Machine Learning-Based DDoS (Distributed Denial of Service) Detection System

Report Name	Project Outline
Author (User Id)	Thomas Roethenbaugh (tpr3)
Supervisor (User Id)	Muhammad Aslam (mua19)
Module	CS39440
Degree Scheme	G401 (Computer Science (with integrated year in industry))
Date	03 February 2025
Revision	1.1
Status	Release

1 Project description

This project aims to create a machine learning (ML) based distributed denial of service (DDoS) detection model that could be used to monitor network traffic with a minimum 99% detection rate, as well as explore the most recent DDoS attacks and their effects. This will be accomplished by using the CIC-DDOS2019 [1] and NSL-KDD [2] datasets, which simulate various types of networking attacks and intrusion simulations such as UDP floods, ICMP floods, SYN floods, fragmented packet attacks and others. Both these datasets are split between training and testing which will ease the creation of said machine learning model, as well as measure its performance. This model will try to try to combine multiple machine learning techniques such as random forest, support vector machine (SVN) and deep learning in an ensemble model simular to the one created in this paper [3] and this paper [4].

This topic holds a great deal of importance and relevance to our modern world. With the rise of cloud computing and web-based software as a service (SaaS), more and more systems have become reliant on high-availability networking. This is something that DDoS attacks are particularly good at disrupting as they flood these services with useless traffic which can incur huge cloud computing costs or take down entire vital software applications. An example would be the 2024 Azure/365 botnet attack [5] which this student experienced first-hand as they were working in the IT department of a multi-national company during their industrial year.

There is even evidence of DDoS attacks rising year after year according to a threat report created by NETSCOUT [6]. To further research in this field, several experiments will be performed with various models in order to determine which would be most effective. To measure this effectiveness, several metrics will be used when training the model.

This will include accuracy, which will measure the correctness of the model by comparing the total number of correct predictions (either true or false) to the total predictions made. Precision, which measures how many predicted detected attacks were correct and how many false alarms there were. Recall, which measures how many DDoS attacks were detected by the model. The F1-score, which is the mean of precision and recall. True positives, the number of correctly detected DDoS attacks and finally, false negatives. The rate at which the model fails to spot a DDoS attack.

2 Proposed tasks

- **Analyse datasets for suitability and notable features.** Before the model can be worked on, some data analysis will need to be conducted on each of the datasets. This will include feature extraction, potentially some data cleaning, as well as full understanding of how the data can be used to train several kinds of ML models.

- **Create and update a GitLab project, adhering to the Srum methodology** This GitLab project will not only be a great way of storing the project's data and code. But will also contain a weekly diary as well as tasks in the form of issues that will demonstrate what is in progress and what is completed. It will also be useful for backups.

- **Research DDoS attacks** Some time will need to be spent fully understanding the types of DDoS attacks and how they damage networking infrastructure. Research networking protocols (UDP, TCP, etc...) and understand how DDoS attacks exploit them enough to explain it succinctly in the final report. This will likely involve reading a few papers about them which will be included in the final report.

- **Train & test random forest, support vector machine (SVN) and deep learning models** Before this student can attempt creating an ensemble ML model, training and testing of the dataset will need to be done in order to understand what models are more or less accurate at detecting DDoS attacks. Some experimentation will also need to be performed on what models work well together to improve the accuracy of the model. The tools that will be used for this will be a Google Colab Jupiter notebook (Python).

- **Assemble multiple machine learning models into a new model** Depending on their effectiveness, several of these models will be merged into one new model that will be more effective at detecting DDoS attacks within the dataset.

- **Continuously write and update the technical report detailing what was achieved and the effectiveness of the model** The technical report will be 12,500 words and will contain all the details about the project, including problems and successes. Definitions and progression over time.

- **Demonstrations** Prepare for both the mid-project and end-of-project demonstration. This will require several of the previous tasks to have been either completed or mostly completed. Some time will also need to be spent on creating a presentation of the work so far, as well as how to explain it in 10 minutes.

- **Create a virtual test environment for the model to be tested in, this will include a software-defined network (SDN)** A stretch goal for this project will be to create a virtual testing environment using VirtualBox, Mininet, Python (+Networking libraries) similar to the ones created in [3]. This would be done to help understand this model's effectiveness against real attacks, as well as understand how it reacts to new data that was not originally from the dataset.

3 Project deliverables

- **An ensemble machine learning model using SVN and/or random forest with a minimum 99 per cent detection rate** - This machine learning model must also be able to detect both benign and malicious traffic in both the datasets as well as a simulated network environment.

- **A data report detailing the model's success in both the datasets and the virtual environment** - For the written aspect of the project, some documentation on how to use the software will be needed, as well as this there will need to be both explanations of data examples and data visualisations to explain the effectiveness of the program.

- **A 12,500 word report detailing progress, research, completed tasks as well as an overall explanation of the project** - This will include detail about the DDoS attacks themselves and how they are detected/mitigated. This will also include the mid-project demonstration report.

- **A GitLab project with code, data and metrics about progress** - The week-by-week process for this project will be an Agile methodology where weekly goals are set out with design, testing and implementation considered before deployment. This will allow for changes in requirements since this project might require changes mid-way through due to experiments or unexpected behaviour from the model.

- **A virtual machine containing a SDN** - A virtual Ubuntu machine containing a Mininet SDN that uses the machine learning model to detect and possibly even stop simulated DDoS attacks.

Refrences

[1] I. Sharafaldin, A. H. Lashkari, S. Hakak, and A. A. Ghorbani, "Developing realistic distributed denial of service (ddos) attack dataset and taxonomy," in *2019 International Carnahan Conference on Security Technology (ICCST)*, 2019, pp. 1–8.

[2] G. Mohi-ud din, "Nsl-kdd," 2018. [Online]. Available: https://dx.doi.org/10.21227/425a-3e55

[3] A. A. Alashhab, M. S. Zahid, B. Isyaku, A. A. Elnour, W. Nagmeldin, A. Abdelmaboud, T. A. A. Abdullah, and U. D. Maiwada, "Enhancing ddos attack detection and mitigation in sdn using an ensemble online machine learning model," *IEEE Access*, vol. 12, pp. 51 630–51 649, 2024.

[4] A. Ahmim, F. Maazouzi, M. Ahmim, S. Namane, and I. B. Dhaou, "Distributed denial of service attack detection for the internet of things using hybrid deep learning model," *IEEE Access*, vol. 11, pp. 119 862–119 875, 2023.

[5] K. O'Flaherty, "Microsoft confirms new outage was triggered by cyberattack," Jul. 31 2024, available: https://www.forbes.com/sites/kateoflahertyuk/2024/07/31/microsoft-confirms-new-outage-was-triggered-by-cyberattack/, Accessed: Jan. 31, 2025.

[6] "Key findings - latest cyber threat intelligence report," Sep. 24 2024, available: https://www.netscout.com/threatreport/key-findings/, Accessed: Jan. 31, 2025.

[7] W. I. Khedr, A. E. Gouda, and E. R. Mohamed, "Fmdadm: A multi-layer ddos attack detection and mitigation framework using machine learning for stateful sdn-based iot networks," *IEEE Access*, vol. 11, pp. 28 934–28 954, 2023.