Detection of Anomalies in Communication Ne

Project Phase I (ECD401)
Mid Semester Examination -1 (September 12, 2022 , Monday)

PROJECT TEAM MEMBERS

1. Gowri V S (BT19ECE033)
2. Sumedha Janani Siriyapuraju (BT19ECE107)
3. Srilikhita Balla (BT19ECE014)
4. Mukesh Kumar Vanika (BT19ECE074)

NAME OF SUPERVISOR

Dr. Abhay Gandhi

Contents

- Introduction
- Problem Statement
- Problem Formulation
- Literature Survey
- Timeline
- Progress
- Results
- Short Term Goals
- Conclusion
- References

Introduction

- Detecting anomalies is a major issue that has been studied for centuries. Nu distinct methods have been developed and used to detect anomalies for diff applications. Anomaly detection refers to “the problem of finding patterns in d do not conform to expected behavior”.
- Normal statistical methods fail in the times when the anomalies/intrusions a following no previous pattern and are not constrained to a particular method
- In our college network too, intrusions into the network could occur both from the college or outside. Although there have been firewalls installed and mea have been adapted for network security, there is still a risk involved, the firew checking anomalies based on fixed protocols and patterns may become outd soon and there is a need for ML based firewalls and IDSs.

Problem Statement

Detection of Anomalies in Communication Networks using Machine Learning.

Problem Formulation

- The rapid growth in the use of computer networks results in the issues of ma
the network availability, integrity, and confidentiality.
- Many techniques have been used to detect anomalies. One of the increasin
significant techniques is Machine Learning (ML), which plays an important r
area.
- Traditional rule based IDSs are not useful for more advanced times where a
and intrusion pattern change. Thus the need for machine learning and deep
based systems arises.
- Hence, we are using Machine Learning Techniques to detect and classify th
anomalies in Communication Networks in our problem statement.

Literature Survey

- Intrusions and the mode of intrusions could be any of the following :
  - Host based or network based intrusions
  - Signature based mechanism or behavioural based mechanism

- Anomaly based intrusion detection
  - Protocol anomaly detection : Refers to exceptions based on protocol format and behavi respect to common practice. Anomalies like - illegitimate command usage, field values a combinations, unusual occurrences of particular commands etc
  - Application Payload anomaly : Eg. presence of shell code in unexpected fields
  - Statistical anomaly detection : includes threshold detection and profile based detection

- Machine Learning
  - In contrast to statistical methods stated above, ML techniques are well suited to learnir with no prior knowledge of the patterns.
  - Clustering and classification are two popular machine learning algorithms used in IDS

A Machine Learning-Based Classification
Prediction Technique for DDoS Attacks

- DDoS attacks

    Distributed network attacks are referred to, usually, as Distributed Denial of Service (DDoS)attacks. Th take advantage of specific limitations that apply to any arrangement asset, such as the framework of th organization's site.

    DDoS attacks are web applications and business websites; and the attacker may have different goals. Some common types DDoS attacks are

    SYN flood,is a form of denial of service attack, in which an attacker rapidly initiates a connection to a se finalizing the connection. The server has to spend resources waiting for half-opened connections, whic consume enough resources to make the system unresponsive to legitimate traffic.

    The UDP flood is a kind of denial-of-service attacks in which numerous User Datagram Protocol (UDP) forwarded to a computer server (targeted) in order to exhaust that server's capability to execute and re

    The HTTP flood is an attack type in which the attacker manipulates HTTP and POST unwanted reques application .

The Smurf attack uses a malware program called smurf to abuse the Internet Protocol (IP) and Interne Message Protocol (ICMP). It imitates the IP address and use ICMP to ping the IP address of the specifie organization.

The Fraggle attack uses a large amount of UDP traffic to transmit to the transmission organization of th This is like a Smurf attack using UDP instead of ICMP.

NTP amplification attack, the attacker abuses a functionality of the Network Time Protocol (NTP) serve devastate a targeted server or network with a large quantity of User Datagram Protocol (UDP) traffic; a result this rendering the destination infrastructure unreachable to regular legitimate users traffic.

Among the machine learning techniques, random forest and XGBoost both are powerful supervised le models.Both are applicable and used for classification problems. The random forest algorithm is appro times faster than other algorithms and best working for classification problems. This should be noted t XGBoost is the ideal algorithm of machine learning because it is approximately 100 times faster than t forest and best for forbid data analysis. Both are simple and faster than other algorithm in terms of ex times.

