International Journal of Engineering Science and Advanced Technology (IJESAT)

IJESAT (Enriching the Research)

Open Access Research Article

Volume: 23 Issue: 07

July, 2023

SEMI SUPERVISED MACHINE LEARNING APPROACHES FOR DDOS ATTACK DETECTION

Mr. K. Venkatesh [1] M. Sai Sri Vasanthi [2], K. Sri Nandini [3], K. Ram Seshu [4], M. Sravya [5]

[1] Professor, Department of CSE, Ramachandra College of Engineering, Eluru, A.P

[2,3,4,5] UG Students, Department of CSE, Ramachandra College of Engineering, Eluru, A.P

ABSTRACT

A network exploit known as a distributed denial of service (DDoS) attack is used to disable a server's regular operations and responses. During the DDoS attack the legitimate users cannot access resources available through internet and allows us to identify unusual network traffic behaviour. One of the biggest risks to the internet is thought to be this attack. Due to this issue we are using various machine learning algorithms to create a detection model in order to resolve the consequences. The unsupervised methods analyze incoming network traffic to find attacks and semi-supervised machine learning (ML) techniques for DDoS detection are provided.(Random Forest, Co-clustering, Support Vector Machine as well as K-Nearest Neighbour).With the unsupervised method, the irrelevant regular traffic data is reduced. A component of this method for DDoS detection allows for a decrease in false positive rates and an improvement in accuracy. On the other hand, the supervised portion significantly lowers the false positive rates of the unsupervised section, and the DDoS traffic is also correctly identified.

1. INTRODUCTION

DDoS attack main objective is to prevent legitimate internet users from accessing resources. The frequency and amount of network traffic delivered to the target controls the attack's intensity. Three distinct types of machine learning approaches exist. They are supervised, unsupervised, and semi-supervised learning-based DDoS detection. Combining supervised and unsupervised algorithms, semi-supervised algorithms can operate on labelled or unlabelled data sets.

For example, in the machine learning semi-supervised technique header features the entropy for the traffic data is estimated. At the unsupervised co-clustering algorithm, the input network is divided into three clusters. The traffic data occurred during the DDoS attack is classified precisely by using of random forest algorithm. The main goal is the detection and removal of irrelevant data on the basis of co clustering, estimation of entropy and the ratio of information gain[random forest algorithm] by improving the accuracy.

Artificial Intelligence, Machine Learning (ML), Pattern Recognition, Statistics, Information Theory are the most used data mining techniques for intrusion detection . Application process of data mining techniques in general and ML techniques more specifically requires five typical steps selection, preprocessing, transformation, mining, and interpretation. Despite that preprocessing and

transformation steps may be trivial for intrusion detection applications, selection, mining and interpretation steps are crucial for selecting relevant data, filtering noisy data and detecting intrusions.
The existing Machine Learning based DDoS detection approaches can be divided into three categories. Supervised ML approaches that use generated labeled network traffic datasets to build the detection model. Two major issues are facing the supervised approaches. First, the generation of labeled network traffic datasets is costly in terms of computation and time.
Without a continuous update of their detection models, the supervised machine learning approaches are unable to predict the new legitimate and attack behaviors.

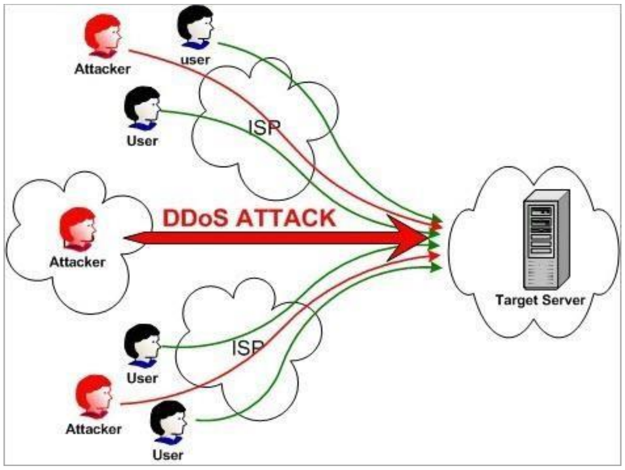


Fig1 DDoS Attack

Second, the the presence of large amount of irrelevant normal data in the incoming network traffic is noisy and reduces the performances of supervised ML classifiers. Unlike the first category, in the unsupervised approaches no labeled dataset is needed to built the detection model. The DDoS and the normal traffics are distinguished based on the analysis of their underlying distribution characteristics. However, the main drawback of the unsupervised approaches is the high false positive rates.
In the high dimensional network traffic data the distance between points becomes meaningless and tends to homogenize. This problem, known as ‘the curse of dimensionality’, prevents unsupervised approaches to accurately detect attacks . The semi-supervised ML approaches are taking advantages of both supervised and unsupervised approaches by the ability to work on labeled and unlabeled datasets.
Also, the combination of supervised and unsupervised approaches allows to increase accuracy and decreases the false positive rates. However, semi-supervised approaches are also challenged by the drawbacks of both approaches. Hence, the semi-supervised approaches require a sophisticated implementation of its components in order to overcome the drawbacks of supervised and unsupervised approaches.

2. LITERATURE SURVEY

Ample research is carried out on the analysis and prediction of sales using various techniques. There are many methods proposed to do so by various researchers. In this section, we will summarize a few of the machine learning approaches.

- ❖ Several approaches have been proposed for detecting DDoS attack. Information theory and machine learning are the most common techniques used in the literature. This section summarizes some of the recent works in DDoS detection. Akilandeswari V. et al. have used a Probabilistic Neural Network to discriminate flash crowd events from DDoS attacks. The method achieves high DDoS detection accuracy with lower false positives rates.

- ❖ Similarly, Ali S. B. et al. have proposed an innovative ensemble of Sugeno type adaptive neuro-fuzzy classifiers for DDoS detection using an effective boosting technique named Marliboost. The proposed technique was tested on the NSL-KDD dataset and have achieved good performance.

- ❖ Mohiuddin A. and Abdun Naser M have proposed an unsupervised approach for DDoS detection based on the co-clustering algorithm. The authors have extended the co-clustering algorithm to handle categorical attributes. The approach was tested on the KDD cup 99 dataset and achieved good performance.

- ❖ Alan S. et al. [19] have proposed a DDoS Detection Mechanism based on ANN (DDMA). The authors used three different topologies of the MLP for detecting three types of DDoS attacks based on the background protocol used to perform each attack namely TCP, UDP and ICMP. The mechanism detects accurately known and unknown, zero day, DDoS attacks.

