PREDICTION AND CLASSIFICATION OF DDOS ATTACKS USING MACHINE LEARNING

T. Sasi Vardhan[1], B. Vyshnavi[2], G. Lahari[2], V. S. S. N. Akhila[2], Y. Shravya[2]

[1]Assistant Professor, [2]UG Scholar, [1,2]Department of CSE-Cyber Security

[1,2]Malla Reddy Engineering College for Women (A), Maisammaguda, Medchal, Telangana.

ABSTRACT

A Distributed denial of Service (DDoS) attack is a non-intrusive internet attack made to take down the targeted website or slow it down by flooding the network, server or application with fake traffic. When against a vulnerable resource-intensive endpoint, even a tiny amount of traffic is enough for the attack to succeed. Distributed Denial of Service (DDoS) attacks are threats that website owners must familiarize themselves with as they are a critical piece of the security landscape. Navigating the various types of DDoS attacks can be challenging and time consuming. To help you understand what a DDoS attack is and how to prevent it, we have written the following guide.

Most types of network interfaces based on the integrated functions, steal users' personal information and start the attack operations. In particular, we consider each HTTP flow generated by mobile apps as a text document, which can be processed by natural language processing to extract text-level features. Later, the use of network traffic is used to create a useful malware detection model.

Keywords: DDoS attack, Machine learning, HTTP.

1. INTRODUCTION

In recent years, widespread adoption of the internet has resulted in to rapid advancement in information technologies. The internet is used by the general population for the purposes such as financial transactions, educational endeavours, and countless other activities. The use of the internet for accomplishing important tasks, such as transferring a balance from a bank account, always comes with a security risk. Today's web sites strive to keep their users' data confidential and after years of doing secure business online, these companies have become experts in information security. The database systems behind these secure websites store non-critical data along with sensitive information, in a way that allows the information owners quick access while blocking break-in attempts from unauthorized users.

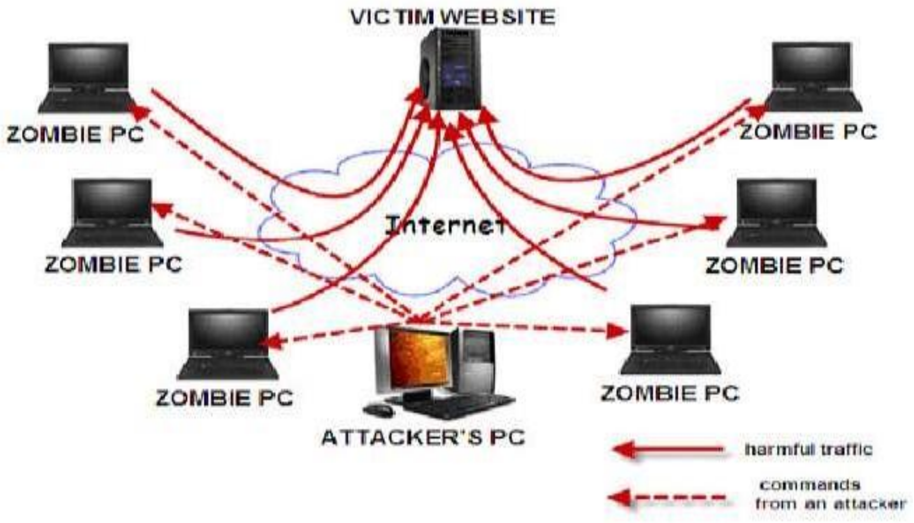


Fig. 1: DDoS architecture.

A Distributed Denial of Service (DDoS) attack is a non-intrusive internet attack made to take down the targeted website or slow it down by flooding the network,server or application with fake traffic. When against a vulnerable resource-intensive endpoint, even a tiny amount of traffic is enough for the attack to succeed.

Distributed Denial of Service (DDoS) attacks are threats that website owners must familiarize themselves with as they are a critical piece of the security landscape. Navigating the various types of DDoS attacks can be challenging and time consuming. To help you understand what a DDoS attack is and how to prevent it, we have written the following guide.

