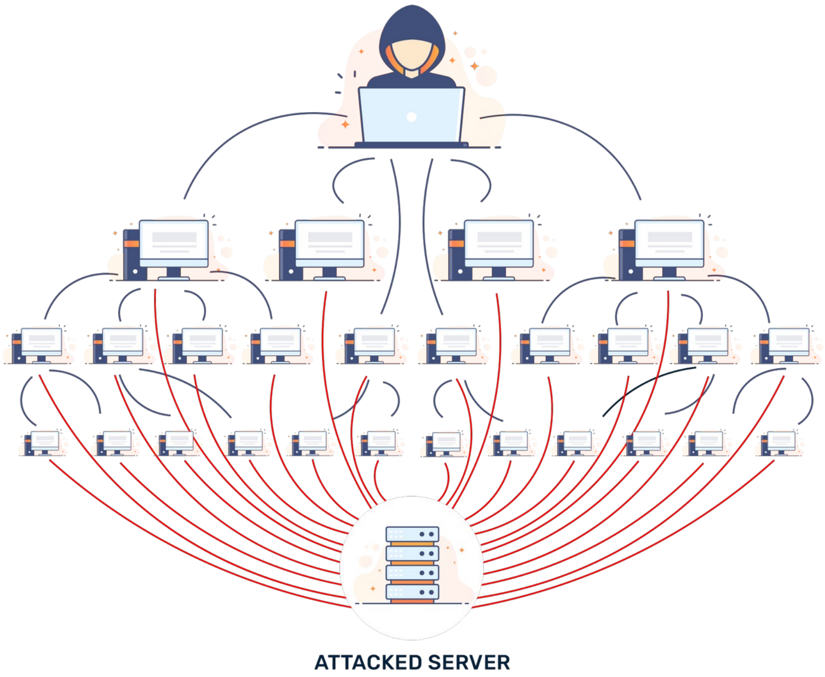
ML4N - Group Project 5

DDoS attacks detection and characterization

Clarifications for this project can be asked to Giordano Paoletti: giordano.paoletti@polito.it



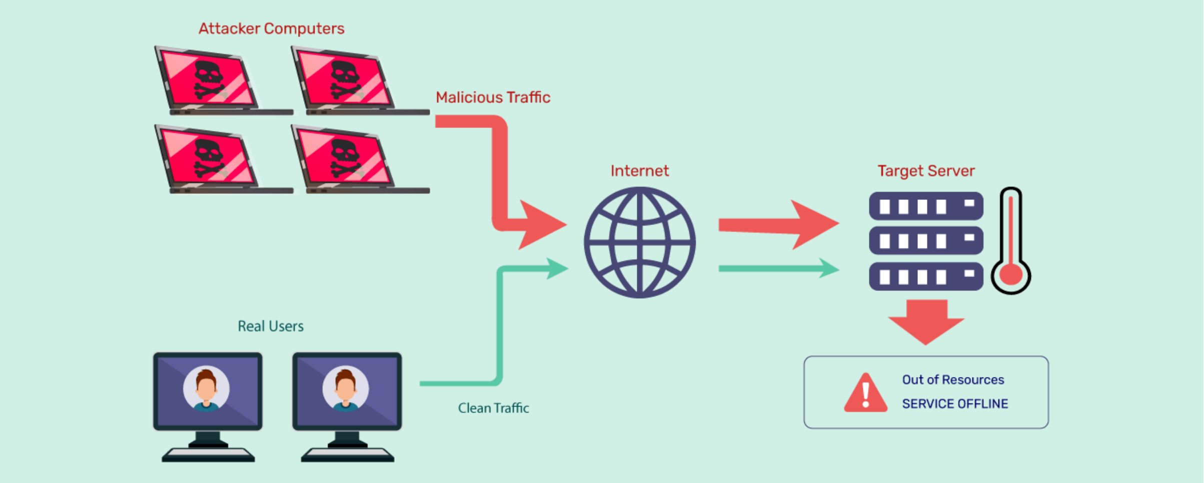
Internet security is one of the most important challenges, especially when the demand for IT services is increasing every day. Among the many existing threats, DDoS (Distributed Denial of Service) attack is a relatively simple but very effective technique to attack intranet and Internet resources. Typically, this attack uses a large number of compromised machines to prevent legitimate users from using web-based services. DDoS attacks can be carried out at the network, transport and application layers using various protocols such as TCP, UDP, ICMP, and HTTP.

By assuming that different DDoS attacks exhibit different traffic patterns, researchers focused on the application of ML algorithms to detect and characterized such patters. Indeed, the automatic detection of DDoS attacks can ease the network monitoring activity of network administrators and allow to quickly take countermeasures.

