JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

INTRUSION DETECTION USING PCA

1.P.Nitesh, 2.S.Bhavick, 3.A.GopiKrishna, 4.P.Krishna, 5.K.Spandana Kumari
B.tech student, B.tech student, B.tech student, B.tech student, Assistant Professor
DEPARTMENT OF CSE
VIDYA JYOTHI INSTITUTE OF TECHNOLOGY, AZIZ NAGAR, HYDERABAD.

Abstract : With the evolution in wireless communication, there are many security threats over the internet. The intrusion detection system (IDS) helps to find the attacks on the system and the intruders are detected. Previously various machine learning (ML) techniques are applied on the IDS and tried to improve the results on the detection of intruders and to increase the accuracy of the IDS . This paper has proposed an approach to develop efficient IDS by using the principal component analysis (PCA) and the random forest classification algorithm. Where the PCA will help to organise the dataset by reducing the dimensionality of the dataset and the random forest will help in classification. Results obtained states that the proposed approach works more efficiently in terms of accuracy as compared to other techniques like S VM, Naïve Bayes, and Decision Tree. The results obtained by proposed method are having the values for performance time (min) is 3.24 minutes, Accuracy rate (%) is 96.78 %, and the Error rate (%) is 0.21 %.

Keywords – IDS, Knowledge Discovery Dataset, PCA,Random Forest.

I. INTRODUCTION

Nowadays, the involvement of the internet in normal life has been increased rapidly. The internet has made a crucial place in everyone’s life. The use of the internet has become very crucial for everyone. So with the increase in the use of internet activities.Different attacks are seen on the system or the network.The attacks like black hole,grey hole,wormhole etc. are seen on the network system. These attacks are to steal the information from the system or to corrupt the data present over any system . To make misuse of the data, the intruders attack the system in various ways, some of the attacks are DoS, probe, snort, r2l etc. So to prevent the system from such attacks, the intrusion detection system was introduced. IDS keep track of attacks on the system and to prevent the system from these attacks . So to detect such attacks, the various works have done earlier by using various techniques. Here an intrusion detection system that makes use of the principal component analysis is used along with the random forest technique. Both the methods work fora special purpose, where the PCA gives the granularity in the data, and the random forest helps the classification between the nodes for attacks.

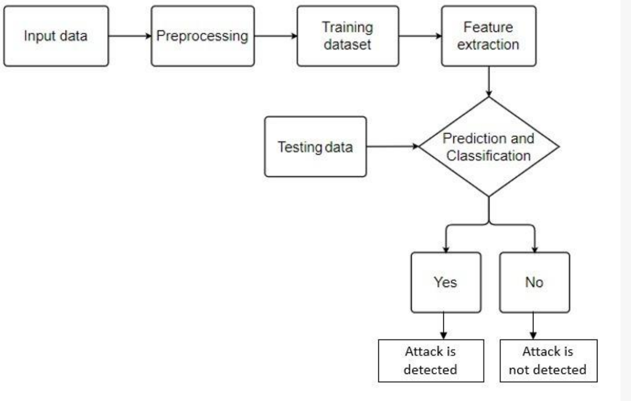


Fig:1 System architecture

•

II. SYSTEM ANALYSIS AND EXISTING SYSTEM

The systems which work over the internet suffer from various malicious activities. The major problem seen in this field is the intrusion in the system for violating the information. This intrusion is detected by creating an intrusion detection system; this system also needs to be accurate and efficient in the detection of the intruders. Various machine learning algorithms were used for intrusion detection; some of them are SVM, Naïve Bayes etc. But the results state that there may be some improvements to be done on terms of accuracy and the detection rates and the false alarm rate. Some other techniques can replace previously applied techniques such as SVM and Naïve Bay es. Also, the study states that the dataset can be improved by using some methods over it. To imp rove the quality of the input to the proposed system