Problem Definition

There are two categories of the DDoS attack namely Direct DDoS attack and Reflection-based DDoS In the Direct DDoS attack the attacker uses the zombie hosts to flood directly the victim host with a large number of network packets. Whereas, in the Reflection based DDoS attack the attacker uses the zombie hosts to take control over a set of compromised hosts called Reflectors. The latter are used to forward a massive amount of attack traffic to the victim host. Recently, destructive DDoS attacks have brought down more than 70 vital services of Internet including Github, Twitter, Amazon, Paypal, etc.

Most of the existing ML-based DDoS detection approaches are under two categories: supervised and unsupervised. Supervised ML approaches for DDoS detection rely on availability of labeled network traffic datasets. Whereas, unsupervised ML approaches detect attacks by analyzing the incoming network traffic. Both approaches are challenged by large amount of network traffic data, low detection accuracy and high false positive rates.

The appearance of malicious apps is a serious threat to the Android platform. Most types of network interfaces based on the integrated functions, steal users' personal information and start the attack operations. In this paper, we propose an effective and automatic malware detection method using the text semantics of network traffic. In particular, we consider each HTTP flow generated by mobile apps as a text document, which can be processed by natural language processing to extract text-level features. Later, the use of network traffic is used to create a useful malware detection model. We examine the traffic flow header using N-gram method from the natural language processing (NLP). Then, we propose an automatic feature selection algorithm based on chi-square test to identify meaningful features. It is used to determine whether there is a significant association between the two variables.We propose a novel solution to perform malware detection using NLP methods by treating mobile traffic as documents. We apply an automatic feature selection algorithm based on N-gram sequence to obtain meaningful features from the semantics of traffic flows. Our methods reveal some malware that can prevent detection of antiviral scanners. In addition, we design a detection system to drive traffic to your own-institutional enterprise network, home network, and 3G / 4G mobile network. Integrating the system connected to the computer to find suspicious network behaviors.

2. LITERATURE SURVEY

Idhammad M, Afdel K, Belouch M (2017) Dos detection methodbased on artificial neural networks. Int J Adv Comput Sci Appl(ijacsa) 8(4):465–471.

DoS attack is a major Internet security problem-DoS is that lots of clients simultaneously send service requests to certain server on the internet such that this server is too busy to provide normal services for others. Attackers using legitimate packets and often changing package information, so that traditional detection methods based on feature descriptions is difficult to detect it. This paper present an artificial intelligence DoS attack detection method based on neural networks. In this method, analysis of server resources and network traffic, To training the ability of detection normal or abnormal, it have better results for detect DoS attack.

Papalexakis EE, Beutel A, Steenkiste P (2014) Network anomalydetection using co- clustering. In: Encyclopedia of social network analysis and mining. Springer, Berlin, pp 1054–1068.

Early Internet architecture design goals did not put security as a high priority. However, today Internet security is a quickly growing concern. The prevalence of Internet attacks has increased significantly, but still the challenge of detecting such attacks generally falls on the end hosts and service providers, requiring system administrators to detect and block attacks on their own. In particular, as social networks have become central hubs of information and communication, they are increasingly the target of attention and attacks. This creates a challenge of carefully distinguishing malicious connections from normal ones. Previous work has shown that for a variety of Internet attacks, there is a small subset of connection measurements that are good indicators of whether a connection is part of an attack or not. In this paper we look at the effectiveness of using two different co-clustering algorithms to both cluster connections as well as mark which connection measurements are strong indicators of what makes any given cluster anomalous relative to the total data set. We run experiments with these co-clustering algorithms on the KDD 1999 Cup data set. In our experiments we find that soft co-clustering, running on samples of data, finds consistent parameters that are strong indicators of anomalous detections and creates clusters, that are highly pure. When running hard co-clustering on the full data set (over 100 runs), we on average have one cluster with 92.44% attack connections and the other with 75.84% normal connections. These results are on par with the KDD 1999 Cup winning entry,

showing that co-clustering is a strong, unsupervised method for separating normal connections from anomalous ones. Finally, we believe that the ideas presented in this work may inspire research for anomaly detection in social networks, such as identifying spammers and fraudsters

3. PROPOSED SYSTEM

It is online sequential semi supervised ML approach for DDoS detection is implemented, A time based sliding window algorithm is used to estimate the entropy of the network header features of the incoming network traffic. 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 the approach allows to reduce the irrelevant normal traffic data for DDoS detection which allows to reduce false positive rates and increase accuracy. Whereas, the supervised part allows to reduce the false positive rates of the unsupervised part and to accurately classify DDoS traffic.