- ❖ Similarly, Boro D. et al. [20] have presented a defense system referred to as DyProSD that combines both the merits of feature-based and statistical approach to handle DDoS flooding attack. The statistical module marks the suspicious traffic and forwards to an ensemble of classifiers for ascertaining the traffic as malicious or normal.

- ❖ Recently, Van Loi C. [21] proposed a novel one class learning approach for network anomaly detection based on combining auto-encoders and density estimation. Authors have tested their method on the NSL-KDD dataset, and obtained satisfactory results. Mohamed I. et al. [22] have proposed a supervised DoS detection method based on a feed- forward neural network. This method consists of three major steps: (1) Collection of the incoming network traffic, (2) selection of relevant features for DoS detection using an unsupervised Correlation-based Feature Selection (CFS) method, (3) classification of the incoming network traffic into DoS traffic or normal traffic.

3. EXISTING SYSTEM

- The first phase of their approach consists of dividing the incoming network traffic into three type of protocols TCP, UDP or Other. Then classifying it into normal or anomaly traffic.
- In the second stage a multi-class algorithm classify the anomaly detected in the first phase to identify the attacks class in order to choose the appropriate intervention.

- The DDoS detection approaches in the literature are under two main categories unsupervised approaches and supervised approaches.
- Depending on the benchmark datasets used, unsupervised approaches often suffer from high false positive rate and supervised approach cannot handle large amount of network traffic data and their performances are often limited by noisy and irrelevant network data.

DISADVANTAGES:

- While some approaches have been used, they have not been well received.
- The main disadvantages are less accuracy rate, classification issue, false positive rates and more network traffic data.

4. PROPOSED SYSTEM

□ The main aim of combining algorithms used in the proposed approach is to reduces noisy and irrelevant network traffic data before preprocessing and classification stages for DDoS detection while maintaining high performance in terms of accuracy, false positive rate and running time, and low resources usage.

□ Our approach starts with estimating the entropy of the FSD features over a time-based sliding window.

□ When the average entropy of a time window exceeds its lower or upper thresholds the co- clustering algorithm split the received network traffic into three clusters.

□ Entropy estimation over time sliding windows allows to detect abrupt changes in the incoming network traffic distribution which are often caused by DDoS attacks. Incoming network traffic within the time windows having abnormal entropy values is suspected to contain DDoS traffic.

ADVANTAGES:

□ The major advantage is that it aids in the removal of extraneous normal traffic data.

DDoS attack detection allows the reduction of false positive rates and also increasing the accuracy compared to previous accuracy values.

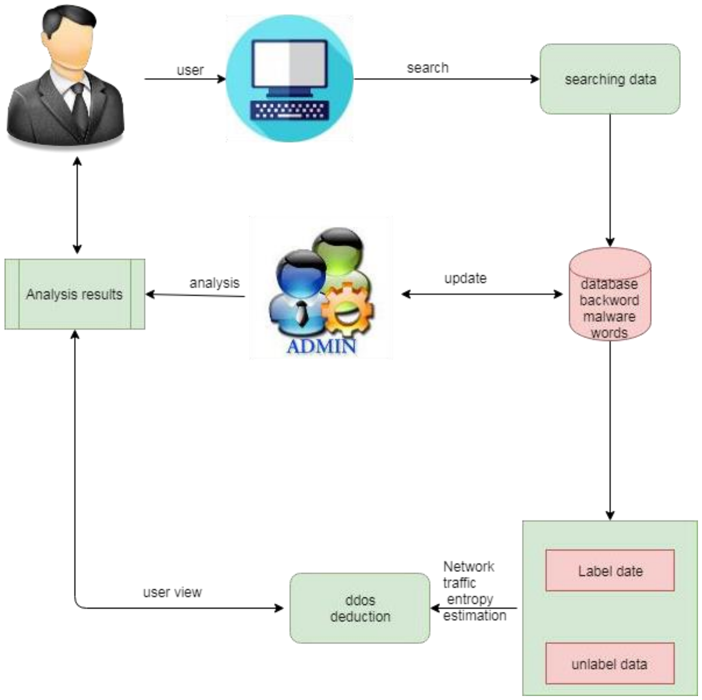


Fig: 1 System Architecture

Adopting a supervised ensemble ML classifiers based on the Random Forest algorithm to accurately classify the anomalous traffic and to reduce the false positive rates. Combining both previous algorithms in a sophisticated semi-supervised approach for DDoS detection. This allows to achieve good DDoS detection performance compared to the state-of-the-art DDoS detection methods. The unsupervised part of our approach allows to reduce the irrelevant and noisy normal traffic data, hence reducing false positive rates and increasing accuracy of the supervised part. Whereas, the supervised part allows to reduce the false positive rates of the unsupervised part and to accurately classify the DDoS traffic.