Detecting Denial of Service Attacks Using Machine Learning Algorithms

- DDos is a type of attack in which the power of multiple malware affected systems a disrupt a network connection.
- It creates a massive amount of traffic and network congestion. Due to this, it becor impossible for data packets to reach its destination. This leads to denial of service f users.
- This is the most significant cyber-threat at the moment. This is because, by inhibitir server's ability to provide resources to legitimate clients, the impacted server's par such as bandwidth and buffer size, are slowed down.

- There are two different models used to identify DDOS attacks.
  - Mathematical Model: In mathematical model, we establish a relation between inter-ar of requests and throughput.

    Throughput (S) = Packet Size(L) *Inter-arrival time of requests

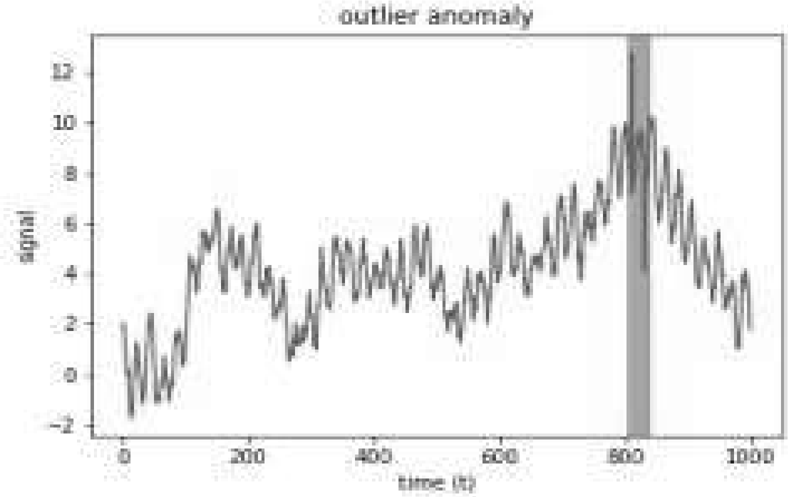
  - We compute the median of thresholds and consider it as the threshold.If throughput is threshold, then it is categorized as an attack or else it is not an attack.
  - Machine Learning Model: In machine learning model, we compare the throughput of th with the threshold. If throughput is above the threshold, then it is categorized as an att it is not an attack.
  - Later, we calculate the Miss Rate of the model.

Unsupervised Anomaly Detection for Communication Networks: an Autoenc

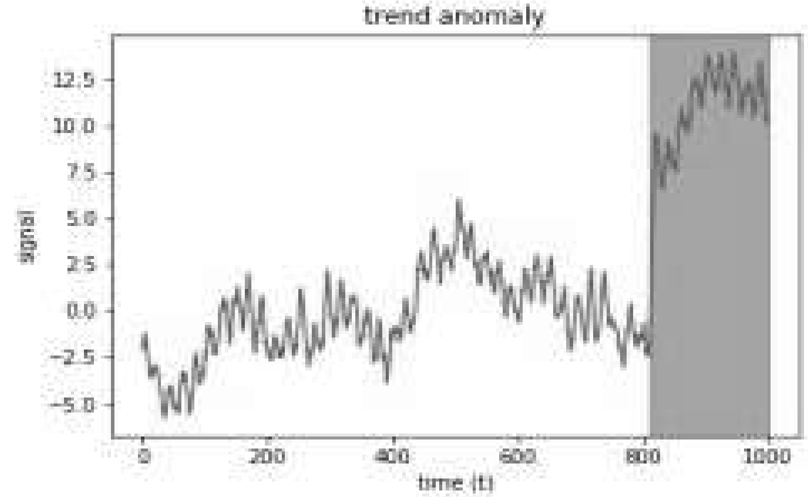
Approach

- Anomaly Detection in Communication Networks is a challenging task as systems ha components and the sheer scale and the dynamic nature of these systems make tra Anomaly Detection Techniques inadequate.
- Many traditional Anomaly Detections are threshold based, it becomes infeasible to the threshold manually, and this leads to a very high rate of false positives.
- Skyline communications has come up with a solution which monitors Communicati Networks and detects upcoming network failures and informs the customers of its high risks.
- But, in real time, we just need to be notified about critical anomalies as the non cri are financially exhausting.
- So, with this Auto Encoder approach, we can predict anomalies when behavior dev normal behavior.

