SEMI-SUPERVISEDMACHINELEARNINGAPPROACHFORDDOS DETECTION

Dr. A. Swetha, Assistant Professor CSE(AI&ML), Vaagdevi College of Engineering (Autonomous), Bollikunta, Khila Warangal (Mandal), Warangal Urban-506005(T.S)
Mrs. A. Nagajyothi, Assistant Professor CSE(AI&ML), Vaagdevi College of Engineering (Autonomous), Bollikunta, Khila Warangal (Mandal), Warangal Urban-506005(T.S)
D. Shiva Sai Kumar (21645A6608), UG Student CSE(AI&ML), Vaagdevi College of Engineering (Autonomous), Bollikunta, Khila Warangal (Mandal), Warangal Urban-506005(T.S)
S. Manoj Kumar (20641A66C3), UG Student CSE(AI&ML), Vaagdevi College of Engineering (Autonomous), Bollikunta, Khila Warangal (Mandal), Warangal Urban-506005(T.S)
N. Saketh Reddy (20641A66E2), UG Student CSE(AI&ML), Vaagdevi College of Engineering (Autonomous), Bollikunta, Khila Warangal (Mandal), Warangal Urban-506005(T.S)
B. Sai Tharun (20641A66E6), UG Student CSE(AI&ML), Vaagdevi College of Engineering (Autonomous), Bollikunta, Khila Warangal (Mandal), Warangal Urban-506005(T.S)

ABSTRACT

The appearance of malicious apps is a serious threat to the Android platform. Most types of networkinterfacesbasedontheintegratedfunctions, stealuserspersonalinformationandstart the attackoperations. Inthis paper, we propose an effective and automatic malware detection methodusingthetextsemanticsofnetworktraffic. Inparticular, weconsidereachHTTPflow 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 fromthenaturallanguageprocessing(NLP).Inaddition,wedesignadetectionsystemtodrive 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.

1.INTRODUCTION

Despite the important evolution of the information security technologies in recent years, the DDoS attack remains a major threat of Internet. The attack aims mainly to deprive legitimate users from Internet resources [1]. The impact of the attack relies on the speed and the amount of the network traffic sent to the victim. Generally, there exist 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 [2]. Attackers have taken advantages of Cloud Computing and Internet of Things technologies to generate a huge amount of attack traffic; more than 665 Gb/s .Analyzing this amount of network traffic at once is inefficient, computationally costly and often leads the intrusion detection systems to fall.

Data mining techniques have been used to develop sophisticated intrusion detection systems for the last two decades. 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 Despite that preprocessingand 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. These three crucial steps are the most challenging of the existing data mining based intrusion detection approaches [3].

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. 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 [4].

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 [5]. 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.

In this paper we present an online sequential semi-supervised ML approach for DDoS detection. A time based sliding window algorithm is used to estimate the entropy of the network header features of the incoming network traffic. When the entropy exceeds its normal range, the unsupervised co-clustering algorithm splits the incoming network traffic into three clusters. Then, an information gain ratio is computed based on the average entropy of the network header features between the network traffic subset of the current time window and each one of the obtained clusters. The network traffic data clusters that produce high information gain ratio are considered as anomalous and they are selected for preprocessing and classification using an ensemble classifiers based on the Extra-Trees algorithm

Our approach constitutes of two main parts unsupervised and supervised. The unsupervised part includes entropy estimation, co-clustering and information gain ratio. The supervised part is the Extra-Trees ensemble classifiers. 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 is used to reduce the false positive rates of the unsupervised part and to accurately classify the DDoS traffic. To better evaluate the performance of the proposed approach three public network traffic datasets are used in the experiment, namely the NSL-KDD the UNB ISCX IDS 2012 dataset and the UNSW-NB15 The experimental results are satisfactory when compared with the state-of-the-art DDoS detection methods.

The main contributions of this paper can be summarized as follows:
Presenting an unsupervised and time sliding window algorithm for detecting anomalous traffic data based on co-clustering, entropy estimation and information gain ratio. This algorithm allows to reduce drastically the amount of network traffic to preprocess and to classify, resulting in a significant improvement of the performance of the proposed approach.
Adopting a supervised ensemble ML classifiers based on the Extra-Trees 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-artDDoS 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.

2.LITERATURE SURVEY

Distributed Denial of Service (DDoS) attacks represent a major threat to uninterrupted and efficient Internet service. In this paper, we empirically evaluate several major information metrics, namely, Hartley entropy, Shannon entropy, Renyi’s entropy, generalized entropy, Kullback–Leibler divergence and generalized information distance measure in their ability to detect both low-rate and high-rate DDoS attacks [6]. These metrics can be used to describe characteristics of network traffic data and an appropriate metric facilitates building an effective model to detect both low-rate and high-rate DDoS attacks. We use MIT Lincoln Laboratory, CAIDA and TUIDS DDoS datasets to illustrate the efficiency and effectiveness of each metric for DDoS detection.