5. RESULTS

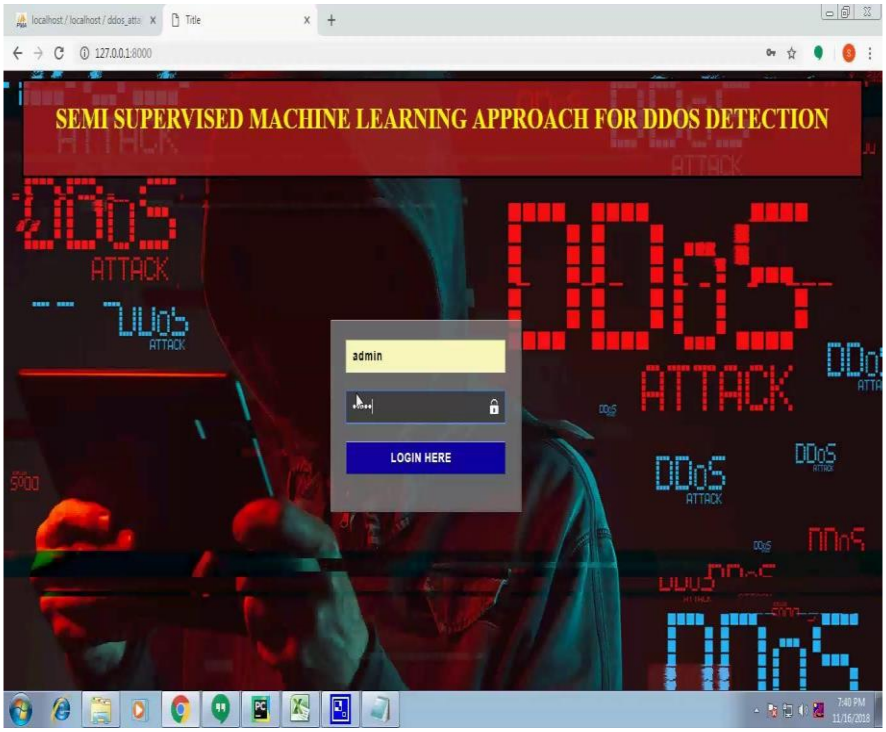
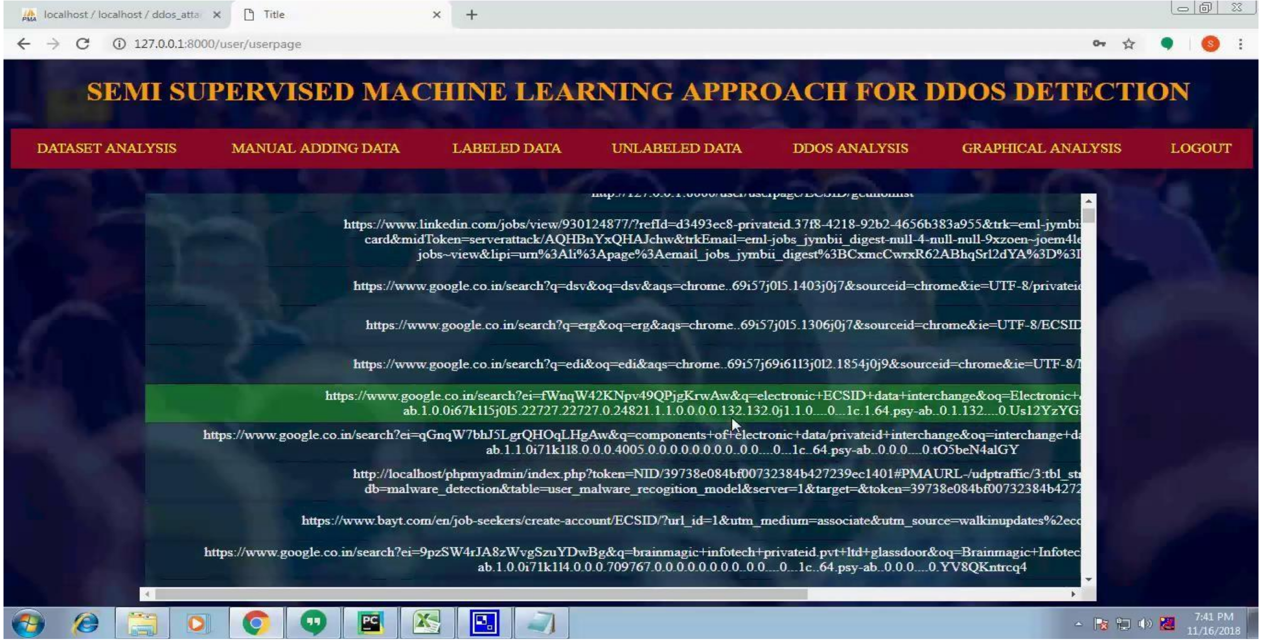
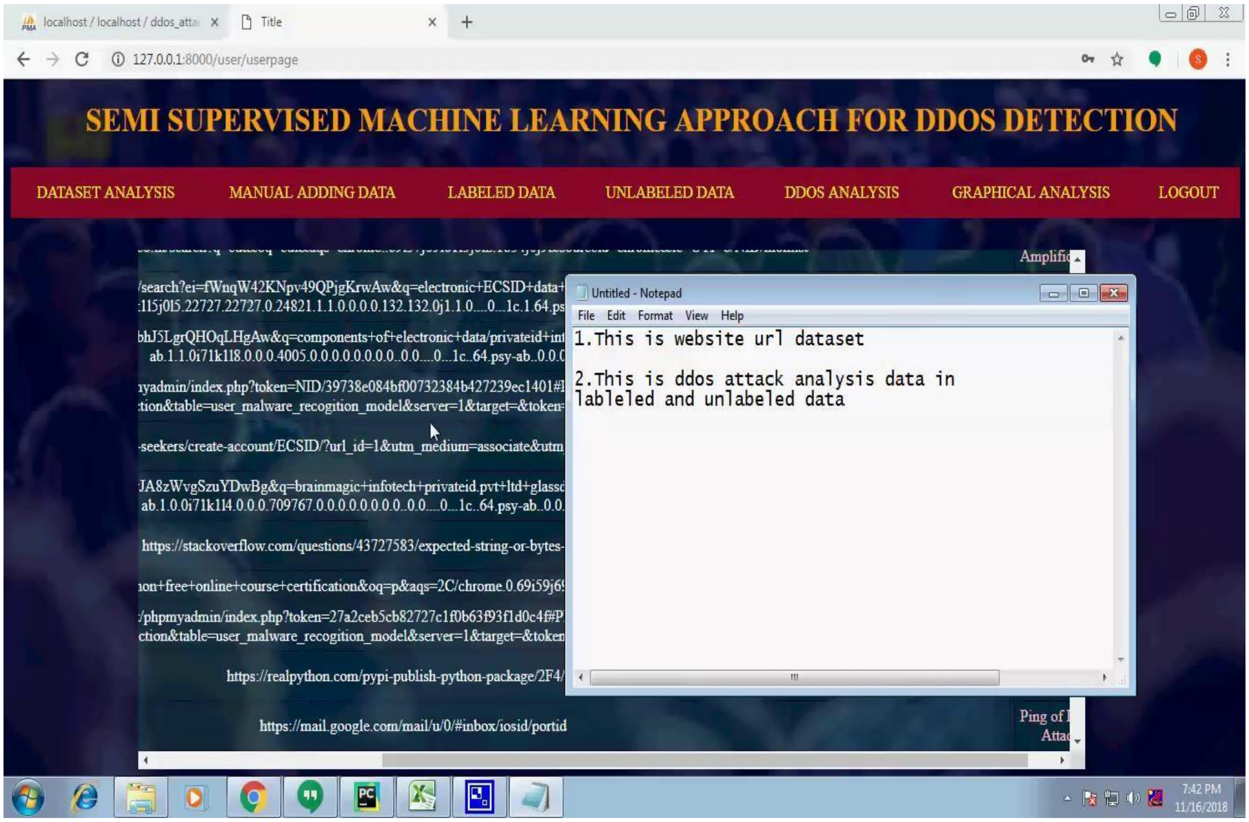

