Science Management Design Journal
Journal Homepage: www.smdjournal.com
ISSN: 2583-925X
Volume: 6
Issue: 2
Pages: 22-28

Dataset for DoS Detection Dataset: A Benchmark for AI-Driven Intrusion Detection and DoS Attack Analysis

Shihab Chilwan[1,*]

[1] Computer Engineering, Vishwakarma University, Pune, 411048, Maharashtra,
*Corresponding Author: Shihab Chilwan: shihab.mujeeb@gmail.com

DoI: https://doi.org/10.70295/SMDJ.2412023

Article history: Received: 12/12/2024, Revised:13/12/2024, Accepted: 14/12/2024, Published Online:15/12/2024

Abstract:

The "DOS ATTACK" dataset, which consists of network traffic data specially prepared to support the research effort into finding the possibility to trace DoS attacks using artificial intelligence and machine learning-based techniques, has more than 820,000 entries. Labeled instances of attack and normal traffic represent this dataset, which are highly useful for training and evaluation of intrusion detection systems. Some of the notable features include source and destination IP addresses, protocol type, packet size, and TCP-specific flags like SYN and ACK, which are actually quite crucial to observe patterns of attacks. The dataset is fairly balanced as about 17 percent of entries fall under the category of attack instances, which gives very realistic settings for evaluating the model. This dataset may also be important for benchmarking AI-based network intrusion detection models, especially for recognition of DoS attacks and general knowledge about network behavior under attack conditions. Salient patterns of distribution in the SYN and ACK flags, packet sizes, and temporal characteristics shown in our analysis could constitute a foundation for more advanced intrusion detection research.

Keywords:

Network intrusion detection, DoS attack, dataset, cybersecurity, machine learning, anomaly detection

1. Introduction:

Cybersecurity is now a huge deal in this modern era of digitalization, where more infrastructure in digital mode is built up and uptrends in sophisticated cyber attacks are at a witness. And, among such attacks, the most widely distributed one has been the DoS network system attack, where some miscreant sends tremendous traffic to any network or service to overwhelm it so that the valid users are desrvedly deprived of using it. The more the technology progresses, the more complex the attacks have become, and traditional methods of protection of the network no longer suffice to prevent such attacks; thus, NIDS seems to find a significant role in the defense against cyber threats in detecting, responding to, and reducing impact in such attacks on the network system. Despite the dramatic role that NIDS plays in the system, it cannot match

Science Management Design Journal
Journal Homepage: www.smdjournal.com
ISSN: 2583-925X
Volume: 6
Issue: 2
Pages: 22-28

with complexity due to changing cyber threats revolutionizing modern cyber threats, so AI-based solutions in cybersecurity are an inevitable step.

1.1 Background and Motivation

DoS attacks and other cyber intrusions threaten organizations' continuity by disrupting service, compromising data integrity, and decreasing the overall level of operational continuity. Given the increasing complexity of networks, such attacks are harder to detect because they are more destructive and require enhanced detection mechanisms, which can quickly and correctly distinguish between bad and good traffic. Traditional NIDSs rely essentially on a signature-based detection technique, where incoming traffic is compared against known attack signatures. Though to some extent effective, it is very inadequate because it depends purely on predefined attack patterns, and completely ineffective in cases of new or modified attacks. The anomaly-based detection approach that tries to identify deviations in behavior for networks appears promising to conquer this limitation but has rather tended to contain high false-positive rates. The bundling of these inadequacies made AI-based solutions an attractive alternative.

1.2 Objectives Of the Study

The new dataset should be presented in this paper, especially designed to develop machine learning-based NIDS, with a strong concentration on DoS attacks, contributing a structured set of network traffic data with labeled instances of attack and nonattack traffic based on attributes such as IP addresses, port numbers, packet size, and protocol. Contributions of this effort include the development of models that can accurately classify network traffic in real time and indicate possible DoS attacks, and thus contribute to much broader fields of automated network security. In addition, the dataset provides a source to compare the differing performance of machine learning models in intrusion detection tasks and guide the identification of optimal algorithms for various network security applications.