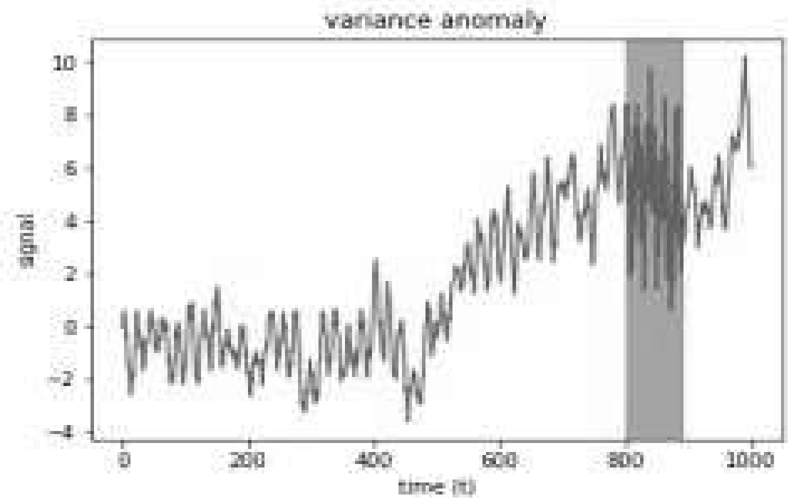
- Anomalies have been classified into 4 types- outliers, trend changes, variance chan level shifts.
- Outliers is when there is a sudden spike in the number of packets of data requestec
- Trend changes is when there is a slow change in behavior over a long time period.
- Variance changes is when the variance of the number of packets requested change
- Level Shifts is when there is a temporary change in the behavior.
- A stream of data with timestamp is created.
- A time-based window function is created and separates the data into different stre
- The reconstruction error is calculated by flagging the thresholds based on the inter representation.
- The time stamp wise division and the RE helps in checking changes that happens in interval and reduces False Positive Rate.
- Various Architectures such as Dense AE, LSTM AE, LSTM seq2seq, TCN AE and TCN s was implemented to classify these anomalies and was observed that TCN seq2seq architectures perform best on the artificial datasets ,while dense AE shows best re the real time use data.