SEMI SUPERVISED MACHINE LEARNING APPROACH FOR DDOS DETECTION
admin
LOGIN HERE

Fig.2 Login Page



```
https://www.linkedin.com/jobs/view/930124877/?refId=d3493ec8-privateid.37f8-4218-92b2-4656b383a955&trk=eml-jymbi
card&midToken=serverattack/AQHBnYxQHAJchw&trkEmail=eml-jobs_jymbii_digest-null-4-null-null-9xzoen~joem4le
jobs~view&lipi=urn%3Ali%3Apage%3Aemail_jobs_jymbii_digest%3BCxmcCwrxR62ABhqSrl2dYA%3D%3I

https://www.google.co.in/search?q=dsv&oq=dsv&aqs=chrome..69i57j015.1403j0j7&sourceid=chrome&ie=UTF-8/privateid

https://www.google.co.in/search?q=erg&oq=erg&aqs=chrome..69i57j015.1306j0j7&sourceid=chrome&ie=UTF-8/ECSID

https://www.google.co.in/search?q=edi&oq=edi&aqs=chrome..69i57j69i6113j012.1854j0j9&sourceid=chrome&ie=UTF-8/I

https://www.google.co.in/search?ei=fWnqW42KNpv49QPjgKrwAw&q=electronic+ECSID+data+interchange&oq=Electronic+
ab.1.0.0i67k115j015.22727.22727.0.24821.1.1.0.0.0.0.132.132.0j1.1.0....0...1c.1.64.psy-ab..0.1.132....0.Us12YzYGI
https://www.google.co.in/search?ei=qGnqW7bhJ5LgrQHOqLHgAw&q=components+of+electronic+data/privateid+interchange&oq=interchange+da
ab.1.1.0i71k118.0.0.0.4005.0.0.0.0.0.0.0.0.0..0.0....0...1c..64.psy-ab..0.0.0....0.tO5beN4alGY
http://localhost/phpmyadmin/index.php?token=NID/39738e084bf00732384b427239ec1401#PMAURL-/udptraffic/3:tbl_str
db=malware_detection&table=user_malware_recogition_model&server=1&target=&token=39738e084bf00732384b4272

https://www.bayt.com/en/job-seekers/create-account/ECSID/?url_id=1&utm_medium=associate&utm_source=walkinupdates%2ecc

https://www.google.co.in/search?ei=9pzSW4rJA8zWvgSzuYDwBg&q=brainmagic+infotech+privateid.pvt+ltd+glassdoor&oq=Brainmagic+Infotec
ab.1.0.0i71k114.0.0.0.709767.0.0.0.0.0.0.0.0.0..0.0....0...1c..64.psy-ab..0.0.0....0.YV8QKntrcq4
```

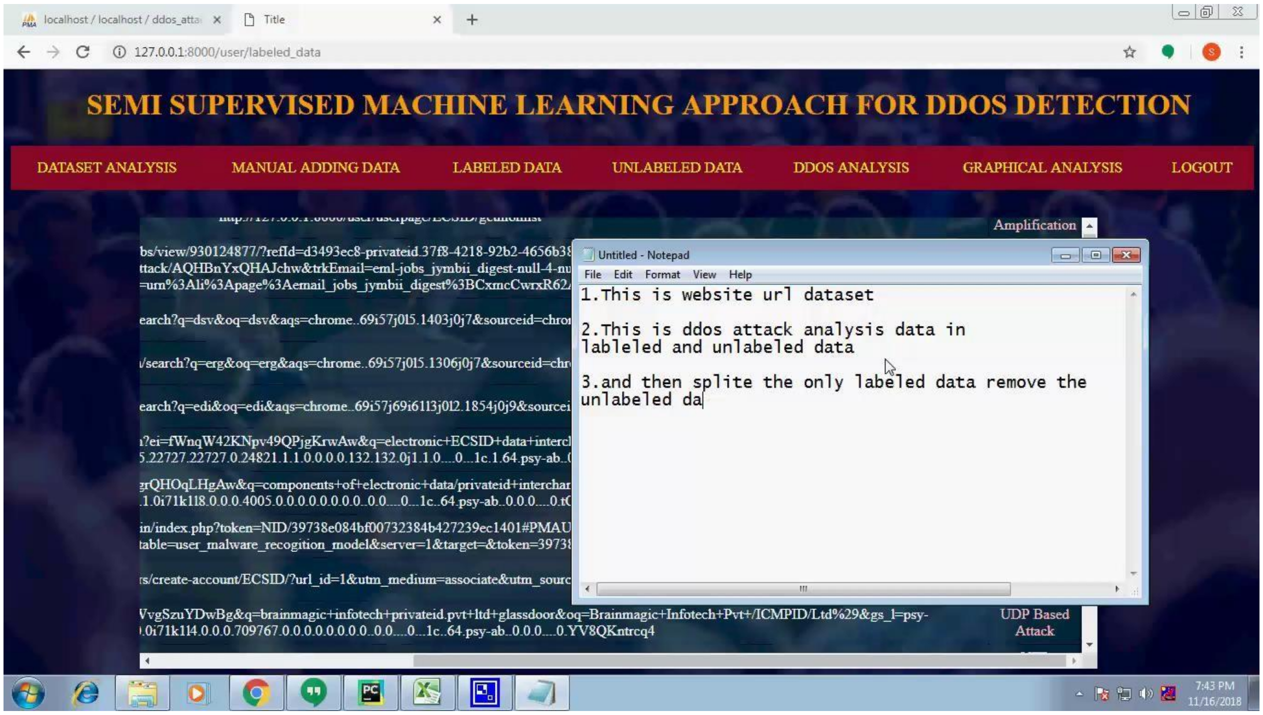
Fig.3 URL Link



```
/search?ei=fWnqW42KNpv49QPjgKrwAw&q=electronic+ECSID+data+
:115j015.22727.22727.0.24821.1.1.0.0.0.0.132.132.0j1.1.0....0...1c.1.64.ps

bhJ5LgrQHOqLHgAw&q=components+of+electronic+data/privateid+int
  ab.1.1.0i71k118.0.0.0.4005.0.0.0.0.0.0.0.0..0.0....0...1c..64.psy-ab..0.0.0

nyadmin/index.php?token=NID/39738e084bf00732384b427239ec1401#
:tion&table=user_malware_recogition_model&server=1&target=&token=

-seekers/create-account/ECSID/?url_id=1&utm_medium=associate&utm

JA8zWvgSzuYDwBg&q=brainmagic+infotech+privateid.pvt+ltd+glassd
 ab.1.0.0i71k114.0.0.0.709767.0.0.0.0.0.0.0.0..0.0....0...1c..64.psy-ab..0.0.

 https://stackoverflow.com/questions/43727583/expected-string-or-bytes-

ion+free+online+course+certification&oq=p&aqs=2C/chrome.0.69i59j6

:/phpmyadmin/index.php?token=27a2ceb5cb82727c1f0b63f93f1d0c4f#P
ction&table=user_malware_recogition_model&server=1&target=&token

          https://realpython.com/pypi-publish-python-package/2F4/

               https://mail.google.com/mail/u/0/#inbox/iosid/portid
```

```
1.This is website url dataset

2.This is ddos attack analysis data in
labl̀eled and unlabeled data
```