1.3 Problem Statement

The challenge of making timely and accurate DoS attack detection and response thus underscores a need for better NIDS solutions. By its definition, a traditional NIDS relies on manual updates and predefined signatures, which is becoming less effective as hackers' tactics evolve. In addition, owing to increased volume and complexity in modern network traffic, anomalies detected result in false positives even today. These weaknesses form the basis for AI-driven methods with high-dimensional data handling capabilities, that are robust against novel patterns of attack, and reliably detect threats. Creating a dataset with the specific purpose of the DoS detection task will allow researchers to train and test various machine learning models, leading eventually to more robust and adaptive NIDS solutions.

Science Management Design Journal
Journal Homepage: www.smdjournal.com
ISSN: 2583-925X
Volume: 6
Issue: 2
Pages: 22-28

2. Material and Methods:

2.1 Dataset Overview

In this experiment, the utilized data set captures network traffic across both normal and several attack scenarios, which include DoS attacks. The experimental design allows us to carry out experiments, train, and validate our machine learning models for monitoring outlying traffic behaviour. The data is from controlled network experiments simulating real-world attack patterns to enhance the confidence and robustness of the outcome.

2.2 Value of Data

- Fully covered: This dataset contains more than 820,000 instances of network traffic, including both attack and normal-labeled entries in support of realistic cybersecurity research.
- In the dataset, attributes used include protocol type, packet length, and TCP flags that are all critical differences between normal traffic and DoS attack traffic.
- Research Utility: This dataset is of high value when used for training, testing, and validating machine learning models toward real-time DoS attack detection.
- Practical Application: By focusing on DoS attack patterns, the data set is helpful to researchers who are developing network intrusion detection systems and may help network administrators and cybersecurity professionals.

2.3 Data Description

The "DOS ATTACK" dataset contains well-structured and labelled network traffic records for cybersecurity research. Features include source IP address, destination IP address, types of protocols, and the length of packets that can help establish a difference between normal and instances of attack cases. Labeling this dataset, distinguishing a different kind of attack traffic from the normal one helps find practical applications for machine learning use cases in DoS attack detection. Another characteristic is the uneven distribution, which reflects real conditions in network traffic, making it more applicable to practical development for intrusion detection systems.

Table 1. Specifications Table

Subject	**Cybersecurity, Machine Learning**
Specific Subject Area	Network Intrusion Detection Systems (NIDS)
Type of Data	Network traffic logs
How Data Were Acquired	Captured from a simulated network environment using monitoring tools with labeled DoS attack patterns.
Data Format	CSV (Comma-Separated Values)
Parameters for Data	Includes features such as source IP, destination IP, protocol types,

Science Management Design Journal
Journal Homepage: www.smdjournal.com
ISSN: 2583-925X
Volume: 6
Issue: 2
Pages: 22-28

Collection	packet sizes, and TCP flags relevant for DoS detection.
Description of Data Collection	Network traffic was recorded in a controlled environment simulating normal and DoS attack scenarios. Essential network metadata was logged to represent attack characteristics.
Data Accessibility	https://drive.google.com/file/d/1kb_7pAqTdrBZN2JV7ZF1weZnEDLXhPdC/view?usp=drive_link

2.4 Experimental Design

Denial-of-Service (DoS) attack traffic was simulated to be captured in the assumed controlled environment. More specifically, the setup consisted of two central systems, the Kali Linux Virtual Machine which will act as the attack source and capturing data packets via wireshark, and the victim system, running Windows operating system.

Here is the step-by-step process:

Attack Simulation: A DoS attack was launched from the Kali Linux Virtual Machine. Creating a flood of packets, the use of the tool hping3 thus simulated the attack situation. In the terminal command was typed hping3 --flood with nearly extremely fast forwarding of packets to the target in a manner that mimics the characteristics of the DoS flood attack. The most common use of hping3 is networking tests, and it has majorly been used to test the configurations of attributes for packets. Therefore, it can be utilized to mirror a vast array of network attacks, including those that characterize high-volume packet floods commonly appearing in typical DoS situations.

