

Analysis of Botnet Attack using Machine Learning

Enamala Yashith Reddy, Grandhe Pranav, Challa Sai Hari Uma Sahith, Datla Yagnith Varma, Ms.R.Deepa
Students, B.Tech, Department of CSE, R.M.K Engineering College, Kavaraipettai, Tamil Nadu, India
Assistant Professor, Department of CSE, R.M.K Engineering College, Kavaraipettai, Tamil Nadu, India

Abstract:

Investing in cutting-edge botnet detection software like DataDome, which can carry out real-time botnet detection and make use of top-notch bot mitigation techniques, is the best way to safeguard websites and web servers from botnet assaults. DataDome's AI-powered solution can conduct real-time behavioral analysis to detect botnet traffic and block all botnet activities before they even reach the web server, even though botnet operators are now very skilled at disguising the identity of the botnet. Even the original server response time can be enhanced by putting bot management and protection into place. DataDome compiles information from thousands of websites, examines billions of requests each day, and uses cutting-edge machine learning to update the algorithm constantly. This enables the botnet prevention system to quickly identify both known botnets and emerging threats.

In order to obtain the experimental results, Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) algorithms were used. The results

demonstrate that performance measures like accuracy, precision, recall, and

confusion matrix. The best aspect is that DataDome does not need any daily maintenance or active botnet mitigation on your part. Set up a list of trusted partner bots to enable, and DataDome will handle all unwanted traffic while you concentrate on more worthwhile projects. Malware from botnets is infecting IoT devices like security cameras, gaming platforms, and more in addition to desktop and laptop computers.

Index terms: Botnet attack, IoT, DataDome, Machine learning, Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN)

1. Introduction:

These days, the Internet of Things (IoT) technology is expanding rapidly, and every minute, many gadgets are connecting to it. Utilizing this technology makes life more organised and easier [1]. For instance, IoT technology was originally only applicable to small offices and homes, but today it is integrated into various sectors for increased

reliability and time savings. Any assault that uses a botnet—a collection of devices and bots linked together to carry out the same task—for distribution and scaling is referred to as a botnet attack [2,3]. Cybercriminals use botnet assaults to conduct intensive scraping, DDoS, and other significant cybercrimes. It is crucial to safeguard businesses from cybercrimes like those that many people have encountered as a result of a hired network. A command and control (C&C or C2) computer and a collection of zombie devices make up every botnet. Sometimes, the words "bot" and "botnet" are used incorrectly to refer to the same thing [4]. To be clear, despite possible similarities, bots and botnets are not the same entity. A botnet is a collection of devices, whereas a bot (or internet bot) is a software application [5,6].

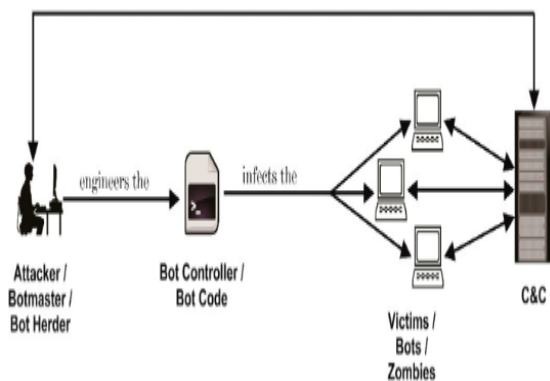


Fig.1. Structure of Botnet attack

A bot is a computer programme created to carry out automated activities online. These jobs are usually straight forward and repeatable, and a bot can complete them much more quickly and precisely than a human user [7]. A Google-owned bot called

Googlebot, for instance, is used to crawl and index webpages, including this one. It is essentially copying and pasting data, a task that can be completed by any human, but it can do so much more quickly and accurately.

2. Related work:

The devices that a bot herder commands are known as "zombie" devices, though they are also sometimes referred to as "bot devices" or simply "bots"—hence the a fore mentioned misunderstanding. A botnet attack is any malicious endeavor made by a hacker or online criminal using the botnet, to state it simply [8]. The DDoS (Distributed Denial of Service) assault is the most prevalent type of botnet attack. In order to prevent a website or web server from properly serving its actual users, the hacker will use the botnet to transmit an incredibly high volume of requests and/or traffic to that location (hence, denial of service).