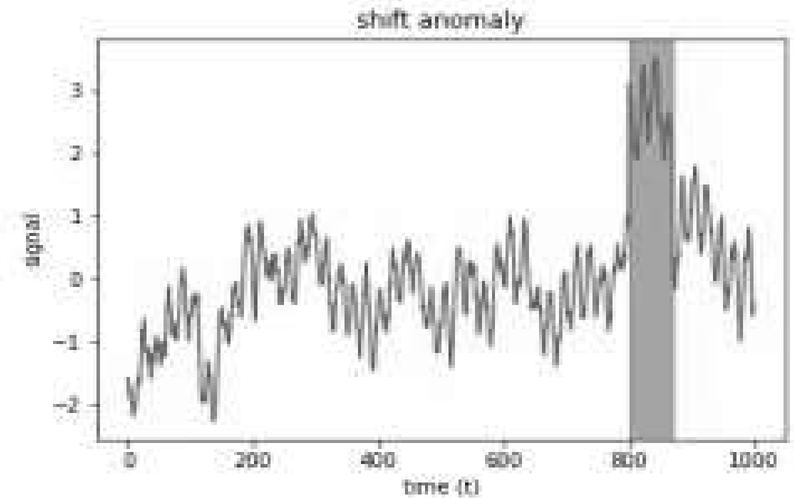
(a) Outlier anomaly



(b) Trend change anomaly



(c) Variance change anomaly



(d) Level shift anomaly

No.	Time	Source	Destination	Protocol	Length	Info
1	0.000000	192.168.137.1	239.255.255.250	SSDP	217	M-SEARCH * HTTP/1.1
2	0.384720	192.168.137.64	142.250.183.14	UDP	1285	55872 → 443 Len=1243
3	0.385031	192.168.137.64	142.250.183.14	UDP	1292	55872 → 443 Len=1250
4	0.385232	192.168.137.64	142.250.183.14	UDP	1292	55872 → 443 Len=1250
5	0.385410	192.168.137.64	142.250.183.14	UDP	1292	55872 → 443 Len=1250
6	0.385586	192.168.137.64	142.250.183.14	UDP	1292	55872 → 443 Len=1250
7	0.385767	192.168.137.64	142.250.183.14	UDP	1292	55872 → 443 Len=1250
8	0.385935	192.168.137.64	142.250.183.14	UDP	1292	55872 → 443 Len=1250
9	0.386113	192.168.137.64	142.250.183.14	UDP	1292	55872 → 443 Len=1250
10	0.386283	192.168.137.64	142.250.183.14	UDP	1292	55872 → 443 Len=1250
11	0.386443	192.168.137.64	142.250.183.14	UDP	267	55872 → 443 Len=225
12	0.389657	192.168.137.64	142.250.183.206	UDP	1285	49422 → 443 Len=1243
13	0.389914	192.168.137.64	142.250.183.206	UDP	1292	49422 → 443 Len=1250
14	0.390098	192.168.137.64	142.250.183.206	UDP	1292	49422 → 443 Len=1250
15	0.390262	192.168.137.64	142.250.183.206	UDP	1292	49422 → 443 Len=1250
16	0.390407	192.168.137.64	142.250.183.206	UDP	1292	49422 → 443 Len=1250
17	0.390556	192.168.137.64	142.250.183.206	UDP	1292	49422 → 443 Len=1250
18	0.390705	192.168.137.64	142.250.183.206	UDP	1292	49422 → 443 Len=1250
19	0.390867	192.168.137.64	142.250.183.206	UDP	1292	49422 → 443 Len=1250
20	0.390993	192.168.137.64	142.250.183.206	UDP	1292	49422 → 443 Len=1250
21	0.391148	192.168.137.64	142.250.183.206	UDP	1292	49422 → 443 Len=1250
22	0.391331	192.168.137.64	142.250.183.206	UDP	1292	49422 → 443 Len=1250
23	0.391506	192.168.137.64	142.250.183.206	UDP	1292	49422 → 443 Len=1250
24	0.391683	192.168.137.64	142.250.183.206	UDP	274	49422 → 443 Len=232
25	0.404643	142.250.183.14	192.168.137.64	UDP	71	443 → 55872 Len=29
26	0.404643	142.250.183.14	192.168.137.64	UDP	72	443 → 55872 Len=30
27	0.404866	142.250.183.206	192.168.137.64	UDP	74	443 → 49422 Len=32
28	0.408002	142.250.183.206	192.168.137.64	UDP	68	443 → 49422 Len=26
29	0.408700	142.250.183.14	192.168.137.64	UDP	68	443 → 55872 Len=26
30	0.409243	192.168.137.64	142.250.183.206	UDP	76	49422 → 443 Len=34
31	0.409679	192.168.137.64	142.250.183.14	UDP	75	55872 → 443 Len=33

Wireshark Dataset

	0	udp	private	SF	105	146	0.1	0.2	0.3	0.4	...	254	1.00.1	0.01	0.00.6	0.00.7	0.00.8	0.00.9	0.00.10	0.00.1
0	0	udp	private	SF	105	146	0	0	0	0	...	254	1.0	0.01	0.00	0.0	0.0	0.0	0.0	0.
1	0	udp	private	SF	105	146	0	0	0	0	...	254	1.0	0.01	0.00	0.0	0.0	0.0	0.0	0.
2	0	udp	private	SF	105	146	0	0	0	0	...	254	1.0	0.01	0.00	0.0	0.0	0.0	0.0	0.
3	0	udp	private	SF	105	146	0	0	0	0	...	254	1.0	0.01	0.01	0.0	0.0	0.0	0.0	0.
4	0	udp	private	SF	105	146	0	0	0	0	...	255	1.0	0.00	0.01	0.0	0.0	0.0	0.0	0.
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	.
311023	0	udp	private	SF	105	147	0	0	0	0	...	255	1.0	0.00	0.01	0.0	0.0	0.0	0.0	0.
311024	0	udp	private	SF	105	147	0	0	0	0	...	255	1.0	0.00	0.01	0.0	0.0	0.0	0.0	0.
311025	0	udp	private	SF	105	147	0	0	0	0	...	255	1.0	0.00	0.01	0.0	0.0	0.0	0.0	0.
311026	0	udp	private	SF	105	147	0	0	0	0	...	255	1.0	0.00	0.01	0.0	0.0	0.0	0.0	0.
311027	0	udp	private	SF	105	147	0	0	0	0	...	255	1.0	0.00	0.01	0.0	0.0	0.0	0.0	0.