Intrusion tolerance is the ability of a system to continue providing (possibly degraded but) adequate services after a penetration. With the rapid development of network technology, distributed denial of service (DDoS) attacks become one of the most important issues today. In this paper, we propose a DDoS ontology to provide a common terminology for describing the DDoS models consisting of the Profile model(the representation of the behaviors of system and users) and the Defense model (the descriptions of Detection and Filter methodologies). Also, the Evaluation strategy based upon current statuses of users' behaviors is used to evaluate the degree of the intrusion tolerance of the proposed models during DDoS attacks. Based upon the ontology, four KCs (Profile model, Evaluation strategy, Detection methodology, and Filter methodology Knowledge Classes) and their relationships are then proposed, where each KC may contain a set of sub-KCs or knowledge represented as a natural rule format. For an arbitrarily given network environment, the default knowledge in the Profile KC and the Evaluation KC, the appropriate detection features in the Detection KC, and the suitable access control list policies in the Filter KC can be easily extracted and adopted by our proposed integrated knowledge acquisition framework. We are now implementing a NORM-based DDoS intrusion tolerance system for DDoS attacks to evaluate the proposed models [7].

Flooding-based distributed denial-of- service (DDoS) attack presents a very serious threat to the stability of the Internet. In a typical DDoS attack, a large number of compromised hosts are amassed to send useless packets to jam a victim or its Internet connection, or both. In the last two years, it is discovered that DDoS attack methods and tools are becoming more sophisticated, effective, and also more difficult to trace to the real attackers [8]. On the defense side, cur-rent technologies are still unable to with stand large-scale attacks. The main purpose of this article is therefore twofold. The first one is to describe various DDoS attack methods, and to present a systematic review and evaluation of the existing defense mechanisms. The second is to discuss a longer-term solution, dubbed the Inter-net- firewall approach, that attempts to intercept attack packets in the Internet core, well before reaching the victim.

This brief provides readers a complete and self-contained resource for information about DDoS attacks and how to defend against them. It presents the latest developments in this increasingly crucial field along with background context and survey material. The book also supplies an overview of DDoS attack issues, DDoS attack detection methods, DDoS attack source trace back, and details on how hackers organize DDoS attacks. The author concludes with future directions of the field, including the impact of DDoS attacks on cloud computing and cloud technology. The concise yet comprehensive nature of this brief makes it an ideal reference for researchers and professionals studying DDoS attacks. It is also a useful resource for graduate students interested in cyberterrorism and networking.

How has the interdisciplinary data mining field been practiced in Network and Systems Management (NSM)? In Science and Technology, there is a wide use of data mining in areas like bioinformatics, genetics, Web, and, more recently, astro informatics. However, the application in NSM has been limited and inconsiderable. In this article, we provide an account of how data mining has been applied in managing networks and systems for the past four decades, presumably since its birth. We look into the field's applications in the key NSM activities—discovery, monitoring, analysis, reporting, and domain knowledge acquisition [9].

3.PROBLEM STATEMENT

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 classifythe anomaly detected in the first phase to identify the attacks class in order to choose the appropriate intervention. Two public datasets are used for experiments in this paper namely the UNSW-NB15 and the NSL-KDDSeveral approaches have been proposed for detecting DDoS attack. Information theory and machine learning are the. The performances of network intrusion detectionapproaches, in general, rely on the distribution characteristics of the underlaying network traffic data used for assessment. 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. Therefore, the need of combining both, supervised and unsupervised approaches arises to overcome DDoS detection issues.

3.1 DISADVANTAGES:

The datasets above are split into train subsets and test subsets using a configuration of 60% and 40% respectively. The train subsets are used to fit the Extra-Trees ensemble classifiers and the test subsets are used to test the entire proposed approach. Before fitting the classifiers the train subsets are normalized using the MinMaxmethod. This section presents the details of the proposed approachand the methodology followed for detecting the DDoSattack. The proposed approach consists of five majorsteps: Datasets preprocessing, estimation of network trafficEntropy, online co-clustering, information gain ratio. The aim of splitting the anomalous network traffic is to reduce the amount of data to be classified by excluding the normal cluster for the classification. For DDoS detection normal traffic records are irrelevant and noisy as the normal behaviors continue to evolve. Most of the time the new unseen normal traffic instances cause the increase of the false positive rate and the decrease of the classification accuracy. Hence, excluding some noisy normal instances of the network traffic data for classification is beneficial in terms of low false positive rates and classification accuracy. Assuming that after the network traffic clustering one cluster contains only normal traffic, a second one contains only DDoS traffic and a third one contains both DDoS and normal traffic.