2.1.Spam attacks:

When a web service that supports SMTP or POP3 becomes integrated into a botnet [9,10]. It can be used to spread adware on the device, send spam and fraudulent emails to the target in order to defraud them, and other things.

2.2. **Cryptocurrency mining:**

In recent years, cryptocurrency mining botnet hijacking has become a prevalent form of cybersecurity threat [10].

2.3. **Fraud traffic:**

Create phony web traffic or fraudulent ad clicks to increase income.

2.4. **Ransom:**

Infect devices with malware and demand money to "unlock" them, or force users to pay to have their device removed from the botnet.

2.5. **Spyware:**

Passwords, credit card numbers, and other private information are monitored by the botnet, which then reports its findings to its owner [11]. The attacker can then advertise this private information on the dark web.

3. Working of Botnet attack:

3.1. Here are some important best solutions for defending against botnet attacks:

3.1.1. **Update Everything ASAP:**

Team should update software solutions on a regular basis, preferably as soon as updates are made available. Most software vendors will publicize known vulnerabilities as soon as they are patched, and security patches are there for a purpose [12]. It may now be a possible entry point for bot herders to infect systems and enlist them into a botnet if the OS is not updated

when vulnerabilities are made known to hackers [13].

3.1.2. **Choose the Right Botnet Detection Solution:**

A professional botnet detection software solution that is made to identify malicious botnet activity swiftly and accurately in real time is required to adequately defend the entire online ecosystem against botnet attacks.



Fig.2. Botnet attack analysis

Traditional (signature-based) fingerprinting and IP reputation detection are no longer adequate defenses against botnets because of their increasingly complex identity-mapping techniques.

3.1.3. **Stop botnet attacks now:**

Powerful machine learning is used by DataDome, a cutting-edge botnet detection and security solution, to safeguard servers and websites from malicious botnet operators [14].

3.2. Educate Users and Employees to Avoid Phishing Attacks:

- Ensure that the organisation, users, and particularly employees who may be more likely to be targeted

due to their position in the organisation have access to sufficient security training and instructional materials about how to prevent phishing attacks [15]. Several fundamental principles include:

- Don't download email attachments unless you are certain of the sender's name. Before clicking on any unexpected file or link, carefully check the sender's email address.
- Purchase reliable antivirus and anti-malware software that can correctly and automatically scan attachments for malware.
- Don't click on any links in any communications (emails, texts, social media direct messages, etc.) unless you are certain that the sender is who they say they are [16]. To prevent DNS cache poisoning when visiting the link, it is preferable to manually type the URL into the address bar of the browser rather than clicking.

- **Use Strong and Unique Passwords:**

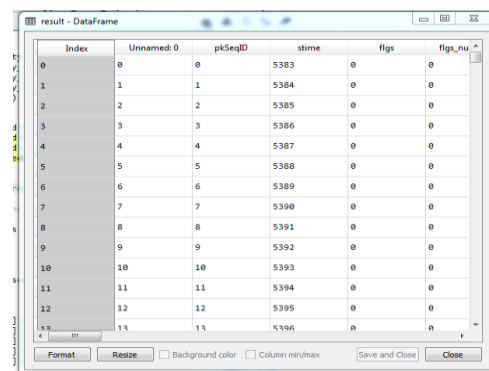
For every account, including admin accounts on all devices that link to other devices or the internet directly, use strong, complicated passwords. Additionally, be positive that each password is specific to a single account.

- **Anti-Malware Software:**

It is critical to engage in a powerful enough antivirus/anti-malware program that can shield your device and system from botnet malware and other online threats because most botnet conversions result from malware infections. Additionally, remember to routinely update your antivirus software.

4. Results:

4.1. Data Selection & Preprocessing:



Index	Unnamed: 0	pkSeqID	stime	flgs	flgs_nu
0	0	0	5383	0	0
1	1	1	5384	0	0
2	2	2	5385	0	0
3	3	3	5386	0	0
4	4	4	5387	0	0
5	5	5	5388	0	0
6	6	6	5389	0	0
7	7	7	5390	0	0
8	8	8	5391	0	0
9	9	9	5392	0	0
10	10	10	5393	0	0
11	11	11	5394	0	0
12	12	12	5395	0	0
13	13	13	5396	0	0

-----Checking Missing Values-----

```

Unnamed: 0      0
pkSeqID         0
stime           0
flgs            0
flgs_number     0
proto           0
proto_number    0
saddr           0
sport           0
daddr           0
dtype: int64