Fig.4 Labeled and Unlabeled Data



```
bs/view/930124877/?refId=d3493ec8-privateid.37f8-4218-92b2-4656b38
ttack/AQHBnYxQHAJchw&trkEmail=eml-jobs_jymbii_digest-null-4-nu
=urn%3Ali%3Apage%3Aemail_jobs_jymbii_digest%3BCxmcCwrxR62
earch?q=dsv&oq=dsv&aqs=chrome..69i57j0l5.1403j0j7&sourceid=chro
/search?q=erg&oq=erg&aqs=chrome..69i57j0l5.1306j0j7&sourceid=chr
earch?q=edi&oq=edi&aqs=chrome..69i57j69i61l3j0l2.1854j0j9&sourcei
1?ei=fWnqW42KNpv49QPjgKrwAw&q=electronic+ECSID+data+intercl
5.22727.22727.0.24821.1.1.0.0.0.0.132.132.0j1.1.0....0...1c.1.64.psy-ab..
grQHOqLHgAw&q=components+of+electronic+data/privateid+interchar
.1.0i71k1l8.0.0.0.4005.0.0.0.0.0.0.0.0.0..0.0....0...1c..64.psy-ab..0.0.0....0.t
in/index.php?token=NID/39738e084bf00732384b427239ec1401#PMAU
table=user_malware_recogition_model&server=1&target=&token=3973
rs/create-account/ECSID/?url_id=1&utm_medium=associate&utm_sourc
VvgSzuYDwBg&q=brainmagic+infotech+privateid.pvt+ltd+glassdoor&oq=Brainmagic+Infotech+Pvt+/ICMPID/Ltd%29&gs_l=psy-
.0i71k1l4.0.0.0.709767.0.0.0.0.0.0.0.0..0.0....0...1c..64.psy-ab..0.0.0....0.YV8QKntrcq4
```

Amplification

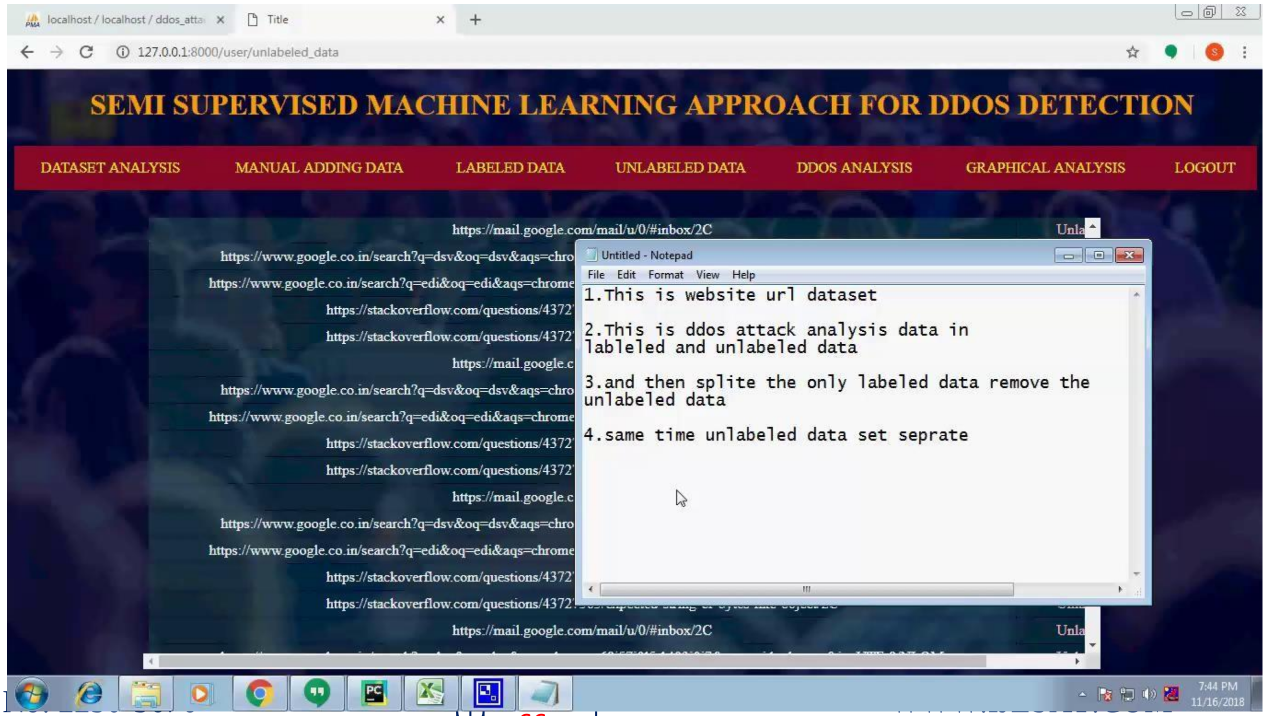
UDP Based Attack

1.This is website url dataset

2.This is ddos attack analysis data in labeled and unlabeled data

3.and then splite the only labeled data remove the unlabeled da

Fig.5 Splitting Labeled data



```
https://mail.google.com/mail/u/0/#inbox/2C                                   Unla
https://www.google.co.in/search?q=dsv&oq=dsv&aqs=chro
https://www.google.co.in/search?q=edi&oq=edi&aqs=chrome
https://stackoverflow.com/questions/4372
https://stackoverflow.com/questions/4372
https://mail.google.c
https://www.google.co.in/search?q=dsv&oq=dsv&aqs=chro
https://www.google.co.in/search?q=edi&oq=edi&aqs=chrome
https://stackoverflow.com/questions/4372
https://stackoverflow.com/questions/4372
https://mail.google.c
https://www.google.co.in/search?q=dsv&oq=dsv&aqs=chro
https://www.google.co.in/search?q=edi&oq=edi&aqs=chrome
https://stackoverflow.com/questions/4372
https://stackoverflow.com/questions/4372
https://mail.google.com/mail/u/0/#inbox/2C                                   Unla
```