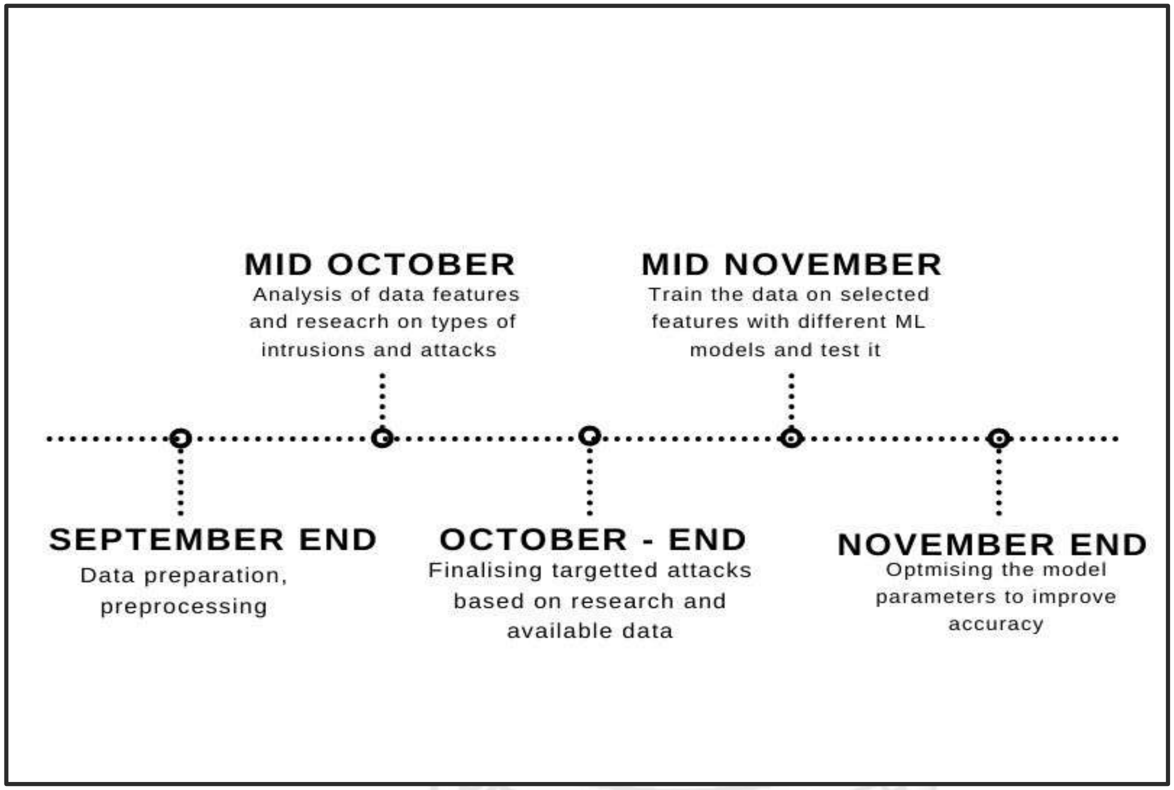
KDDCup99 Dataset

	Time	Blade	Action	Type	Interface	Origin	Source	Source User Name	Destination	Service	Application Risk	Application Name
0	Sep 1, 2022 3:32:46 PM	Application Control	Accept	Session	NaN	gateway.vnit.ac.in	172.25.192.101	NaN	104.199.240.237	TCP_1-1024	Low	Spotify
1	Sep 1, 2022 3:32:46 PM	Application Control,URL Filtering	Accept	Session	NaN	gateway.vnit.ac.in	10.19.0.25	NaN	129.226.27.10	http	Unknown	in.teddymobile.cn
2	Sep 1, 2022 3:32:46 PM	Application Control	Accept	Session	NaN	gateway.vnit.ac.in	192.168.17.37	NaN	52.98.57.162	https	Very Low	Office365-Outlook-web
3	Sep 1, 2022 3:32:46 PM	Application Control	Accept	Session	NaN	gateway.vnit.ac.in	10.15.7.244	NaN	172.217.160.208	https	Low	Google Cloud Platform
4	Sep 1, 2022 3:32:46 PM	Content Awareness,Application Control	Accept	Session	NaN	gateway.vnit.ac.in	10.10.1.5	NaN	91.108.56.177	TCP_1-1024	Medium	Telegram

