iJRASET
International Journal For Research in Applied Science and Engineering Technology



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 5 **Issue:** IV **Month of publication:** April 2017

**DOI:** http://doi.org/10.22214/ijraset.2017.4270

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

Volume 5 Issue IV, April 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

# An Intrusion Detection System for Identifying Attacks using Classification Technique

Aanchal Tiwari[1], Rohit Miri[2], Amit Kumar Dewangan[3]
[1,2,3] *Dept of CSE, Dr. C. V. Raman University, Kota, Bilaspur, Chhattisgarh, India*

***Abstract:*** ***Information security is one of the important role to protect the information from unauthorized person. Intrusion detection is a classifier to classify the data as normal and various types of attacks. Data mining based decision tree algorithm play very important role to develop the robust IDS to classify the attacks which is harmful for our system. In this research work, used decision tree techniques as classifier to classify the attacks. We have also develop the robust ensemble model which is combination of C4.5, Simple CART and decision tree that gives better accuracy. Our proposed ensemble model gives 99.70% with 80-20% training –testing partition. We have also applied the feature selection technique to computationally increase the performance of model. Our proposed model gives 99.80% in 11 features with info gain feature selection technique while 98.80% in 16 features with gain ratio feature selection technique.***
***Keywords: Intrusion Detection System, Classifier, Attack.***

## I. INTRODUCTION

Now a day's data is increasing day by day in every organization. To secure such information is one of the most important issue for every organization. Information security is one of the important issues to protect the information from unauthorized access. An intrusion detection system is used to detect several types of malicious behaviors that can compromise the security and trust of computer system. Classification plays very important role to classify the unwanted data. Actually IDS is a classifier that classifies the different types of attacks and normal data. In this research work, we have used decision tree algorithm to develop the classifier which classify the normal and different types of attacks. Various researcher have worked to develop the IDS. J. Jabez et al. [2] have presented the details of new approach called Outlier Detection approach to detect the intrusion in the computer network. Their training model was consist of big datasets with distributed environment that improves the performance of Intrusion detection system. There proposed approach was also been tested with the KDD data sets that are received from real world. Experimental results shows that proposed IDS system took less execution time and storage to predict and also the performance of proposed IDS is better than that of other existing machine learning approaches and can significantly detect almost all anomaly data in the computer network. Y. Maleh et al. [3] have evaluated the performance of their intrusion detection model using KDDcup'99 database and they studied the variations in detection rate and false positive, when the number of IDS increases in the network. They proposed a hybrid intrusion detection model for WSN. Their IDS used a learning algorithm based on the SVM and a detection technique based on the attack signatures and their combined model of IDS achieved a higher rate of intrusion detection almost 98% with a number very reduces false alarms near 2%. N. Sharma et al. [4] have observed several research works and they have compared the resulting discussions by their techniques. This paper provided a direction in the face of Intrusion detection improvement and suggested just like "The detection approach can be better at detecting R2L and U2R attacks more efficiently as well as anomaly detection approach, which is better at detecting attacks at the absence of match signatures as provided in the misuse rule files", "Hybridization of Association and Optimization can provide better detection. J. Reddy et al. [5] have proposed a new virtual Honeynet architecture that implements virtual honeynet collaboration systems (VHCS) which can be able to overcome the honeypot module and security module problem. Here, they studied and used different types of honeypots, intrusion detection systems and related analysis tools. When honeypots was implemented, log file was generated. By the help of the data gathered, it was found that most of the attacks were on protocols which are based on TCP/IP. HTTP port was one of the most vulnerable port. Another vulnerable port found was FTP port. It was also found that the number of vulnerabilities increased when this port was opened. Also, there exists Proxy scan attempt, IIS attempt using the get command. This approach can be worked with anti-spam technology to achieve real time detection and prevention system to minimize the attack and sources.

## II. METHODOLOGY

A decision tree [1] is a flow chart like tree structure, where each internal node denote a test on an attribute, each branch represent an


©IJRASET: All Rights are Reserved

outcome of the test, and each leaf node hold a class label. The topmost node in a tree is the root node. Decision tree can handle high dimensional data. Their representation of acquired knowledge in tree from is intuitive and generally easy to assimilate to human. The learning and classification steps of decision tree induction are simple and fast. Decision tree algorithm is simple and fast. These tree classifiers have good accuracy. Decision tree induction algorithms have been used for classification in many application areas such as medicine, manufacturing, and production, Financial Analysis, astronomy, and molecular Biology. Decision tree are the basic of Several Commercial rule induction System. In this research work we will use various data mining based decision tree algorithm like C4.5, SimpleCART and Random tree.

We have used ensemble technique to combine the two are more models. The main motive of assembling models to develop a robust model and improve the accuracy of models.in this research work ,ensemble the C4.5, SimpleCART and Random tree to achieve the better classification accuracy.