4. RESULTS

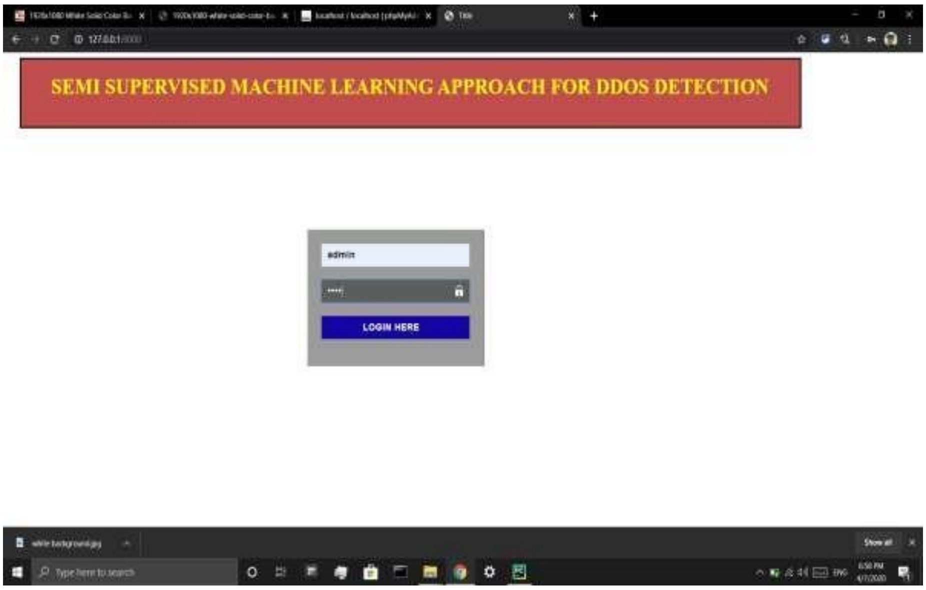
Executing consists of the processes used to complete the work defined in the project plan to accomplish the project's requirements. Execution process involves coordinating people and resources, as well as integrating and performing the activities of the project in accordance with the project plan.

Command to run the project is:

```
“Python manage.py runserver”
```

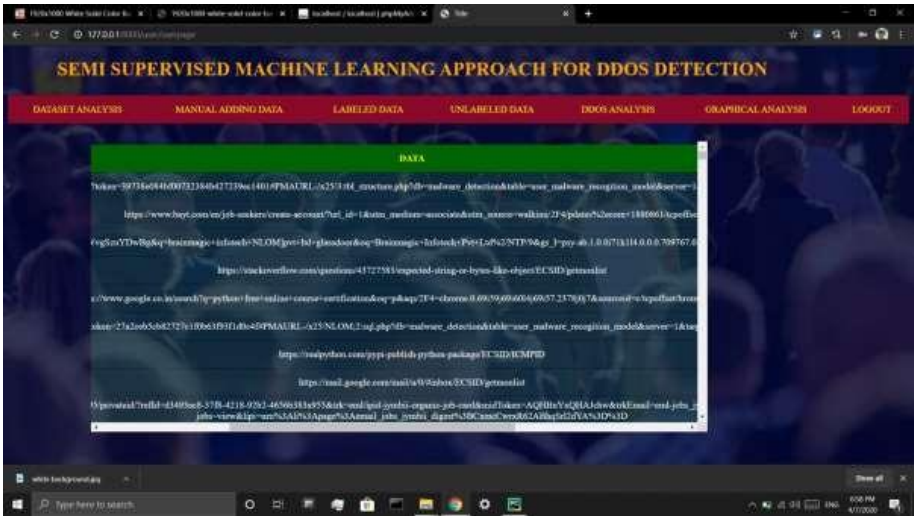
output screens

Following are the screenshots that will display after execution of our project Source Code:



SEMI SUPERVISED MACHINE LEARNING APPROACH FOR DDOS DETECTION
admin
••••
LOGIN HERE

Fig. 2: Login page.