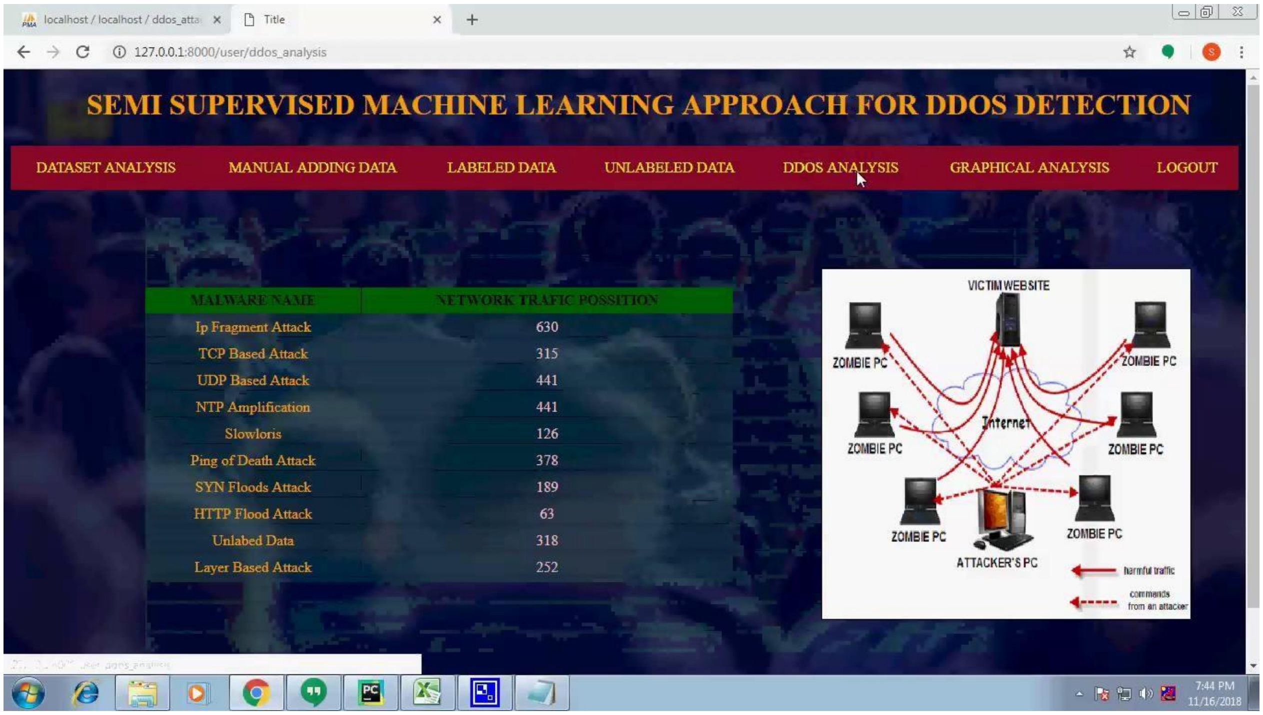
1.This is website url dataset

2.This is ddos attack analysis data in labeled and unlabeled data

3.and then splite the only labeled data remove the unlabeled data

4.same time unlabeled data set seprate

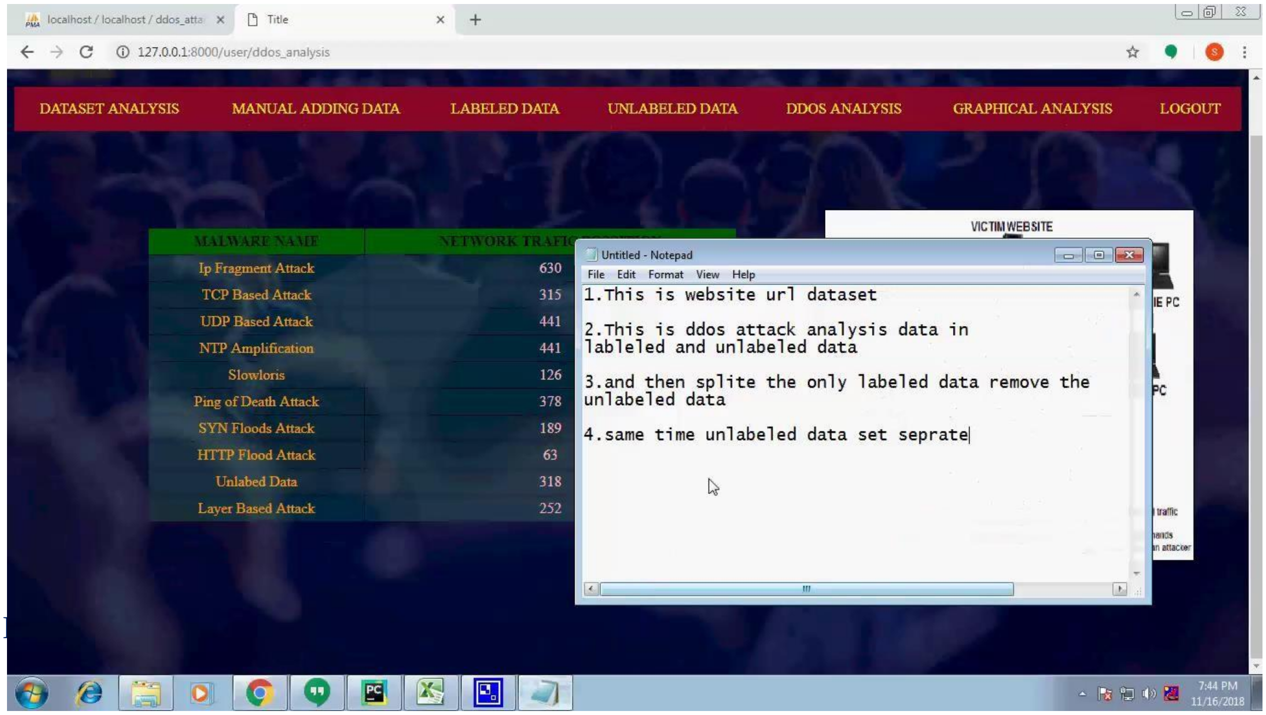
Fig..6 Separate Unlabeled Dataset



SEMI SUPERVISED MACHINE LEARNING APPROACH FOR DDOS DETECTION

MALWARE NAME	NETWORK TRAFIC POSSITION
Ip Fragment Attack	630
TCP Based Attack	315
UDP Based Attack	441
NTP Amplification	441
Slowloris	126
Ping of Death Attack	378
SYN Floods Attack	189
HTTP Flood Attack	63
Unlabed Data	318
Layer Based Attack	252