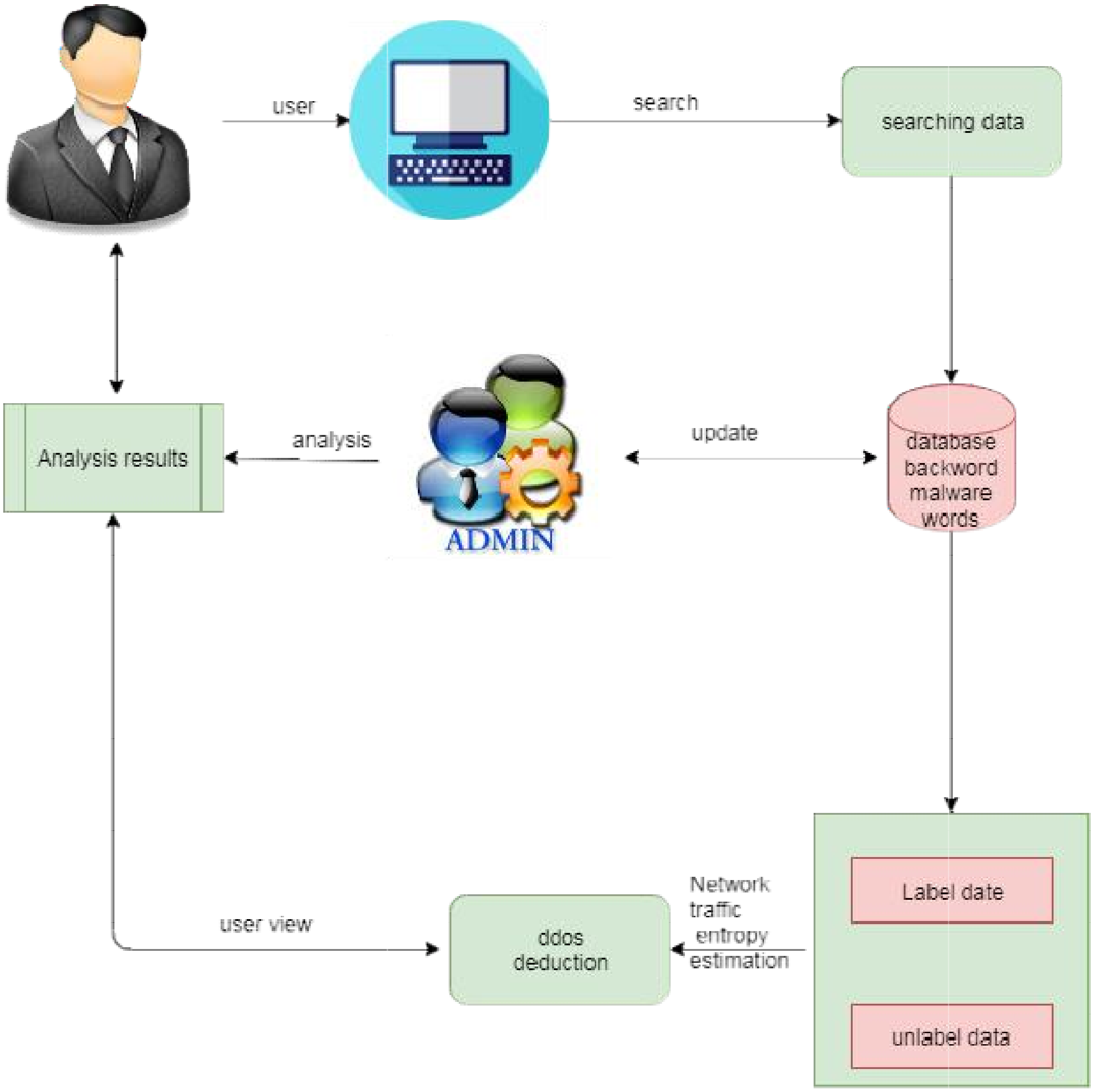
4. PROPOSED SYSTEM

This sections introduces our methodology to detect the DDoS attack. The five-fold steps application process of data mining techniques in network systems discussed in characterizes the followed methodology.The main aim of combining algorithms used in the proposedapproach 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 [10] 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. The focus only on the suspected time windows allows to filter important amount of network traffic data, therefore only relevant data is selected for the remaining steps of the proposed approach. Also, important resources are saved when no abnormal entropy occurs. In order to determine the normal cluster, we estimate the information gain ratio based on the average entropy of the FSD features between the received network traffic data during the current time window and each one of the obtained clusters. As discussed in the previous section during a DDoS period the generated amount of attack traffic is largely bigger than the normal traffic. Hence, estimating the information gain ratio based on the FSD features allows to identify the two cluster that preserve more information about the DDoS attack and the cluster that contains only normal traffic. Therefore, the cluster that produce lower information gain ratio is considered as normal and the remaining clusters are considered as anomalous. The information gain ratio is computed for each cluster as follows:

4.1ADVANTAGE:

Where subset represents the received subset of networkdata during the time window w, Ci (i= 1, 2, 3) are the obtained clusters from subset and |Ci | is the size of theithcluster. avgH(subset) is the average entropy of the FSDfeatures of the input subset and |subset | represents the size The clustering of the incoming network traffic dataallows to reduce important amount of normal and noisy databefore the preprocessing and classification steps. More than6% of a whole traffic dataset can be filtered .

5. SYSTEM ARCHITECTURE



6. IMPLEMENTATION

6.1.User Apps:

User handling for some various times of smart phones, desktops, laptops and tablets. If any kind of devices attacks for some unauthorized Malware softwares.In this Malware on threats for user personal dates includes for personal contact, bank account numbers and any kind of personal documents are hacking in possible.

6.2. DDOS Attack Deduction:

User search the any link Notably, not all network traffic data generated by malicious apps correspond to malicious traffic. Many malware take the form of repackaged benign apps; thus, Malware can also contain the basic functions of a benign app. Subsequently, the network traffic they generate can be characterized by mixed benign and malicious network traffic. We examine the traffic flow header using Co-clustering algorithm from the natural language processing (NLP).