We have also used feature selection techniques to computationally improve the performance of model. The main motive of feature selection technique is reduce the feature subset using remove the irrelevant feature subset from data set. In this research we have used info gain and gain ratio feature selection technique [1].

## III. DATA SET

NSL- KDD is collected from UCI [6] repository and one of the publicly available data set for the evaluation of intrusion detection system which is solving some of the inherent problems of the KDD'99 data set. One of the most important efficiencies in the KDD data set is the huge number of redundant records, which causes the learning algorithms to be biased towards the frequent records, and thus prevent them from learning infrequent records which are usually more harmful to networks such as U2R and R2L attacks. This research work have used 25192 records of NSL-KDD data set .In this research work, we have used two types of NSL-KDD data set : one is binary class and second is multiclass data set. Binary class contains normal and attack types of samples while multiclass data set contains one type of normal and four type of attacks data Like DoS, R2L, U2R and Probe. All the features of NSL-KDD data set same as features of KDD99 data set.

## IV. EXPERIMENT RESULTS

In this experiment, we have decision tree based classification techniques to develop the IDS. We have used NSL-KDD data set and applied on various data mining techniques with different data partitions. When we focus on the individual models, then C4.5 model gives better classification accuracy as 99.52% of accuracy as better accuracy compare to other individual models like SimpleCart and Random tree as shown in table 1. We have also proposed ensemble model which is combination of C4.5, SimpleCART and Random tree which give better accuracy with all data partition. The proposed ensemble model gives 99.70% as best IDS to classify the various types of attacks. Table 1 shows that classification accuracy of various models with different data partitions.

Table 1.1 Accuracy of models with different data partition

| Models | 60-40 % | 70-30 % | 80-20 % |
|---|---|---|---|
| C4.5 | 99.39 | 99.51 | 99.52 |
| Simple CART | 99.32 | 99.39 | 99.50 |
| Randam Tree | 99.19 | 99.14 | 99.34 |
| C4.5 + SimpleCART+ Random Tree | 99.50 | 99.61 | 99.70 |

Feature selection is very important role to improve the performance of model.We have used two feature selection techniques like info gain and gain ratio to select the relevant feature from data set. Table 2 shows that accuracy of proposed model with different feature subset in case of info gain feature selection. Our proposed model gives satisfactory results as 99.80% with 11 features in case of info gain feature selection.

Similarly, we have also used gain ratio feature selection technique to competently improve the performance of model. Table 3 shows that accuracy of model with different feature subset. We have achieved best classification accuracy with 16 feature subset in case of gain ratio feature selection.



©IJRASET: All Rights are Reserved

Volume 5 Issue IV, April 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Table 2 : Accuracy of best model (C4.5 +SimpleCART+ Random Tree) in best partition (80-20%) with feature different subset with Info Gain

| Number of feature Subset | Accuracy |
|---|---|
| 37 | 99.72 |
| 31 | 99.72 |
| 23 | 99.78 |
| 16 | 99.78 |
| 13 | 99.78 |
| 11 | 99.80 |

Table 3: Accuracy of best model (C4.5 +SimpleCART+Random Tree) in best partition (80-20%) with feature different subset with Gain Ratio

| Number of feature Subset | Accuracy |
|---|---|
| 37 | 99.76 |
| 27 | 99.76 |
| 21 | 99.76 |
| 18 | 99.82 |
| 16 | 99.80 |