Fig.7 Network Traffic Position



MALWARE NAME	NETWORK TRAFIC POSSITION
Ip Fragment Attack	630
TCP Based Attack	315
UDP Based Attack	441
NTP Amplification	441
Slowloris	126
Ping of Death Attack	378
SYN Floods Attack	189
HTTP Flood Attack	63
Unlabed Data	318
Layer Based Attack	252

```
1.This is website url dataset

2.This is ddos attack analysis data in
lableled and unlabeled data

3.and then splite the only labeled data remove the
unlabeled data

4.same time unlabeled data set seprate
```

Fig.8 Data Seperation

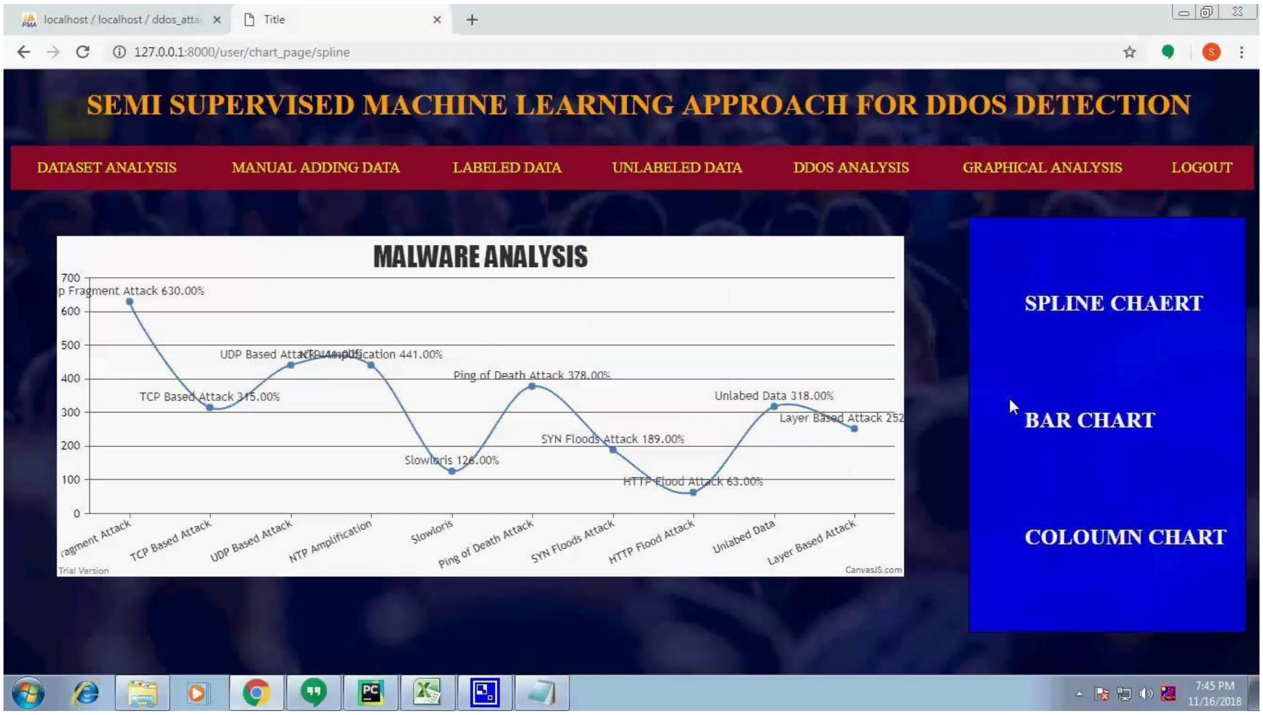


Fig.9 Spline Chart

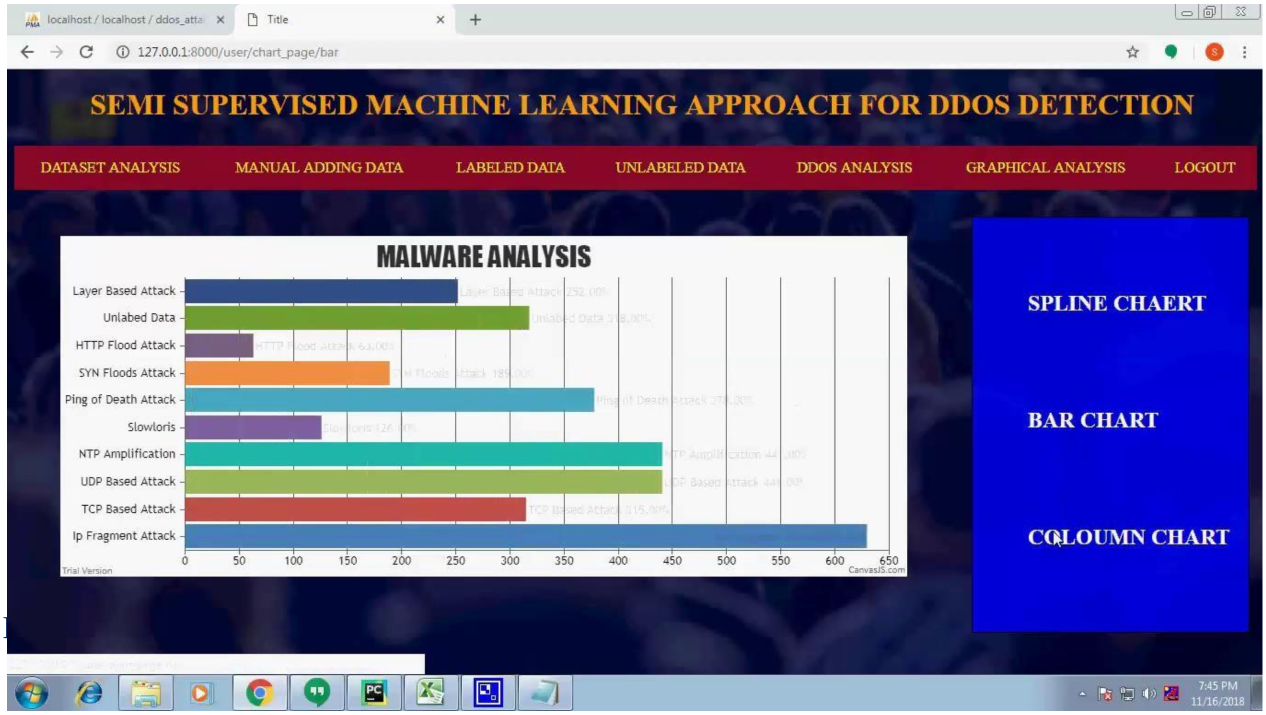


Fig.10 Column Chart

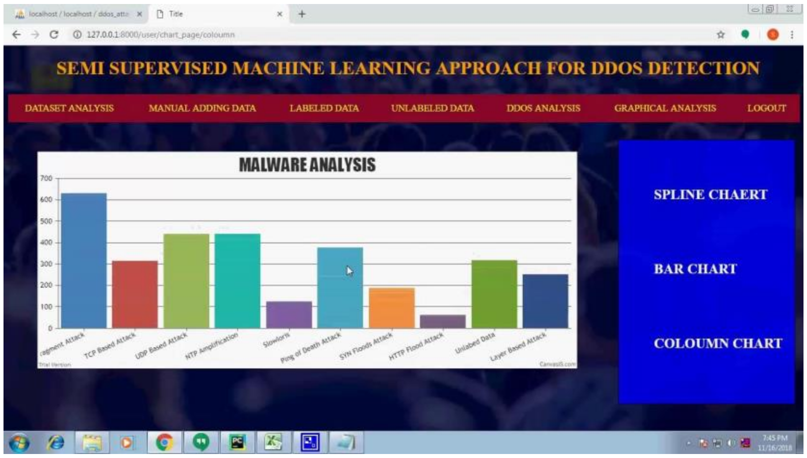
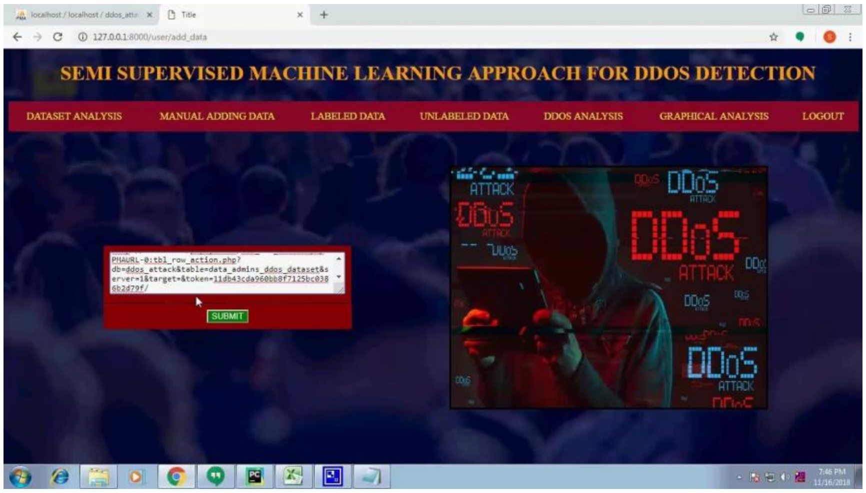


Fig.11 Bar Chart



```
PHAURL-0:tbl_row_action.php?
db=ddos_attack&table=data_admins_ddos_dataset&s
erver=1&target=&token=11db43cda960bb8f7125bc038
6b2d79f/
```

Fig.12 Adding of Manual Data

6. CONCLUSION

In recent years, DDoS attacks have become increasingly common and sophisticated, making it challenging for traditional detection methods to keep up with the evolving threat landscape. As a result, semi-supervised learning approaches have emerged as a promising technique for detecting DDoS attacks in network traffic.Semi-supervised learning is an approach that combines labeled and unlabeled data to improve the accuracy of a model while reducing the need for manual labeling. This approach has been applied in the context of DDoS detection, where it has shown promising results.The use of semi-supervised learning algorithms such as Random Forest, Co-clustering, Support Vector Machine, and K-Nearest Neighbors has been shown to be effective in detecting DDoS attacks in network traffic. These algorithms can be used in combination with both labeled and unlabeled data to improve the accuracy of DDoS attack detection while reducing the need for manual labeling. The performance of semi-supervised learning algorithms for DDoS detection has been evaluated on various datasets, and the results show that these approaches can achieve high accuracy rates.Overall, the use of semi-supervised learning for DDoS detection holds great promise for improving the accuracy of detection and reducing the need for manual labeling of data. However, there are still challenges that need to be addressed, such as the selection of appropriate algorithms and features, the quality and quantity of labeled and unlabeled data, and the need for real- time detection in high-speed networks. Nonetheless, as the threat landscape continues to evolve, the use of semi-supervised learning for DDoS detection is likely to become increasingly important for maintaining the security and integrity of computer networks.