?token=39738e0848d00732384b427239ec1401#PMAURL-/x25!3 tbl_structure.php?db=malware_detection&table=user_malware_recognition_model&server=1

https://www.bayt.com/en/job-seekers/create-account/?url_id=1&utm_medium=associate&utm_source=walkins/2F4/pdater%2ecent=1880861/tcpeffie

/vgSzuYDwBg&q=brainmagic+infotech+NLOM/gvt=bd=glassdoor&oq=Brainmagic+Infotech+Pvt+Ltd%2/NTP/9&gs_l=psy-ab.1.0.0i71k1l4.0.0.0.709767.0

https://stackoverflow.com/questions/45727585/expected-string-or-bytes-like-objectECSID/getmonlist

://www.google.co.in/search?q=python+free+online+course+certification&oq=pkaqs/2F4=chrome.0.69i59j69i60l4j69i57.2378j0j7&sourceid=chrome

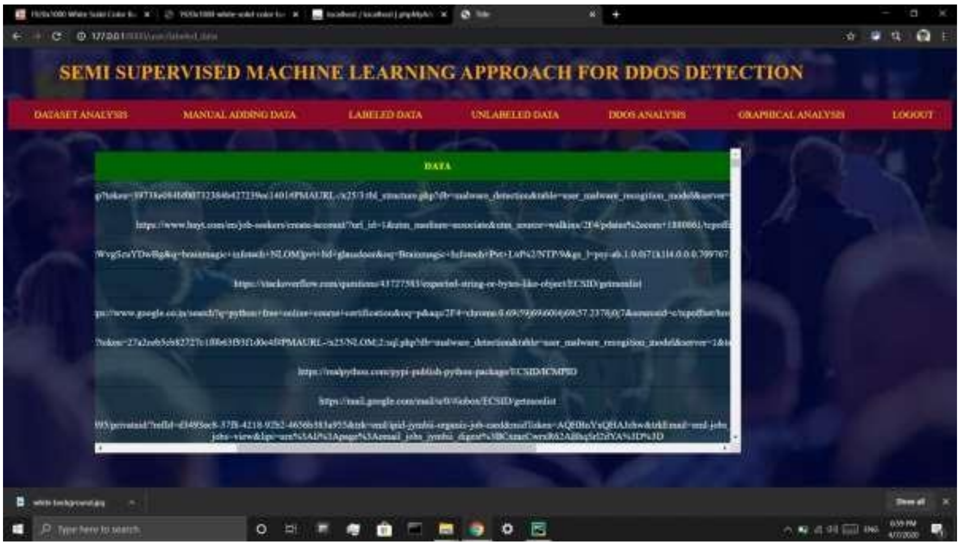
oken=27a2eeb5eb82727e1f0b63f93f1d0e45#PMAURL-/x25/NLOM,2:sql.php?db=malware_detection&table=user_malware_recognition_model&server=1&target

https://realpython.com/pypi-publish-python-package/ECSIDMCMPID

https://mail.google.com/mail/u/0/#inbox/ECSID/getmonlist

5/privateid/?refld=d3493ae8-37f8-4218-92b2-4656b383a955&trk=eml/ipid-jymbii-organic-job-card&midToken=AQHBbYaQHAJdw&trkEmail=eml-jobs_jobs-view&lipi=urn%3Ali%3Apage%3Aemail_jobs_jymbii_digest%3BCamdCwrdR62ABbq5d2dYA%3D%3D

Fig. 2: DDoS dataset analysis.



p?token=39738e0848d00732384b427239ec1401#PMAURL-/x25!3 tbl_structure.php?db=malware_detection&table=user_malware_recognition_model&server=

https://www.bayt.com/en/job-seekers/create-account/?url_id=1&utm_medium=associate&utm_source=walkins/2F4/pdater%2ecent=1880861/tcpeffi