```

Fig.3. Data selection with preprocessing operations on data

The process of choosing the appropriate data source, type, and tools to gather the data is known as data selection. Data selection happens before real data

collection is done. Any type of processing done on raw data to get it ready for another data processing method is referred to as data preprocessing, a part of data preparation. It has historically been a crucial first step in the data mining procedure.

```
dtype: int64
```

-----Before Label Encoding-----						
	Unnamed: 0	pkSeqID	stime	...	attack	category subcategory
0	1650261	1650261	1.528103e+09	...	1	DDoS HTTP
1	1650262	1650262	1.528103e+09	...	1	DDoS HTTP
2	1650263	1650263	1.528103e+09	...	1	DDoS HTTP
3	1650264	1650264	1.528103e+09	...	1	DDoS HTTP
4	1650265	1650265	1.528103e+09	...	1	DDoS HTTP
5	1650266	1650266	1.528103e+09	...	1	DDoS HTTP
6	1650267	1650267	1.528103e+09	...	1	DDoS HTTP
7	1650268	1650268	1.528103e+09	...	1	DDoS HTTP
8	1650269	1650269	1.528103e+09	...	1	DDoS HTTP
9	1650270	1650270	1.528103e+09	...	1	DDoS HTTP

[10 rows x 47 columns]

-----After Label Encoding-----						
	Unnamed: 0	pkSeqID	stime	...	attack	category subcategory
0	0	0	5383	...	1	0 0
1	1	1	5384	...	1	0 0
2	2	2	5385	...	1	0 0
3	3	3	5386	...	1	0 0
4	4	4	5387	...	1	0 0
5	5	5	5388	...	1	0 0
6	6	6	5389	...	1	0 0
7	7	7	5390	...	1	0 0
8	8	8	5391	...	1	0 0
9	9	9	5392	...	1	0 0

Fig.4. Preprocessing operation before and after label encoding the data

4.2. Data Normalization & Data Splitting:

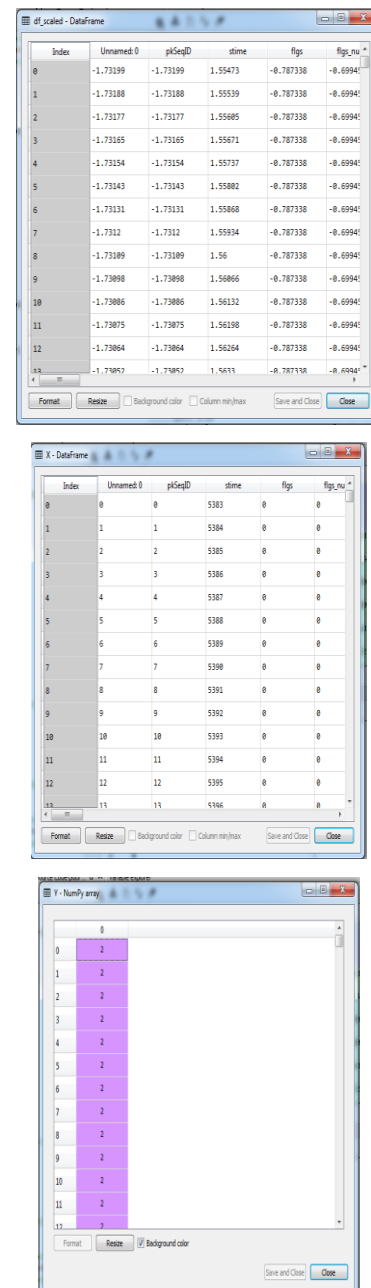


Fig.5. Data Normalization & Data Splitting process

In order to make information simpler to locate, group, and analyze, data entries are organized through the process of normalization, which ensures that they look similar across all fields and records. There are numerous methods and regulations for data standardization. When data is

separated into two or more subsets, this is known as data splitting. Typically, a two-part split is used to train the model while the other portion is used to test or evaluate the data. Data separation is a crucial component of data science, especially when building models from data.