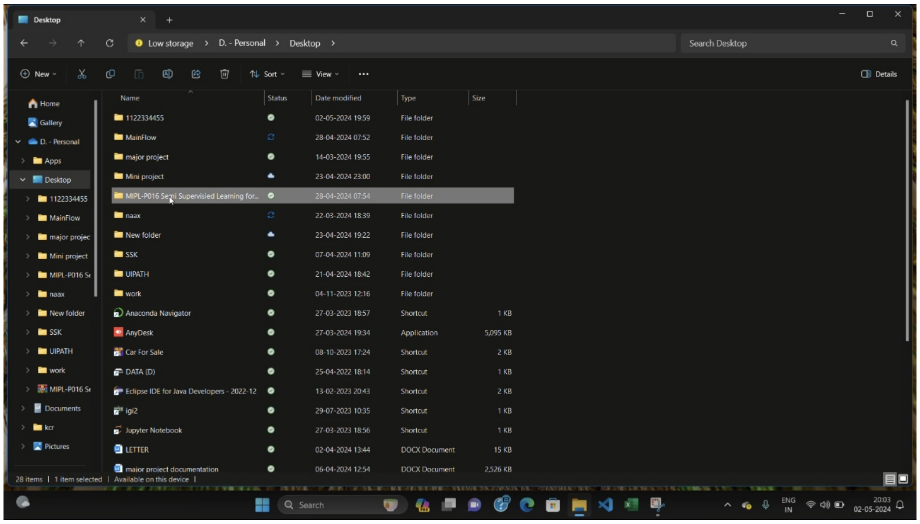
6.3.Classifications of DDOS Attack:

Here, we compare the classification performance of Co-clustering algorithm with other popular machine learning algorithms. We have selected several popular classification algorithms. For all algorithms, we attempt to use multiple sets of parameters to maximize the performance of each algorithm. Using Co-clustering algorithm algorithms classification for malware bag-of- words weightage.

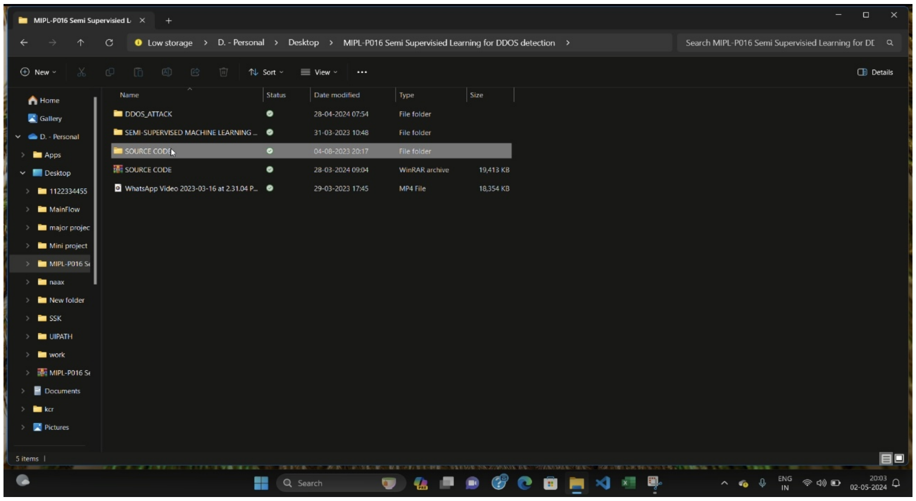
6.4.Graphical analysis:

The graph analysis is done by the values taken from the result analysis part and it can be analyzed by the graphical representations. Such as pie chart, pyramid chart and funnel chart here in this project.