WvgSzuYDwBg&q=brainmagic+infotech+NLOM/gvt=bd=glassdoor&oq=Brainmagic+Infotech+Pvt+Ltd%2/NTP/9&gs_l=psy-ab.1.0.0i71k1l4.0.0.0.709767

https://stackoverflow.com/questions/45727585/expected-string-or-bytes-like-objectECSID/getmonlist

ps://www.google.co.in/search?q=python+free+online+course+certification&oq=pkaqs/2F4=chrome.0.69i59j69i60l4j69i57.2378j0j7&sourceid=chrome

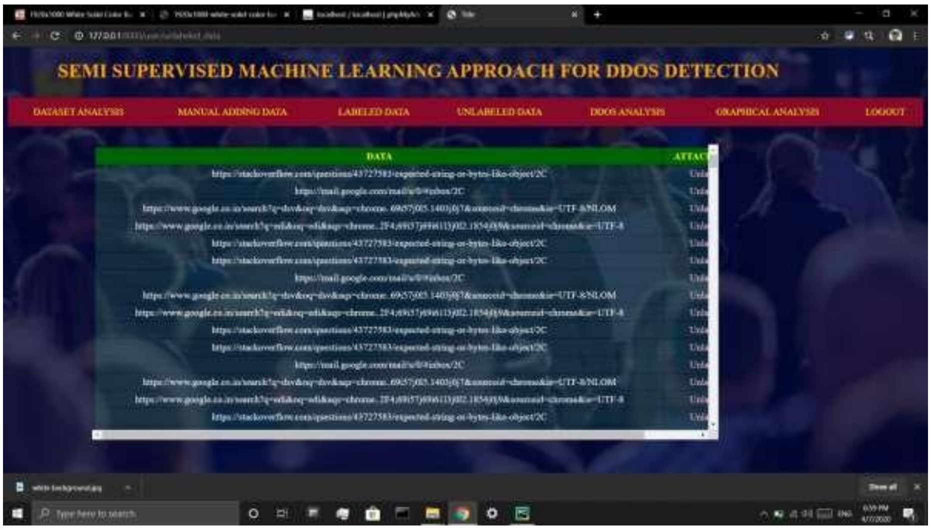
?token=27a2eeb5eb82727e1f0b63f93f1d0e45#PMAURL-/x25/NLOM,2:sql.php?db=malware_detection&table=user_malware_recognition_model&server=1&t

https://realpython.com/pypi-publish-python-package/ECSIDMCMPID

https://mail.google.com/mail/u/0/#inbox/ECSID/getmonlist

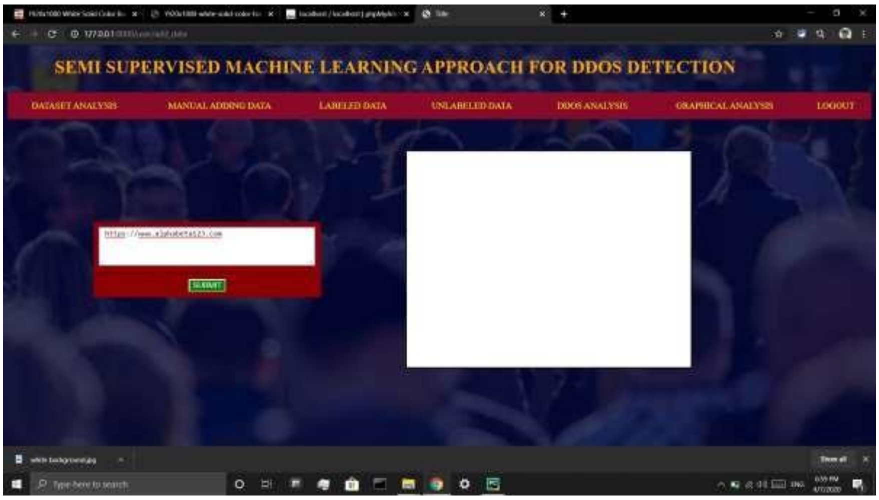
95/privateid/?refld=d3493ae8-37f8-4218-92b2-4656b383a955&trk=eml/ipid-jymbii-organic-job-card&midToken=AQHBbYaQHAJdw&trkEmail=eml-jobs_jobs-view&lipi=urn%3Ali%3Apage%3Aemail_jobs_jymbii_digest%3BCamdCwrdR62ABbq5d2dYA%3D%3D

Fig. 3: Labeled data.