Traffic Capture: All the incoming traffic was logged and captured on the victim system using Wireshark. As Wireshark is quite an in vogue network protocol analyzer, all information about packets like source and destination IP addresses, protocol type, and size of packets along with TCP flags like SYN and ACK, were recorded. These would prove to be overall logs of attack traffic and normal traffic which was later differentiated upon and analyzed.

Exporting the data and labeling: After the simulated DoS attack was over, Wireshark was used for exporting captured traffic in CSV format. The exported data is extremely easy to handle and natively compatible with machine learning models. In this phase, the dataset rows were labeled, based on the category; packets from the flood command hping3 were labeled as "attack," and all the other packets were considered to be "normal." Labeled data forms the basis used in most of the model training, testing, and evaluation in research studies for intrusion detection, especially DoS patterns. This experimental design ensured the collection of the dataset representative of attack and benign traffic under totally controlled conditions for robust development of cybersecurity models.

2.5 Data Processing and labelling

Science Management Design Journal
Journal Homepage: www.smdjournal.com
ISSN: 2583-925X
Volume: 6
Issue: 2
Pages: 22-28

This is during the data pre-processing stage in which network traffic logs are cleaned to make all features uniform. For instance, numerical variables such as packet size were normalized in the same way. Records were classified as either an attack or normal instance according to predefined criteria associated with DoS attacks, which made binary classification in machine learning tasks easy. The data was then formatted into CSV format and hence made accessible and ready for use in the training and validation of the NIDS model.

3. Results and Discussion:

The dataset contains network traffic data with multiple entries relating to TCP packets between two IP addresses, namely 192.168.1.12 and 192.168.1.11. The packets depict both SYN (synchronize) flags and ACK (acknowledge) flags being set, thus showing a pattern of the TCP handshake process. Most of the entries in the dataset have marked as "Malicious" with high indication of malicious activity. The column "High_Frequency" has predominantly marked as "1, which indicates that these packets constitute a repeated communication pattern.

Majority of the entries, have "Malicious" activity. SYN packets sent from the destination 192.168.1.11 to 192.168.1.12 had an arbitrary number of repeated SYN requests to the same port (135), which is a characteristic feature of a SYN flood attack, one of the common Denial-of-Service attacks. After eliminating these malicious entries, the rest of this dataset reflects mostly normal network behavior with no significant deviation from typical communication behavior.

Result Analysis
The result analysis does reveal that most of the traffic does not come through as valid, and within a short period of time, the high-frequency SYN packets could be scanning or attacking activities but perhaps especially on port 135. There is an alert to be investigated or watched for systems that monitor such anomalies typical of network scanning or flooding attempts.

4. Conclusion:

In summary, the "DOS ATTACK" dataset is one of the most comprehensive and valuable sources used to assist in the development and study of AI-based NIDS. With more than 820,000 instances of data, it offers not only normal network traffic but numerous patterns of denial-of-service attacks, like the common SYN flood DoS attack type. Such major attributes of the dataset, such as source and destination IP addresses, protocol types, packet sizes, and TCP flags, are used for distinguishing between benign and malicious network behavior. The imbalance of the dataset in which only 17 percent of the instances represent attack traffic, thus reflecting real-world conditions, would make it a much more realistic tool for machine learning model evaluation. Detailed analysis of the dataset will highlight specific patterns, including unusual frequencies of SYN packet anomalies, that can lay the basis for developing even better and more adaptive intrusion detection systems. The dataset, therefore, will increase real-time attack detection capabilities as well as diminish false positives to strengthen cybersecurity measures against evolving network threats.

Science Management Design Journal
Journal Homepage: www.smdjournal.com
ISSN: 2583-925X
Volume: 6
Issue: 2
Pages: 22-28

Acknowledgement: I would like to express my sincere gratitude to all those who contributed to the development of this research and dataset. My deepest appreciation goes to the professors and colleagues for their invaluable guidance and support throughout this project. Special thanks to the team responsible for gathering and curating the dataset, whose efforts have provided a crucial foundation for this study. I would also like to acknowledge the tools and resources, such as Wireshark and Kali Linux, which made the data collection and analysis process possible. Lastly, I appreciate my family and friends for their continuous encouragement and understanding during this research.