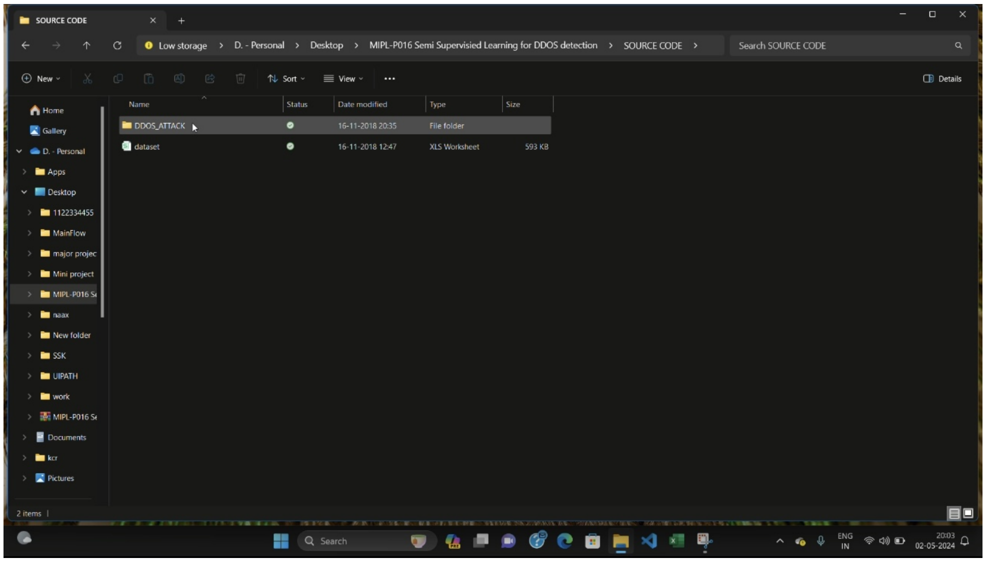
7. OUTPUT RESULTS



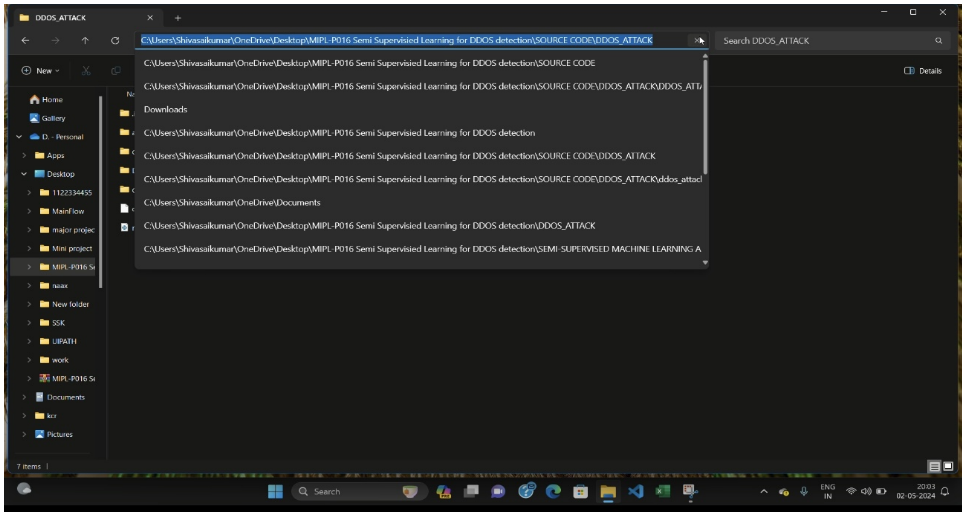
Name	Status	Date modified	Type	Size
1122334455	✓	02-05-2024 19:59	File folder	
MainFlow	↻	28-04-2024 07:52	File folder	
major project	✓	14-03-2024 19:55	File folder	
Mini project	☁	23-04-2024 23:00	File folder	
MIPL-P016 Semi Supervisied Learning for...	✓	28-04-2024 07:54	File folder	
naax	↻	22-03-2024 18:39	File folder	
New folder	☁	23-04-2024 19:22	File folder	
SSK	✓	07-04-2024 11:09	File folder	
UIPATH	✓	21-04-2024 18:42	File folder	
work	✓	04-11-2023 12:16	File folder	
Anaconda Navigator	✓	27-03-2023 18:57	Shortcut	1 KB
AnyDesk	✓	27-03-2024 19:34	Application	5,095 KB
Car For Sale	✓	08-10-2023 17:24	Shortcut	2 KB
DATA (D)	✓	25-04-2022 18:14	Shortcut	1 KB
Eclipse IDE for Java Developers - 2022-12	✓	13-02-2023 20:43	Shortcut	2 KB
igi2	✓	29-07-2023 10:35	Shortcut	1 KB
Jupyter Notebook	✓	27-03-2023 18:56	Shortcut	1 KB
LETTER	✓	02-04-2024 13:44	DOCX Document	15 KB
major project documentation	✓	06-04-2024 12:54	DOCX Document	2,526 KB



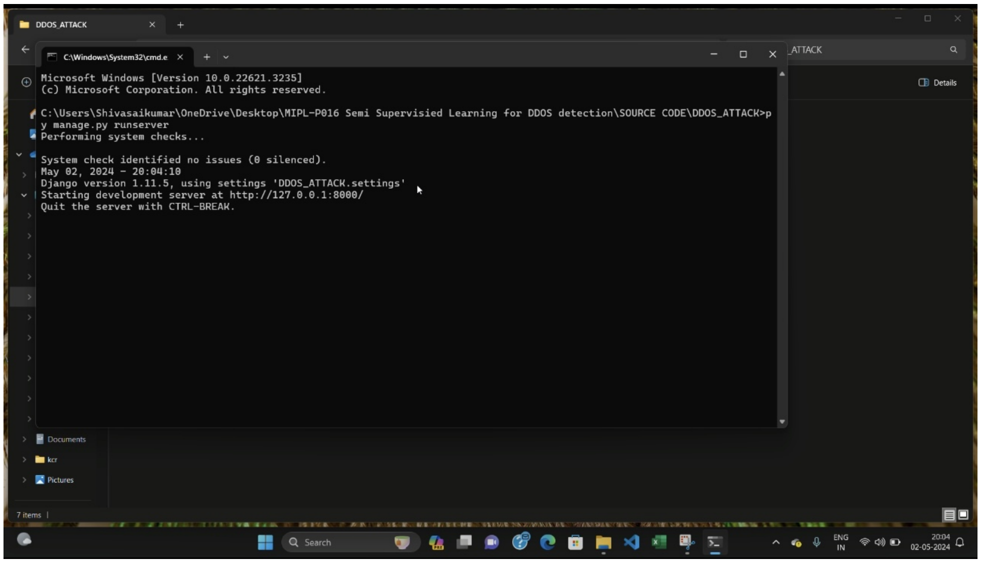
Name	Status	Date modified	Type	Size
DDOS_ATTACK	✓	28-04-2024 07:54	File folder	
SEMI-SUPERVISED MACHINE LEARNING ...	✓	31-03-2023 10:48	File folder	
SOURCE CODE	✓	04-08-2023 20:17	File folder	
SOURCE CODE	✓	28-03-2024 09:04	WinRAR archive	19,413 KB
WhatsApp Video 2023-03-16 at 2.31.04 P...	✓	29-03-2023 17:45	MP4 File	18,354 KB