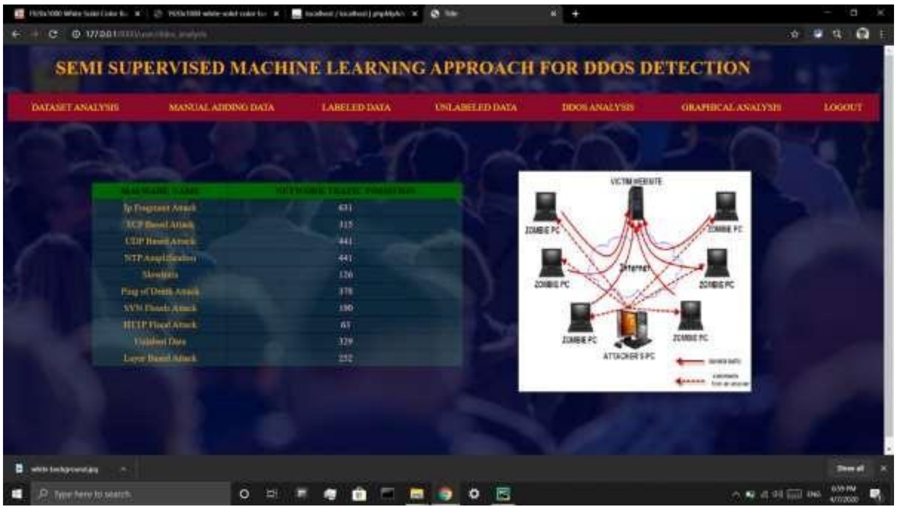
DATA	ATTAC
https://stackoverflow.com/questions/43727583/expected-string-or-bytes-like-object/2C	Unla
https://mail.google.com/mail/u/0/#inbox/2C	Unla
https://www.google.co.in/search?q=dsv&oq=dsv&aqs=chrome..69i57j0l5.1403j0j7&sourceid=chrome&ie=UTF-8/NLOM	Unla
https://www.google.co.in/search?q=sdi&oq=sdi&aqs=chrome..2F4,69i57j69i61l3j0l2.1854j0j9&sourceid=chrome&ie=UTF-8	Unla
https://stackoverflow.com/questions/43727583/expected-string-or-bytes-like-object/2C	Unla
https://stackoverflow.com/questions/43727583/expected-string-or-bytes-like-object/2C	Unla
https://mail.google.com/mail/u/0/#inbox/2C	Unla
https://www.google.co.in/search?q=dsv&oq=dsv&aqs=chrome..69i57j0l5.1403j0j7&sourceid=chrome&ie=UTF-8/NLOM	Unla
https://www.google.co.in/search?q=sdi&oq=sdi&aqs=chrome..2F4,69i57j69i61l3j0l2.1854j0j9&sourceid=chrome&ie=UTF-8	Unla
https://stackoverflow.com/questions/43727583/expected-string-or-bytes-like-object/2C	Unla
https://stackoverflow.com/questions/43727583/expected-string-or-bytes-like-object/2C	Unla
https://mail.google.com/mail/u/0/#inbox/2C	Unla
https://www.google.co.in/search?q=dsv&oq=dsv&aqs=chrome..69i57j0l5.1403j0j7&sourceid=chrome&ie=UTF-8/NLOM	Unla
https://www.google.co.in/search?q=sdi&oq=sdi&aqs=chrome..2F4,69i57j69i61l3j0l2.1854j0j9&sourceid=chrome&ie=UTF-8	Unla
https://stackoverflow.com/questions/43727583/expected-string-or-bytes-like-object/2C	Unla

Fig. 4: Unlabeled data.



https://www.alphabeta123.com

Fig. 5: Manual adding data.



MALWARE NAME	NETWORK TRAFIC POSITION
Ip Fragment Attack	631
TCP Based Attack	315
UDP Based Attack	441
NTP Amplification	441
Slowloris	126
Ping of Death Attack	378
SYN Floods Attack	190
HTTP Flood Attack	63
Unlabel Data	329
Layer Based Attack	252

Fig. 6: DDoS analysis.