III.PROPOSED SYSTEM

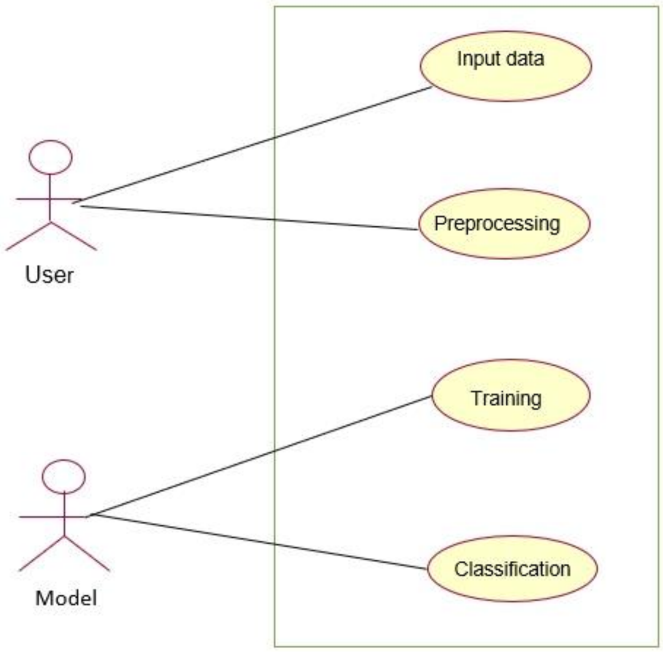
The intrusion detection system works for the improvement of the system, which is affected by the intruders. This system can do the detection of the intruders. The proposed system tries to eliminate the existing problems related to the previous work. The proposed system consists of the two methods that are principal component analysis, and the other one is the random forest. The principal component analysis is used for the reduction of the dimension of the dataset; by this method, the dataset quality will be improved as the dataset may contain correct attributes. After this, the random forest algorithm will be applied for the detection of the intruders, w hich provide both the detection rate and the false alarm rate in an improved manner as compared to SVM

IV.IMPLEMENTATION MODULE DESCRIPTION

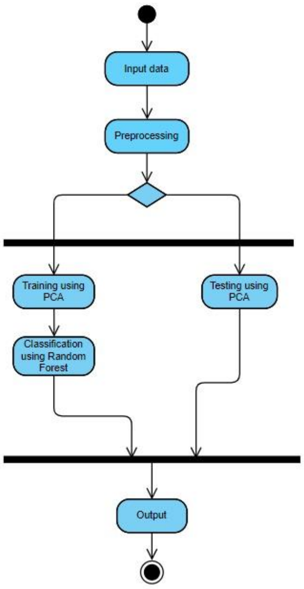
- Firstly, the Input data is collected from the sources like Kaggle and KDD.
- The collected input data is pre-processed using several pre-processing techniques which converts network into series of observations.
- After pre-processing the datasets are obtained further improved the quality of datasets using PCA (Principal Component Analysis). The datasets are trained.
- From trained datasets the improved quality datasets are extracted and then the datasets are predicted and classified using Random Forest algorithm.
- Even the Testing data is also Predicted and Classified using Random Forest and finally the required output is obtained, it de tects whether there is an attack on the system or not.

V.SYSTEM DESIGN IMPLEMENTATION

1)USECASE DIAGRAM:



2)ACTIVITY DIAGRAM:



VI.RESULTS AND DISCUSSION

1.First the user should enter the login details.

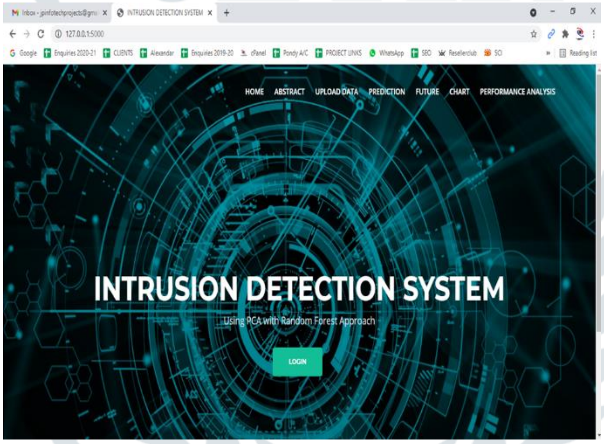


Fig:2 Home page

2.The user should enter the dataset .

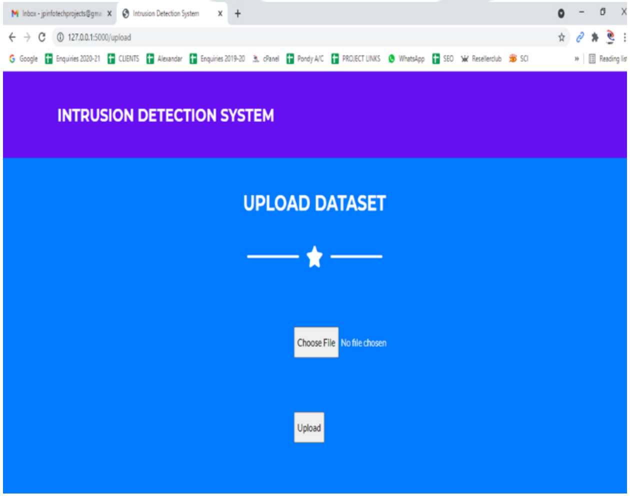
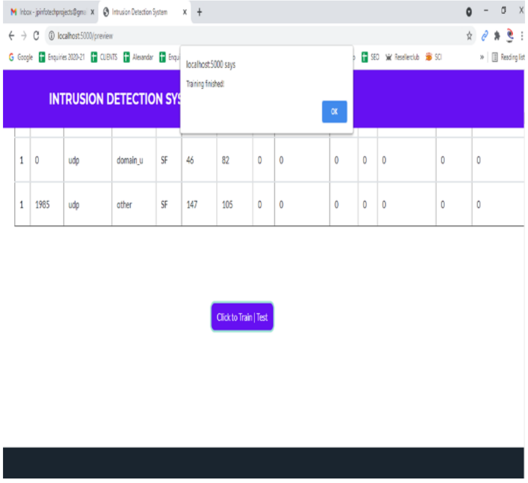