4.3. PCA:

One of the most popular unsupervised machine learning methods, Principal Component Analysis (PCA) is used in a wide range of applications, including exploratory data analysis, dimensionality reduction, information compression, data de-noising, and many others

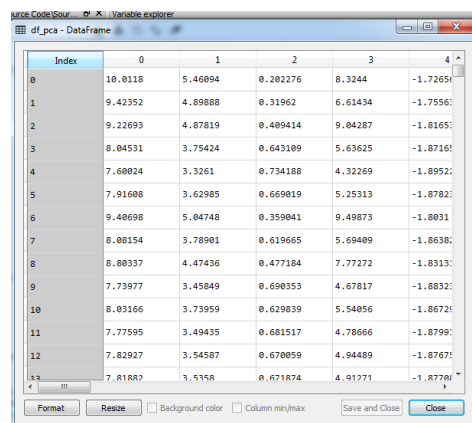


Fig.6. PCA analysis

4.4. Auto Encoder & LSTM:

```

train on 381 samples, validate on 96 samples
epoch 1/10
381/381 [=====] - 3s 7ms/sample - loss: 0.2540 - val_loss: 0.2822
epoch 2/10
381/381 [=====] - 1s 2ms/sample - loss: 0.2534 - val_loss: 0.2815
epoch 3/10
381/381 [=====] - 0s 210us/sample - loss: 0.2527 - val_loss: 0.2808
epoch 4/10
381/381 [=====] - 0s 173us/sample - loss: 0.2521 - val_loss: 0.2800
epoch 5/10
381/381 [=====] - 0s 123us/sample - loss: 0.2514 - val_loss: 0.2792
epoch 6/10
381/381 [=====] - 0s 126us/sample - loss: 0.2507 - val_loss: 0.2785
epoch 7/10
381/381 [=====] - 0s 121us/sample - loss: 0.2500 - val_loss: 0.2777
epoch 8/10
381/381 [=====] - 0s 134us/sample - loss: 0.2493 - val_loss: 0.2769
epoch 9/10
381/381 [=====] - 0s 123us/sample - loss: 0.2485 - val_loss: 0.2761

```

Model: "sequential_1"

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 46, 50)	10480
dropout (Dropout)	(None, 46, 50)	0
lstm_1 (LSTM)	(None, 5)	1120
dropout_1 (Dropout)	(None, 5)	0
dense_10 (Dense)	(None, 3)	18
dense_11 (Dense)	(None, 1)	4
Total params: 11,542		
Trainable params: 11,542		
Non-trainable params: 0		

Long Short term Memory

PERFORMANCE METRICS

```

1.Confusion Matrix [[ 390  0  0  0]
[ 186  0  0  0]
[11607  0  0  0]
[  57  0  0  0]]

```

2.Accuracy 67.70833333333334 %

3.Precision for LSTM 100.0 %

4.Recall for LSTM 67.70833333333334 %

Fig.7. LSTM with Auto encoding process

An autoencoder is a neural network unsupervised learning method that trains the network to ignore signal "noise" in order to learn effective data representations (encoding). Image denoising, image compression, and, in some instances, even the creation of image data can all be done with autoencoders. Long short-term memory networks, or LSTMs, are employed in deep learning. A variety of recurrent neural networks (RNNs) can acquire long-term dependencies,

particularly in problems involving sequence prediction.

4.5.CNN:

Convolutional Neural Networks, also known as CNNs, are a subset of artificial neural networks used in deep learning and are frequently employed for object and image identification and classification. As a result, Deep Learning uses a CNN to identify items in an image.

Model: "model_1"		
Layer (type)	Output Shape	Param #
input_2 (InputLayer)	[(None, 46, 1)]	0
conv1d (Conv1D)	(None, 45, 2)	6
max_pooling1d (MaxPooling1D)	(None, 22, 2)	0
flatten (Flatten)	(None, 44)	0
dense_12 (Dense)	(None, 1)	45
Total params: 51		
Trainable params: 51		
Non-trainable params: 0		
None		

Convolutional Neural Network		
PERFORMANCE METRICS		
1.Confusion Matrix	[[398 148 10662 57]	
[0 38 945 0]	
[0 0 0 0]	
[0 0 0 0]]	
2.Accuracy 100.0 %		
3.Precision 72.4907063197026 %		
4.Recall 100.0 %		

Fig.8. Convolution Neural Network process

5. Conclusion:

The most sophisticated and deadly types of cyber security threats are botnets, which are a major worry for organizations, people, and even governments. Although defending

against botnets and botnet assaults can be challenging, it is not impossible. You can successfully protect equipment from becoming zombie devices and reduce the risk of system/network being impacted by various botnet attacks by using the advice given above and selecting a comprehensive solution that specializes in bot protection.