References:

[1] Anderson, R. (2020). *Security Engineering: A Guide to Building Dependable Distributed Systems* (3rd ed.). Wiley.

[2] Shaikh, K., Chilwan, S., & Shaikh, R. (2024). Evaluation of Machine Learning Techniques for Network Intrusion Detection Systems.

[3] Bai, X., & Chen, W. (2019). "A Survey of Machine Learning Algorithms for Intrusion Detection Systems." *Journal of Cyber Security and Privacy*, 1(2), 109-124. https://doi.org/10.1002/cyber.1134

[4] Benassi, S., & Brooks, R. (2021). "The Application of AI and Machine Learning in Network Intrusion Detection." *International Journal of Computer Applications*, 12(3), 56-70. https://doi.org/10.5121/ijca.2021.12603

[5] Burrell, P., & Jones, S. (2018). "Detecting Distributed Denial-of-Service Attacks Using Machine Learning Techniques." *Proceedings of the 5th International Conference on Network Security*, 125-138. https://doi.org/10.1109/ICNS.2018.00022

[6] Chand, R., & Kumar, P. (2020). "Anomaly-Based Intrusion Detection Systems: A Comprehensive Survey." *Computer Networks and Communication Systems*, 8(1), 100-115. https://doi.org/10.1016/j.comnet.2019.07.015

[7] Dhanasekaran, R., & Rajasekaran, M. (2022). "Machine Learning Approaches in Cybersecurity for DoS Detection." *Journal of Computational Security*, 18(4), 433-450. https://doi.org/10.1145/3501300

[8] Glover, S., & Anderson, T. (2019). "Evaluating AI Techniques for DoS Detection in Large Scale Networks." *Journal of Artificial Intelligence Research*, 42(3), 245-259. https://doi.org/10.1007/978-3-319-95789-1_32

[9] Khan, M., & Iqbal, M. (2020). "A Novel Machine Learning-Based Approach for DoS Attack Detection in IoT Networks." *IEEE Transactions on Network and Service Management*, 17(2), 1081-1094. https://doi.org/10.1109/TNSM.2020.2962478

[10] Li, Z., & Zhang, Q. (2021). "Deep Learning for DoS Attack Detection and Prevention." *Journal of Network Security*, 10(3), 203-210. https://doi.org/10.1016/j.jnse.2021.03.007

[11] Liu, Y., & Liu, S. (2019). "Intrusion Detection in Network Traffic Using Machine Learning: A Review." *Journal of Computer Science and Technology*, 34(2), 120-134. https://doi.org/10.1007/s11390-019-1902-5

Science Management Design Journal
Journal Homepage: www.smdjournal.com
ISSN: 2583-925X
Volume: 6
Issue: 2
Pages: 22-28

[12] Mahajan, A., & Sharma, V. (2021). "A Study on Denial of Service Attack Detection Using Machine Learning Algorithms." *International Journal of Computer Applications*, 174(6), 18-28. https://doi.org/10.5120/ijca2021921861

[13] Mavridis, N., & Papanikolaou, K. (2020). "Cyber Threat Detection Using Machine Learning: The Case of DoS Attacks." *Proceedings of the 2nd International Conference on Artificial Intelligence for Cyber Security*, 92-105. https://doi.org/10.1007/978-3-030-31292-7_9

[14] Suryawanshi, Y., Patil, K., & Chumchu, P. (2022). VegNet: Dataset of vegetable quality images for machine learning applications. *Data in Brief*, *45*, 108657.

[15] Yadav, S., & Gupta, M. (2021). "Detection of DoS Attacks Using AI Techniques: A Review." *International Journal of Network Security*, 23(4), 505-518. https://doi.org/10.38025/ijns.2021.0050