Fig:3 Page after login

3.The output is displayed here the user clicks on train to train the data.



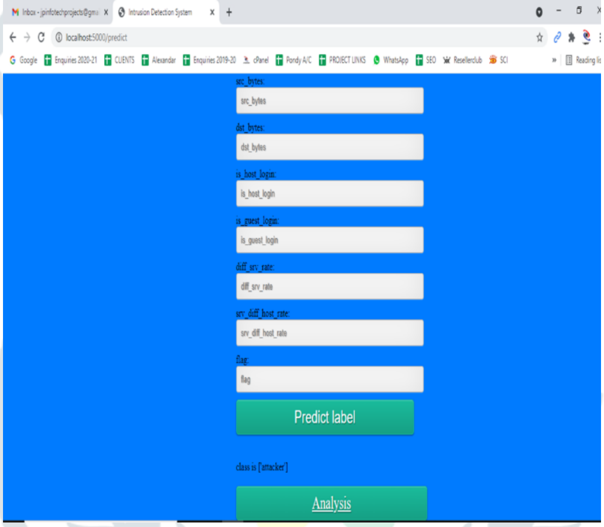
localhost:5000 says

Training finished!

1	0	udp	domain_u	SF	46	82	0	0	0	0	0	0	0
1	1985	udp	other	SF	147	105	0	0	0	0	0	0	0

Fig:4 Ouput for the above page details

4.The next page shows the trained data.



src_bytes: src_bytes

dst_bytes: dst_bytes

is_host_login: is_host_login

is_guest_login: is_guest_login

diff_srv_rate: diff_srv_rate

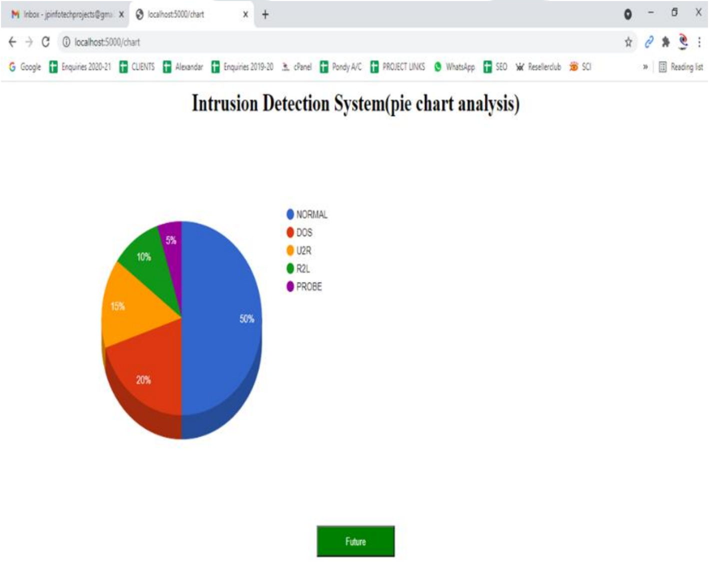
srv_diff_host_rate: srv_diff_host_rate

flag: flag

class is ['attacker']

Fig:5 Output of trained data

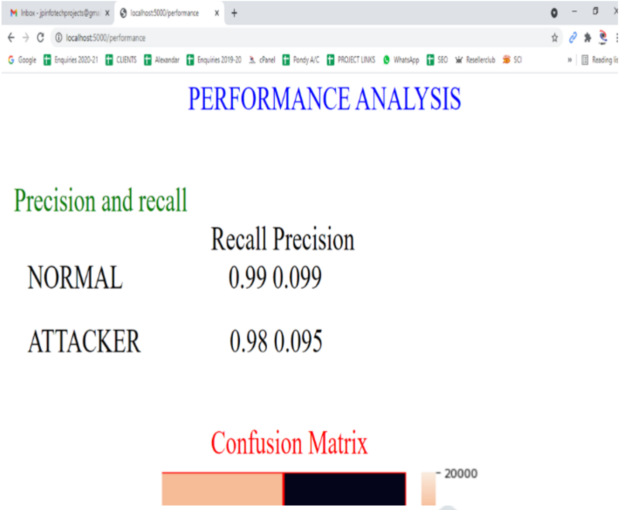
5.The output is displayed as what type of attack happened at what percent.