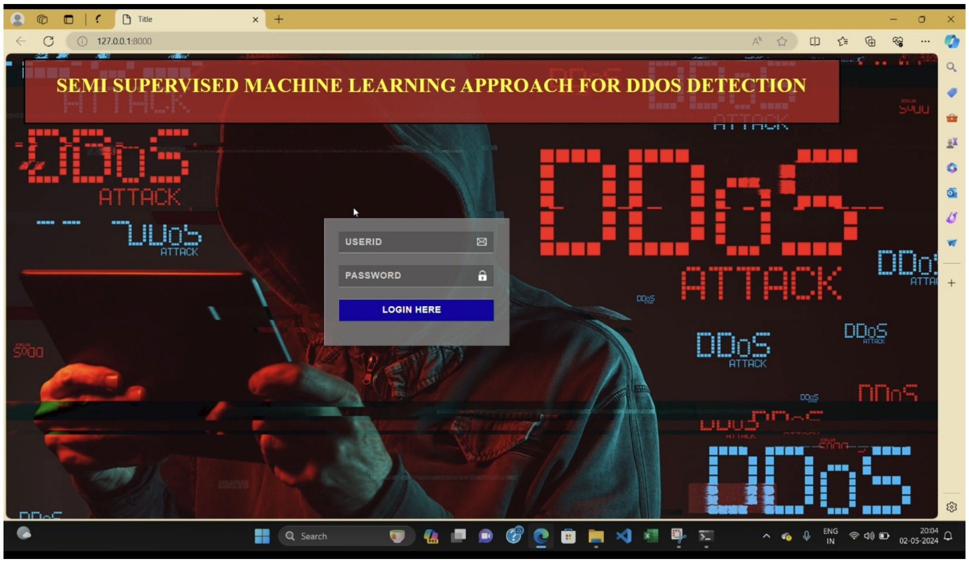
Name	Status	Date modified	Type	Size
DDOS_ATTACK	✓	16-11-2018 20:35	File folder	
dataset	✓	16-11-2018 12:47	XLS Worksheet	593 KB

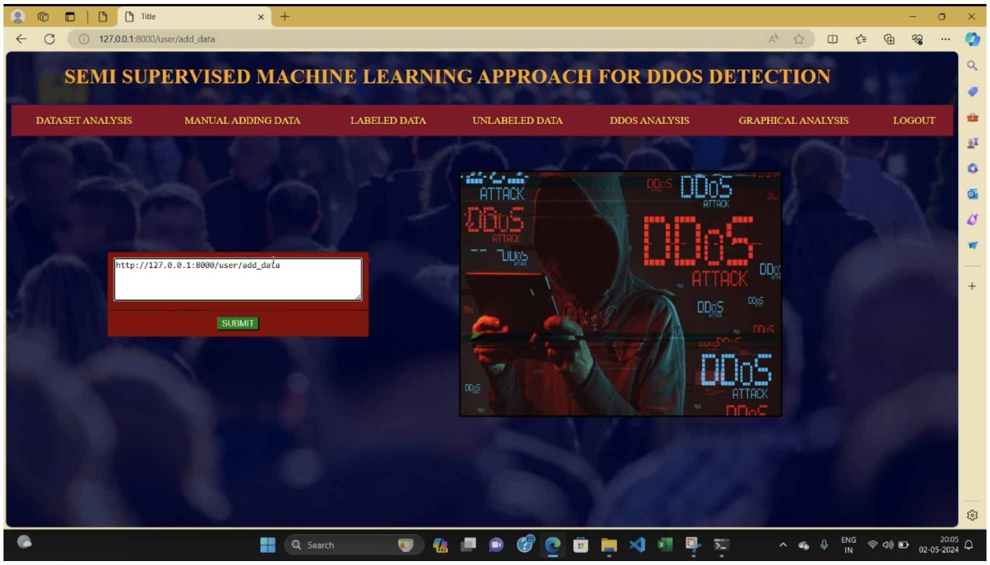


```
C:\Users\Shivasaikumar\OneDrive\Desktop\MIPL-P016 Semi Supervisied Learning for DDOS detection\SOURCE CODE\DDOS_ATTACK
C:\Users\Shivasaikumar\OneDrive\Desktop\MIPL-P016 Semi Supervisied Learning for DDOS detection\SOURCE CODE
C:\Users\Shivasaikumar\OneDrive\Desktop\MIPL-P016 Semi Supervisied Learning for DDOS detection\SOURCE CODE\DDOS_ATTACK\DDOS_ATT/
Downloads
C:\Users\Shivasaikumar\OneDrive\Desktop\MIPL-P016 Semi Supervisied Learning for DDOS detection
C:\Users\Shivasaikumar\OneDrive\Desktop\MIPL-P016 Semi Supervisied Learning for DDOS detection\SOURCE CODE\DDOS_ATTACK
C:\Users\Shivasaikumar\OneDrive\Desktop\MIPL-P016 Semi Supervisied Learning for DDOS detection\SOURCE CODE\DDOS_ATTACK\ddos_attacl
C:\Users\Shivasaikumar\OneDrive\Documents
C:\Users\Shivasaikumar\OneDrive\Desktop\MIPL-P016 Semi Supervisied Learning for DDOS detection\DDOS_ATTACK
C:\Users\Shivasaikumar\OneDrive\Desktop\MIPL-P016 Semi Supervisied Learning for DDOS detection\SEMI-SUPERVISED MACHINE LEARNING A
```