The goal of this project is to develop a complete Machine Learning pipeline to automatize the detection and analysis of flows generated during DDoS attacks solving two tasks: (i) one supervised task, i.e., classification and (ii) one unsupervised task i.e., clustering. Through the analysis of the results coming from the two tasks you should be able to identify and characterize flows generated during a specific DDoS attack understanding, when possible, the attack behaviors.

The provided dataset contains benign and the most recent common DDoS attacks that resemble the real world data. It also includes the results of network traffic analysis using CICFlowMeter-V3 (https://www.unb.ca/cic/research/applications.html#CICFlowMeter) with labeled flows-based on timestamps, source and destination IPs, source and destination ports, protocols and attacks. The dataset has been built replicating the behaviour of 25 users based on the HTTP, HTTPS, FTP, SSH and email protocols.
In this dataset, you have different modern reflective DDoS attacks such as NetBIOS, LDAP, MSSQL, UDP, UDP-Lag, SYN, NTP, DNS and SNMP. Attacks were subsequently executed during the data acquisition period period.
The ground truth (GT) is made of the name of the attack a considered flow is referred to.

For more details regarding the features check the file readme.md.



Section 1 – Data exploration and pre-processing

Data characterization and features engineering: explore the dataset and learn about the behavior of features at different levels, e.g., flow, ip, ports, etc. Then,generate the features allowing to succesfully accomplish the other tasks.

The first task of the project is to present the dataset through various data visualization techniques and statistical analysis.

- Investigate the provided dataset.
- Produce different visualizations and statistical analysis both at the generic traffic level (e.g., number of flows, etc.) and GT level. (e.g., distribution of features, GT class characterization, ECDF of ports, flows, etc.)
- Generate additional features e.g., quantifying the traffic related to each flow on the basis of the previous analysis (e.g. avg, min, max, quantiles, etc.)

- Perform correlation analaysis and visualization through PCA. If you think it could improve the tasks solution, you can do dimensionality reduction for generating the features used in the next tasks.
- Evaluate if you need to scale or standardize data
- Characterize the new final features, by producing plots regarding distributions of features (EPDF or ECDF), and correlation analysis.

Section 2 – Supervised learning – classification

The second task consists on classifying the flows according to the attack – supervised classification. You are provided by a ground truth containing the label of the attacks. In the provided GT you can find different names of the attacks.

1. Perform a split to segment the dataset into training and test dataset, in a stratified way with respect to the labels.
2. Choose at least 3 ML methods, and perform the model training, with default parameter configuration, evaluating the performance on both training and test set. Output the confusion matrix and classification report. Do you observe overfitting or under-fitting?
3. Tune the hyper-parameters of the models through cross-validation. How do performance vary? Which model generates the best performance?
4. Investigate the False Positive and False Negative. Can you draw considerations about the misclassification in terms of features? Report your analysis and findings for the ones you consider the most notable samples.

Section 3 – Unsupervised learning – clustering

In this task you will group flows that produce similar/correlated/coordinated patterns. The clustering will be done in an unsupervised fashion, independently of the labels (i.e., the attack label) used in Section 2. (clustering, unsupervised task). The goal is to understand if there exist simialr “families” of attacks.

Choose at least 2 Clustering Algorithms, and for each of them:

1. Determine the number of clusters: This can be done using methods like the elbow method or silhouette analysis. Explain your reasoning.
2. Find the best hyper-parameters, if any
3. Evaluate the clusters through clustering metrics and performance indicators
4. Report a coarse analysis of the detected clusters (e.g. ECDF of number flows per cluster, silhouette, etc.).

Section 4 – Clusters explainability and analysis

Cluster explainability and analysis: characterize the found clusters in terms of features distribution and activity patters drawing considerations about darknet traffic analysis.

The fourth and final task is to examine the detected clusters and try to find new patterns.

1. Do clusters reflect the GT labels ? What is ECDF of number of clusters assigned to each class? Are there pure clusters where all elements belong to a single class? Remember that clustering is unsupervised, hence the GT is not used in the clustering algorithm. Is there benign traffic with similar characteristics to malicious one?
2. What are the most important features in the obtained clusters? (you can use a method that provide feature importance or use an explainability technique (https://en.wikipedia.org/wiki/Explainable_artificial_intelligence)
3. Can you spot sub-attacks? Because of which feature? Can you find new groups or similar clusters? Why? According to which feature?
4. Which of the attacks are more similar? According to which feature?

Project acknowledgment: Luca Gioacchini luca.gioacchini@polito.it