Intrusion Detection System(pie chart analysis)

Fig:6: Output of the analysis

6.The next page is the performance analysis page.



PERFORMANCE ANALYSIS

Precision and recall

	Recall	Precision
NORMAL	0.99	0.099
ATTACKER	0.98	0.095

Confusion Matrix

Fig:7 Performance Analysis

7.Confusion Matrix.

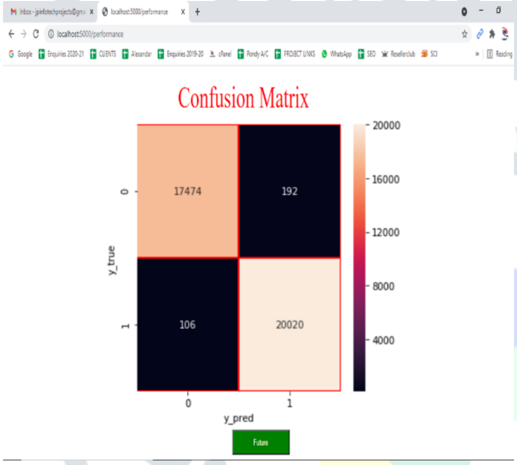


Fig:8 Confusion Matrix

VII. CONCLUSION

As the involvement of the systems over the internet increasing rapidly, the security concerns have also seen. The proposed approach deals with the detection of intruders over the internet efficiently. The proposed algorithm has performed well as compared to the previously applied algorithms such as SVM, Naïve Bayes, and Decision Tree. The detection rates and the false error rates can be improved at a great extent by the proposed approach. The dataset used here is the knowledge discovery dataset. The results obtained by our proposed method having the values for Performance time (min)is 3.24 minutes, Accuracy rate (%) is 96.78 %, and the Error rate (%) is 0.21 %.

V111. REFERENCES

1. JafarAbo Nada; Mohammad Rasmi Al-Mosa, 2018 International Arab Conference on Information Technology (ACIT), A Proposed Wireless Intrusion Detection Prevention and Attack System

2. Kinam Park; Youngrok Song; Yun-Gyung Cheong, 2018 IEEE Fourth International Conference on Big Data Computing Service and Applications (BigData Service), Classification of Attack Types for Intrusion Detection Systems Using a Machine Learning Algorithm

3. S. Bernard, L.Heutte and S. Adam "On the Selection of Decision Trees in Random Forests" Proceedings of International Joint Conference on Neural Networks, Atlanta, Georgia, USA, June 14-19, 2009, 978-1-4244-3553-1/09/$25.00 ©2009 IEEE

4. A. Tesfahun, D. Lalitha Bhaskari, "Intrusion Detection using Random Forests Classifier with SMOTE and Feature Reduction" 2013 International Conference on Cloud & Ubiquitous Computing & Emerging Technologies, 978-0-4799-2235-2/13 $26.00 © 2013 IEEE

5. Le, T.-T.-H., Kang, H., & Kim, H. (2019). The Impact of PCA-Scale Improving GRU Performance for Intrusion Detection. 2019 International Conference on Platform Technology and Service (PlatCon). Doi:10.1109/platcon.2019.8668960

6. Anish Halimaa A, Dr K.Sundarakantham: Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019) 978-1-5386-9439-8/19/$31.00 ©2019 IEEE "MACHINE LEARNING BASED INTRUSION DETECTION SYSTEM."

7. Mengmeng Ge, Xiping Fu, Naeem Syed, Zubair Baig, Gideon Teo, Antonio Robles-Kelly (2019). Deep Learning-Based Intrusion Detect ion for IoT Networks, 2019 IEEE 24th Pacific Rim International Symposium on Dependable Computing (PRDC), pp. 256-265, Japan.

8. R.Patgiri, U. Varshney, T.Akutota, and R. Kunde, "An Investigation on Intrusion Detection System Using Machine Learning" 978-1-5386-9276-9/18/$31.00 c2018IEEE.

9. Rohit Kumar Singh Gautam, Er. Amit Doegar; 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence) "An Ensemble Approach for Intrusion Detect ion System Using Machine Learning Algorithms."

10. Kazi Abu Taher, Billal Mohammed Yasin Jisan, Md. Mahbubur Rahma, 2019 International Conference on Robot ics, Electrical and Signal Processing Techniques (ICREST) "Network Intrusion Detect ion using Supervised Machine Learning Technique with Feature Selection."