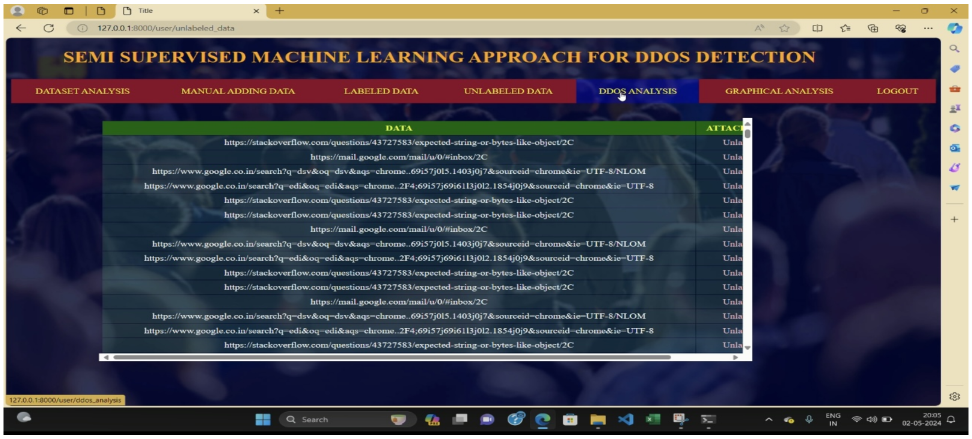


```
Microsoft Windows [Version 10.0.22621.3235]
(c) Microsoft Corporation. All rights reserved.

C:\Users\Shivasaikumar\OneDrive\Desktop\MIPL-P016 Semi Supervisied Learning for DDOS detection\SOURCE CODE\DDOS_ATTACK>p
y manage.py runserver
Performing system checks...

System check identified no issues (0 silenced).
May 02, 2024 - 20:04:10
Django version 1.11.5, using settings 'DDOS_ATTACK.settings'
Starting development server at http://127.0.0.1:8000/
Quit the server with CTRL-BREAK.
```

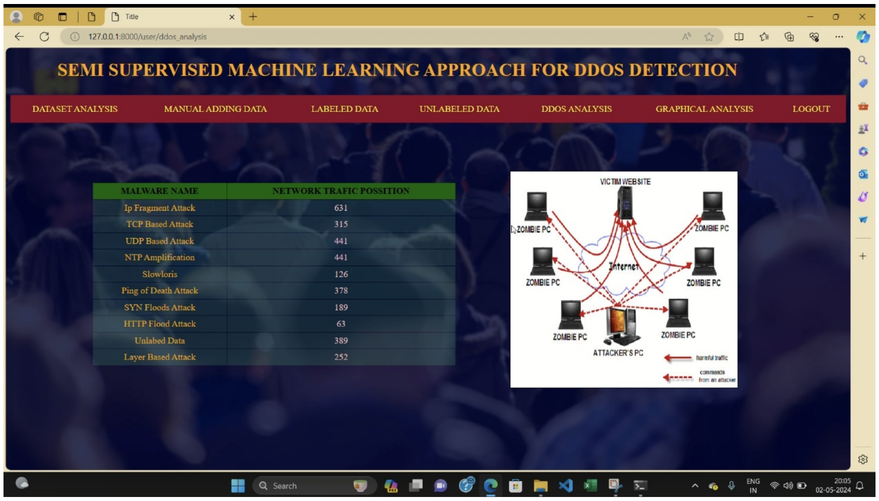

127.0.0.1:8000
SEMI SUPERVISED MACHINE LEARNING APPROACH FOR DDOS DETECTION
USERID
PASSWORD
LOGIN HERE



http://127.0.0.1:8000/user/add_data

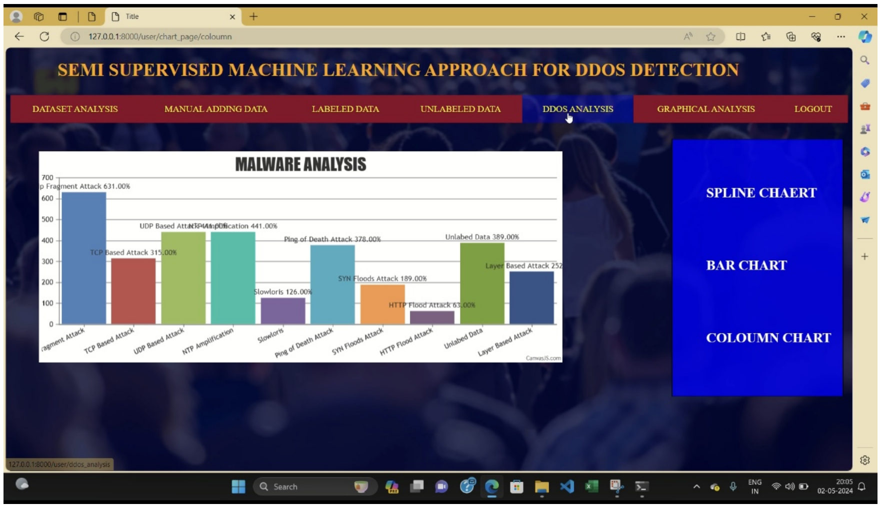

127.0.0.1:8000/user/unlabeled_data
SEMI SUPERVISED MACHINE LEARNING APPROACH FOR DDOS DETECTION
DATASET ANALYSIS
MANUAL ADDING DATA
LABELED DATA
UNLABELED DATA
DDOS ANALYSIS
GRAPHICAL ANALYSIS
LOGOUT
127.0.0.1:8000/user/ddos_analysis