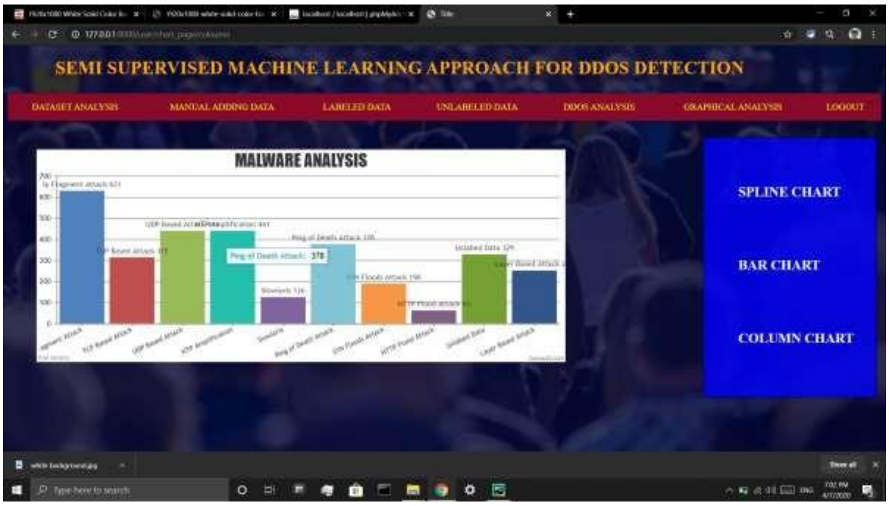
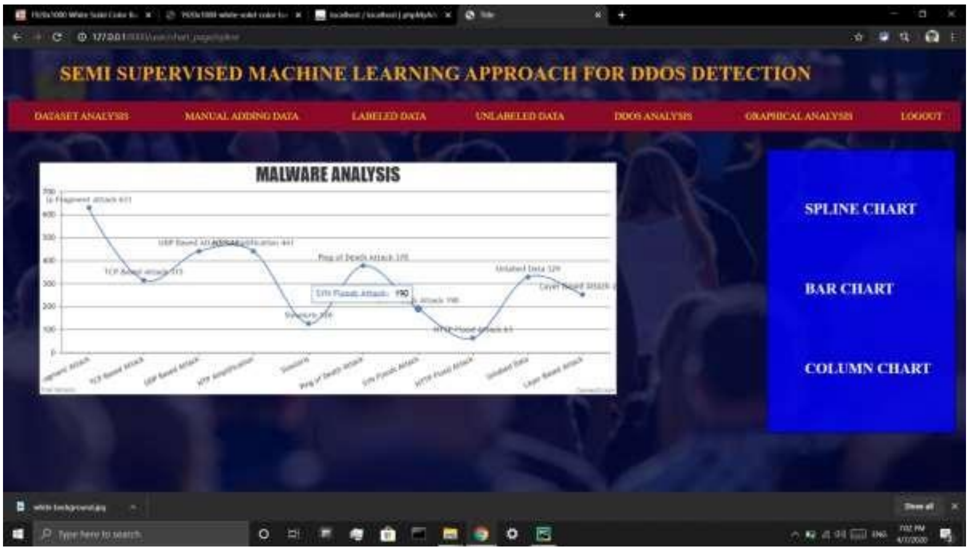
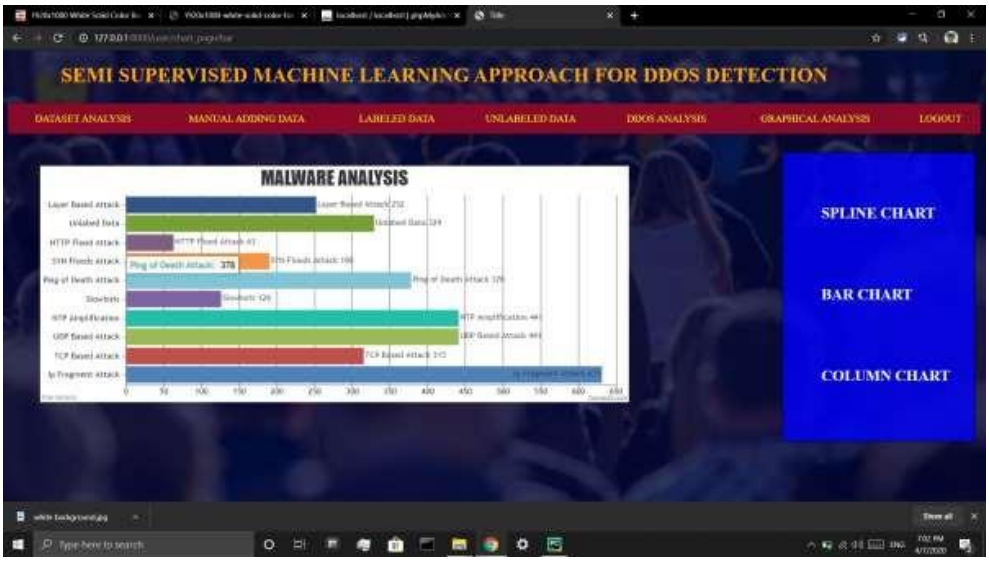