REFERENCES

1. Bhuyan MH, Bhattacharyya DK, Kalita JK (2015) An empirical evaluation of information metrics for low rate and high-rate ddos attack detection. Pattern RecognLett 51:1–7.
2. Akilandeswari V, Shalinie SM (2020) Probabilistic neural network-based attack traffic classification. In: 2012 fourth international conference on advanced computing (ICoAC). IEEE, pp 1–8.
3. Saied A, Overill RE, Radzik T (2019) Detection of known and unknown ddos attacks using artificial neural networks. Neuro computing 172:385–393.
4. Liu T, Wang Z, Wang H, Lu K (2019) An entropy based method for attack detection in large scale network. Int J Comput Commun Control 7(3):509–517.
5. Boro D, Bhattacharyya DK (2018) Dyprosd: a dynamic protocol specific defense for high- rate ddos flooding attacks. Microsyst Technol 23:1–19.
6. Idhammad M, Afdel K, Belouch M (2017) Dos detection method based on artificial neural networks. Int J Adv Comput Sci Appl(ijacsa) 8(4):465–471.
7. Mustapha B, Salah EH, Mohamed I (2017) A two stage classifier approach using rep tree algorithm for network intrusion detection .Int J Adv Comput Sci Appl (ijacsa) 8(6):389–394.
8. Boroujerdi AS, Ayat S (2020) A robust ensemble of neuro fuzzy classifiers for ddos attack detection. In: 2013 3rdinternational conference on computer science and network technology (ICCSNT). IEEE, pp 484– 487.
9. Ahmed M, Mahmood AN (2020) Novel approach for network traffic pattern analysis using clustering based collective anomaly detection. Ann Data Sci 2(1):111– 130.
10. Nicolau M, McDermott J et al (2021) A hybrid auto encoder and density estimation model for anomaly detection. In: International conference on parallel problem solving from nature.
11. Springer, pp717– 726.
12. Jaiganesh V., Sumathi P. and MangayarkarasiS.,"An Analysis of Intrusion Detection System using Back Propagation Neural Network",IEEE 2013 publication.
13. Han C., Yi Lv, Yang D., Hao Y., "An Intrusion Detection System Based on NeuralNetwork",2011 International Conference on Mechatronic Science, Electric Engineering and Computer, August 19-22, 2011, Jilin, China, IEEE Publication.