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=edi&oq=edi&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=edi&oq=edi&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=edi&oq=edi&aqs=chrome..2F4;69i57j69i61l3j0l2.1854j0j9&sourceid=chrome&ie=UTF-8	Unla
https://stackoverflow.com/questions/43727583/expected-string-or-bytes-like-object/2C	Unla



SEMI SUPERVISED MACHINE LEARNING APPROACH FOR DDOS DETECTION

MALWARE NAME	NETWORK TRAFIC POSSITION
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	189
HTTP Flood Attack	63
Unlabed Data	389
Layer Based Attack	252



SEMI SUPERVISED MACHINE LEARNING APPROACH FOR DDOS DETECTION

8. CONCLUSION

Android is a new and fastest growing threat to malware. Currently, many research methods and antivirus scanners are not hazardous to the growing size and diversity of mobile malware. As a solution, we introduce a solution for mobile malware detection using network traffic flows, which assumes that each HTTP flow is a document and analyzes HTTP flow requests using NLP string analysis. The N-Gram line generation, feature selection algorithm, and SVM algorithm are used to create a useful malware detection model. Our evaluation demonstrates the efficiency of this solution, and our trained model greatly improves existing approaches and identifies malicious leaks with some false warnings. The harmful detection rate is 99.15%, but the wrong rate for harmful traffic is 0.45%. Using the newly discovered malware further verifies the performance of the proposed system. When used in real environments, the sample can detect 54.81% of harmful applications, which is better than other popular anti-virus scanners. As a result of the test, we show that malware models can detect our model, which does not prevent detecting other virus scanners. Obtaining basically new malicious models Virus total detection reports are also possible. Added, Once, new tablets are added to training samples, we will Please re-train and refresh and update the new malware.

9. FUTURE SCOPE

The future scope for semi-supervised machine learning approaches in DDoS (Distributed Denial of Service) detection is quite promising.With the proliferation of IoT devices and the expansion of network infrastructure, there is an exponential increase in data generated. Much of this data can be utilized for DDoS detection, making semi-supervised learning approaches more relevant.DDoS attacks are becoming increasingly sophisticated, making traditional rule-based detection methods less effective. Semi-supervised learning can adapt to the evolving nature of attacks by leveraging both labeled and unlabeled data to detect anomalies.Semi-supervised learning techniques can scale better compared to fully supervised methods since they can make use of large amounts of unlabeled data, which is often readily available in network traffic monitoring.By combining the strengths of both supervised and unsupervised learning, semi-supervised approaches can potentially achieve higher accuracy in DDoS detection. They can learn from labeled data while also discovering patterns and anomalies in unlabeled data.Labeling network traffic data for DDoS detection can be time-consuming and expensive. Semi-supervised learning can reduce the need for manual annotation by leveraging unlabeled data, thus lowering the overall cost of implementing DDoS detection systems.Semi-supervised learning models can adapt to changing network environments and attack strategies more effectively compared to traditional methods. They can continuously learn from incoming data without the need for frequent retraining.

10. REFERENCES

1. Bhuyan MH, Bhattacharyya DK, Kalita JK (2015) An empiricalevaluation 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 forddos intrusion tolerance. Exp Syst Appl 27(3):379–390
3. ChangRKC(2002)Defendingagainstflooding-baseddistributeddenial-of-serviceattacks: a tutorial. IEEE Commun Mag 40(10):42–51
4. YuS (2014) Distributed denialofserviceattackand defense.Springer,Berlin
5. Wikipedia(2016)2016dyncyberattack.https://en.wikipedia.org/wiki/2016Dyn cyberattack.(Online;accessed10Apr 2017)
6. theguardian (2016) Ddos attack that disrupted internet was largestof its kind in history, expertssay.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. HanJ,PeiJ,KamberM(2006)Whatisdatamining.Datamining:conceptsandtechniques. Morgan Kaufinann
9. Berkhin P (2006) A survey of clustering data mining techniques.In: Grouping multidimensional data. Springer, pp 25–71
10. MoriT(2002)Informationgainratioastermweight:thecaseofsummarizationofirresults. In: Proceedings of the 19th international conference on computational linguistics, vol 1. Association for Computational Linguistics, pp 1–7
11. GeurtsP,ErnstD,WehenkelL(2006)Extremelyrandomizedtrees.MachLearn63(1):3–42
TavallaeeM,BagheriE,LuW,GhorbaniA-A(2009)Adetailedanalysisofthekddcup99data set.In:ProceedingsofthesecondIEEEsymposiumoncomputationalintelligence for security and defence applications 2009.