	Time	Creation Time	Interface Direction	Policy Rule UID	Type	Index Time	App Protocol	Policy Date
0	"Sep 1	2022 3:31:56 PM"	2022-09-01T10:01:56Z	inbound	08739ec8-c102-4f82-b6ce-6f38c4ef635f	Session	2022-09-01T08:48:48Z	HTTPS
1	Sep 1, 2022 3:31:56 PM	2022-09-01T10:01:56Z	inbound	1000869c-1ede-4ea8-8953-f806b202e6d7	Session	2022-09-01T08:48:48Z	HTTPS	2022-08-30T04:11:05Z
2	Sep 1, 2022 3:31:56 PM	2022-09-01T10:01:56Z	inbound	1000869c-1ede-4ea8-8953-f806b202e6d7	Session	2022-09-01T08:48:48Z	HTTPS	2022-08-30T04:11:05Z
3	Sep 1, 2022 3:31:56 PM	2022-09-01T10:01:56Z	inbound	1000869c-1ede-4ea8-8953-f806b202e6d7	Session	2022-09-01T08:48:48Z	HTTPS	2022-08-30T04:11:05Z
4	Sep 1, 2022 3:31:56 PM	2022-09-01T10:01:56Z	inbound	4a6d8a91-6b95-4070-be89-fc5e432f6fcf	Session	2022-09-01T08:48:48Z	HTTPS	2022-08-30T04:11:05Z
...	...	...	...	...	...	...	...	...
995	Sep 1, 2022 3:31:20 PM	2022-09-01T10:01:20Z	outbound	08739ec8-c102-4f82-b6ce-6f38c4ef635f	Session	2022-09-01T08:48:48Z	NaN	2022-08-30T04:11:05Z

Firewall dataset obtained from network centre

Timeline for the Project Completio



Work Done

- Researched various real time problem statements which can be of use in ou day lives

- Some such topics were
  - Calculation of path loss in Wireless Communication Networks
  - Prediction of Video quality using ML
  - Detection of anomalies in Communication Networks

- Explored various papers and understood the concept
- Finalised our problem statement based on the importance and relevance of problems
- Acquired sample Firewall data from Network Centre

Results

- Formulated the problem statement after researching on the need to have an efficie
learning based system to detect anomalies
- Researched on the existing methods for anomaly detection and need for machine l
based algorithms
- Researched the different types of attacks possible in the communication network a
the importance of this project specifically to the college network

Work to be Done by the Next Evaluation

- Prepare our dataset using packet sniffer
- To simulate an attack using softwares - like Slowloris
- Analyse our own dataset and analyse the features needed to process the da
  comparing it with already publicly available datasets like KDCup1999

Conclusion

- The numerous variety of attacks make it a necessity to have an intelligent learning
algorithm to detect the attacks.
- Firewalls and other current IDS use statistical methods to detect anomalies and intr
- Need to find the best parameters/features to train the model on to get best accura

References

- A. L. G. Rios, Z. Li, K. Bekshentayeva and L. Trajković, "Detection of Denial of Servic
in Communication Networks," 2020 IEEE International Symposium on Circuits and S
(ISCAS), 2020, pp. 1-5, doi: 10.1109/ISCAS45731.2020.9180445.
- Ismail et al., "A Machine Learning-Based Classification and Prediction Technique fo
Attacks," in IEEE Access, vol. 10, pp. 21443-21454, 2022, doi: 10.1109/ACCESS.202
- Bonte, Pieter & Hautte, Sander & Lejon, Annelies & Ledoux, Veerle & De Turck, Fili
Hoecke, Sofie & Ongenae, Femke. (2020). Unsupervised Anomaly Detection for
Communication Networks: An Autoencoder Approach. 10.1007/978-3-030-66770-
- Kumari, Kimmi, and M. Mrunalini. "Detecting Denial of Service Attacks Using Mach
Learning Algorithms - Journal of Big Data." SpringerOpen, 28 Apr. 2022,