratory, Nat'l Inst. of Standards and Technology", 2001.

[26] Joo, D., Hong, T. and Han, I. "The Neural Network Models for IDS based on The asymmetric Costs of False Negative Errors and False Positive Errors", Expert Systems with Applications, Vol. 25, pp. 69-75, 2003.

[27] Liu, Y., Tian, D. and Wangannids, A. "Intrusion Detection System based on Artificial Neural Network," In Proceedings of the Second International Conference on Machine Learning and Cybernetics, Wan, pp.2-5, 2003.

[28] Moradi, M. and Zulkernine, M. "A Neural Network based System for Intrusion Detection and Classification

of Attacks," in Proceedings of IEEE International Conference on Advances in Intelligent Systems – Theory and Applications, Luxembourg, Vol. 148, pp. 1-6, 2004.

[29] Sarasamma, S., Zhu, Q. and Huff, J. "Hierarchical Kohonen Net for Anomaly Detection in Network Security", IEEE Transactions on System, Man, Cybernetics, Part B, Cybernetics, Vol. 35, No. 2, pp. 302-312, 2005.

[30] Amini, M., Jalili, R. and Shahriari, H.R. "RT-UNNID: A practical solution to real-time network-based intrusion detection using unsupervised neural networks", Computers and Security Science Direct, Vol. 25, No.6, pp. 459-468, 2006.

[31] Koutsoutos, S., Christou, I.T. and Efremidis, S. "A classifier ensemble approach to intrusion detection for network-initiated attacks," in Proceedings of the International Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies, Vol.160, pp.307-319, 2007.