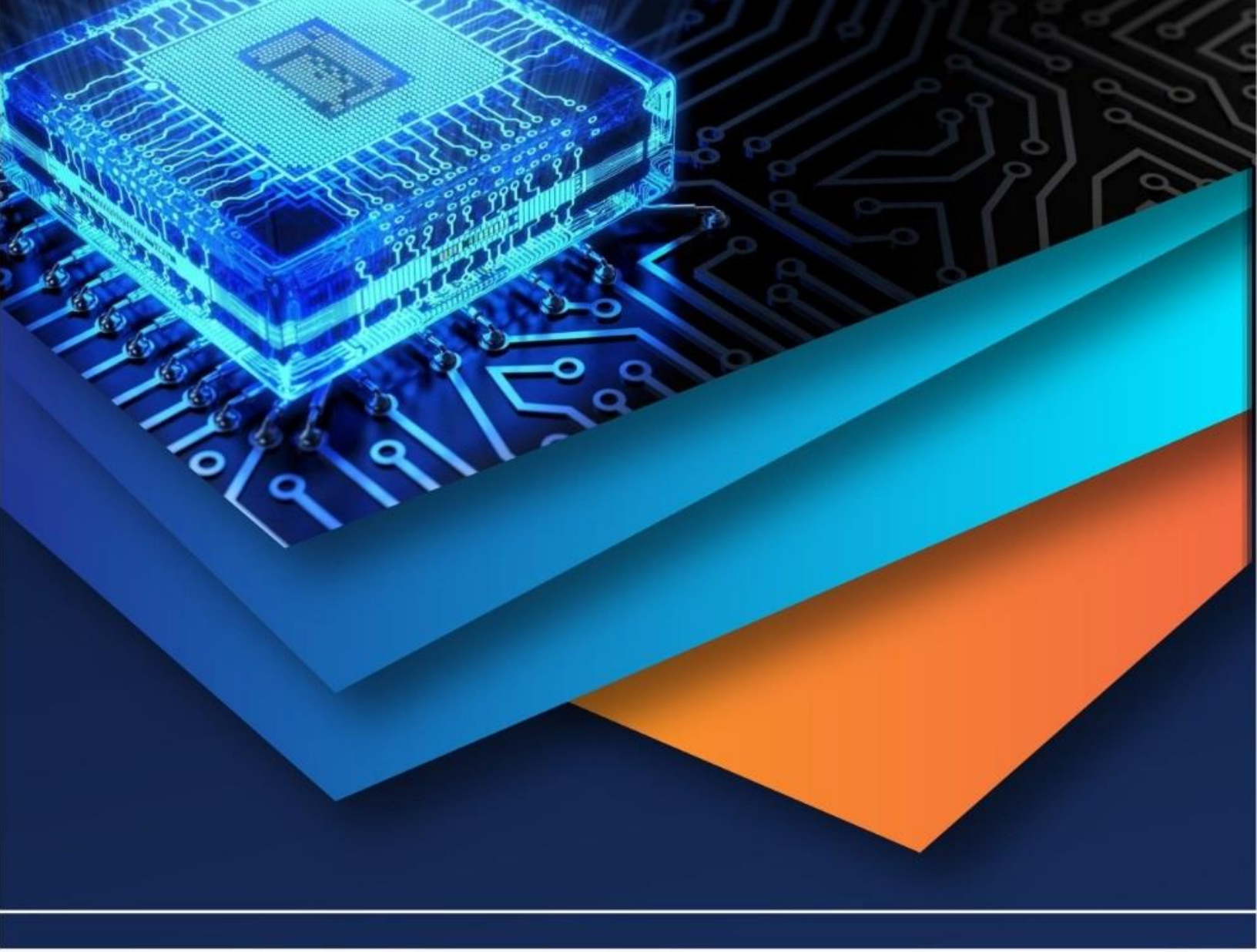
## V. CONCLUSIONS

Intrusion detection is a mechanism that is used to prevent the unauthorized person to access the information. The main aim of this research work is to develop the robust classifier as IDS which classify the various types of attacks and normal data. In this research work, focus on the decision tree like C4.5, SimpleCART and Random tree to develop the robust IDS. The proposed IDS is robust and efficient classifier for classification of attacks. We have also used feature selection techniques which contributes to improve the classification accuracy with less computational time.

## REFERENCES

[1] Han, J. and Kamber, M., Data Mining Concepts and Techniques. Morgan Kaufmann, San Francisco. 2nd ed., ISBN: 13: 978-1-55860-901-3,2006.
[2] J. Jabez and B. Muthukumar (2015),Intrusion Detection System (IDS): Anomaly Detection Using Outlier Detection Approach, International Conference on Intelligent Computing, Communication & Convergence organized by Inderscience Institute of Management and Technology Bhubaneswar, Odisha, India. Procedia Computer Science , Vol. 48 , pp. 338 – 346,2015.
[3] Y. Maleh , A. Ezzati , Y. Qasmaoui and M. Mbida, A Global Hybrid Intrusion Detection System for Wireless Sensor Networks, The 5th International Symposium on Frontiers in Ambient and Mobile Systems (FAMS 2015).,Procedia Computer Science, Vol. 52, pp.1047 – 1052,2015
[4] N. Sharma and B. Gaur, An approach for efficient intrusion detection for KDD dataset: a survey, International Journal of Advanced Technology and Engineering Exploration, Vol 3(18) , pp. 72-76, 2016.
[5] J. Reddy , Bharti., S. K. Mishra and K. S. Babu , Honeypot-Based Intrusion Detection System: A Performance Analysis, 3rd International Conference on Computing for Sustainable Global Development, 16th - 18th March, 2016 BharatiVidyapeeth's Institute of Computer Applications and Management (BVICAM) New Delhi (INDIA)..3947-3951,2016.
[6] UCI Repository of Machine Learning Databases, University of California at Irvine, Department of Computer Science. Available: http://www.ics.uci.edu/~mlearn/databases/ (Browsing date: 16 feb 2017).


©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)