6. REFERENCES:

1. Li, S.; Xu, L.D.; Zhao, S. The internet of things: A survey. *Inf. Syst. Front.* 2014, 17, 243–259.
2. Al-Rushdan, H.; Shurman, M.; Alnabelsi, S.H.; Althebyan, Q. Zero-Day Attack Detection and Prevention in Software-Defined Networks. In *Proceedings of the International Arab Conference on Information Technology (ACIT)*, Al Ain, United Arab Emirates, 3–5 December 2019.
3. Amin, B.R.; Taghizadeh, S.; Rahman, M.S.; Hossain, M.J.; Varadharajan, V.; Chen, Z. Cyber-attacks in smart grid–dynamic impacts, analyses, and recommendations. *IET Cyber-Phys. Syst. Theory Appl.* 2020, 5, 321–329.
4. Almudaires, F.; Almaiah, M. Data an Overview of Cybersecurity Threats on Credit Card Companies and Credit Card Risk Mitigation. In *Proceedings of the International Conference on Information Technology (ICIT)*, Amman, Jordan, 14–15 July 2021.

5. Yaacoub, J.P.; Noura, H.; Salman, O.; Chehab, A. Security analysis of drones systems: Attacks, limitations, and recommendations. *Internet Things* 2020, 11, 100218.
6. Ettredge, M.; Guo, F.; Li, Y. Trade secrets and cyber security breaches. *J. Account. Public Policy* 2018, 37, 564–585.
7. Outpost24 Product Managers. Cyber Security in 2020 and beyond. Available online: <https://outpost24.com/blog/CyberSecurity-in-2020-and-beyond> (accessed on 1 February 2022).
8. Pandey, A.K.; Tripathi, A.K.; Kapil, G.; Singh, V.; Khan, M.W.; Agrawal, A.; Kumar, R.; Khan, R.A. Trends in Malware Attacks. In *Advances in Digital Crime, Forensics, and Cyber Terrorism*; IGI Global: Hershey, PA, USA, 2020; pp. 47–60.
9. Suresh, P.; Daniel, J.V.; Parthasarathy, V.; Aswathy, R.H. A state-of-the-art review on the Internet of Things (IoT) history, technology, and fields of deployment. In *Proceedings of the International Conference on Science Engineering and Management Research (ICSEMR)*, Chennai, India, 27–29 November 2014.
10. International Telecommunication Union. ITU Internet Report 2005: The Internet of Things; Technical Report; International Telecommunication Union: Geneva, Switzerland, 2005.
11. Alieyan, K.; Almomani, A.; Abdullah, R.; Almutairi, B.; Alauthman, M. Botnet and Internet of Things (IoTs). In *Security, Privacy, and Forensics Issues in Big Data*; IGI Global: Hershey, PA, USA, 2020; pp. 304–316.
12. Blythe, J.M.; Sombatruang, N.; Johnson, S.D. What security features and crime prevention advice are communicated in consumer IoT device manuals and support pages? *J. Cybersecur.* 2019, 5, tyz005.
13. Malhotra, P.; Singh, Y.; Anand, P.; Bangotra, D.K.; Singh, P.K.; Hong, W.C. Internet of Things: Evolution, Concerns and Security Challenges. *Sensors* 2021, 21, 1809.
14. Grizzard, J.B.; Sharma, V.; Nunnery, C.; Kang, B.B.; Dagon, D. Peer-to-Peer Botnets: Overview and Case Study. In *First Workshop on Hot Topics in Understanding Botnets (HotBots 07)*; USENIX Association: Cambridge, MA, USA, 2007.
15. Beltrán-García, P.; Aguirre-Anaya, E.; Escamilla-Ambrosio, P.J.; Acosta-Bermejo, R. IoT Botnets. In *Communications in Computer and Information Science*; Springer International Publishing: Merida, Mexico, 2019; pp. 247–257.
16. Wazzan, M.; Algazzawi, D.; Bamasaq, O.; Albeshri, A.; Cheng, L. Internet of Things Botnet Detection Approaches: Analysis and Recommendations for Future Research. *Appl. Sci.* 2021, 11, 5713.