Fig. 7: Graphical analysis.





5. CONCLUSION

However, today Internet security is a quickly growing concern. The prevalence of Internet attacks has increased significantly, but still the challenge of detecting such attacks generally falls on the end hosts and service providers, requiring system administrators to detect and block attacks on their own.

Distributed Denial of Service (DDoS) attack is a very serious problem of web applications. Finding the efficient solution of this problem is essential. Researchers have developed many techniques to detect and prevent this vulnerability.

REFERENCE

[1] Bhuyan MH, Bhattacharyya DK, Kalita JK (2015) An empirical evaluation of information metrics for low-rate and high- rate ddos attack detection. Pattern Recogn Lett 51:1–7

[2] Lin S-C, Tseng S-S (2004) Constructing detection knowledge for ddos intrusion tolerance. Exp Syst Appl 27(3):379–390

[3] Chang RKC (2002) Defending against flooding-based distributed denial-of-service attacks: a tutorial. IEEE Commun Mag 40(10):42–51

[4] Yu S (2014) Distributed denial of service attack and defence. Springer, Berlin

[5] Wikipedia (2016) 2016 dyn cyberattack. https://en.wikipedia.org/ wiki/2016 Dyn cyberattack. (Online; accessed 10 Apr 2017)

[6] theguardian (2016) Ddos attack that disrupted internet was largest of its kind in history, experts say. https://www.theguardian.com/ technology/2016/oct/26/ddos-attack-dyn- mirai-botnet. (Online; accessed 10 Apr 2017)

[7] Kalegele K, Sasai K, Takahashi H, Kitagata G, Kinoshita T (2015) Four decades of data mining in network and systems management. IEEE Trans Knowl Data Eng 27(10):2700–2716

[8] Han J, Pei J, Kamber M (2006) What is data mining. Data mining: concepts and techniques. Morgan Kaufinann

[9] Berkhin P (2006) A survey of clustering data mining techniques. In: Grouping multidimensional data. Springer, pp 25–71

[10] Mori T (2002) Information gain ratio as term weight: the case of summarization of ir results. In: Proceedings of the 19th international conference on computational linguistics, vol 1. Association for Computational Linguistics, pp 1–7

[11] Geurts P, Ernst D, Wehenkel L (2006) Extremely randomized trees. Mach Learn 63(1):3–42

[12] Tavallaee M, Bagheri E, Lu W, Ghorbani A-A (2009) A detailed analysis of the kdd cup 99 data set. In: Proceedings of the second IEEE symposium on computational intelligence for security and defence applications 2009

[13] Shiravi A, Shiravi H, Tavallaee M, Ghorbani AA (2012) Toward developing a systematic approach to generate benchmark datasets for intrusion detection. Comput Secur 31:357–374

[14] Moustafa N, Slay J (2015) Unsw-nb15: a comprehensive data set for network intrusion detection systems (unsw-nb15 network data set). In: Military communications and information systems conference (MilCIS), 2015. IEEE, pp 1–6

[15] Moustafa N, Slay J (2016) The evaluation of network anomaly detection systems: statistical analysis of the unsw- nb15 data set and the comparison with the kdd99 data set. Inf Secur J: Glob Perspect 25:18– 31s