Applications of Machine Learning to DDoS Attack Detection and Prevention

Presentation Date: November 20, 2024

CSE 543 Group 16
Team Leader: Laela Olsen (1220948755)
Deputy Leader: Anchala Balaraj (1233434009)
Team Members: Jin Heo (1212846750), Ahmad Karimi (1233868989), Yesha Modi (1232679307), Vignesh Mohan (1233212372), Nikhil Swami (1233379331), Advaith Venkatsubramanian (1233367072)

Arizona State University

Background

1. DDoS Threat Landscape

- DDoS attacks cause service disruption by overwhelming systems with excessive traffic from distributed sources.

1. Types of DDoS Attacks

- Volumetric attacks
- Protocol attacks
- Application layer attacks

3. Why Machine Learning for DDoS?

- Ability to identify complex patterns
- Can adapt to evolving attack patterns

4. Traditional Detection Methods

- Pattern matching identification fails against new variants
- Threshold-based systems struggle with attack complexity and detection speed

Problem Statement

- Distributed Denial of Service (DDoS) attacks can cause large scale damage to systems, economies and reputations.
- Our objective is to analyze a variety of machine learning algorithms which can detect malicious patterns based on a collection of attributes.
- In addition to leveraging various algorithms, the objective was also to conduct a deep-dive into the data and look for attributes which can contribute significantly to the possibility of a DDoS attack.
- DDoS attacks are polymorphic, hence dynamic and flexible methods need to be researched to keep such attacks in check.

Proposed Idea & Methodology

Question: What approach to take, practical or theoretical?

Answer: For the given problem statement, the study focuses more on practical approach.

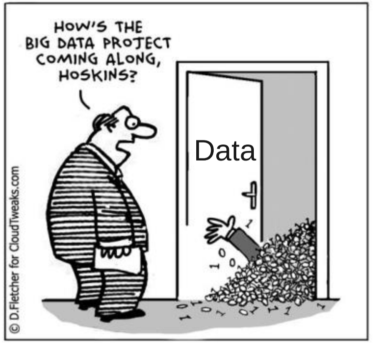
Practical

Theoretical

Proposed Idea & Methodology

Choosing the right dataset

- Using the CICIDS2017 dataset because it contains large amounts of data.
- The dataset also contains realistic simulated network traffic.
- This dataset is crucial as it includes both benign and malicious traffic patterns, allowing for robust model training and testing.


HOW'S THE BIG DATA PROJECT COMING ALONG, HOSKINS?
Data
© D.Fletcher for CloudTweaks.com

Proposed Idea & Methodology

Choosing the right ML algorithm

- The optimal machine learning algorithm for this task is not yet determined.
- Therefore, selecting the most appropriate machine learning algorithm becomes the primary challenge.
- Given the nature of the dataset, classification algorithms appear to be the most suitable choice.
- Our study primarily focuses on classification algorithms and also incorporates neural network approaches.

Methodology

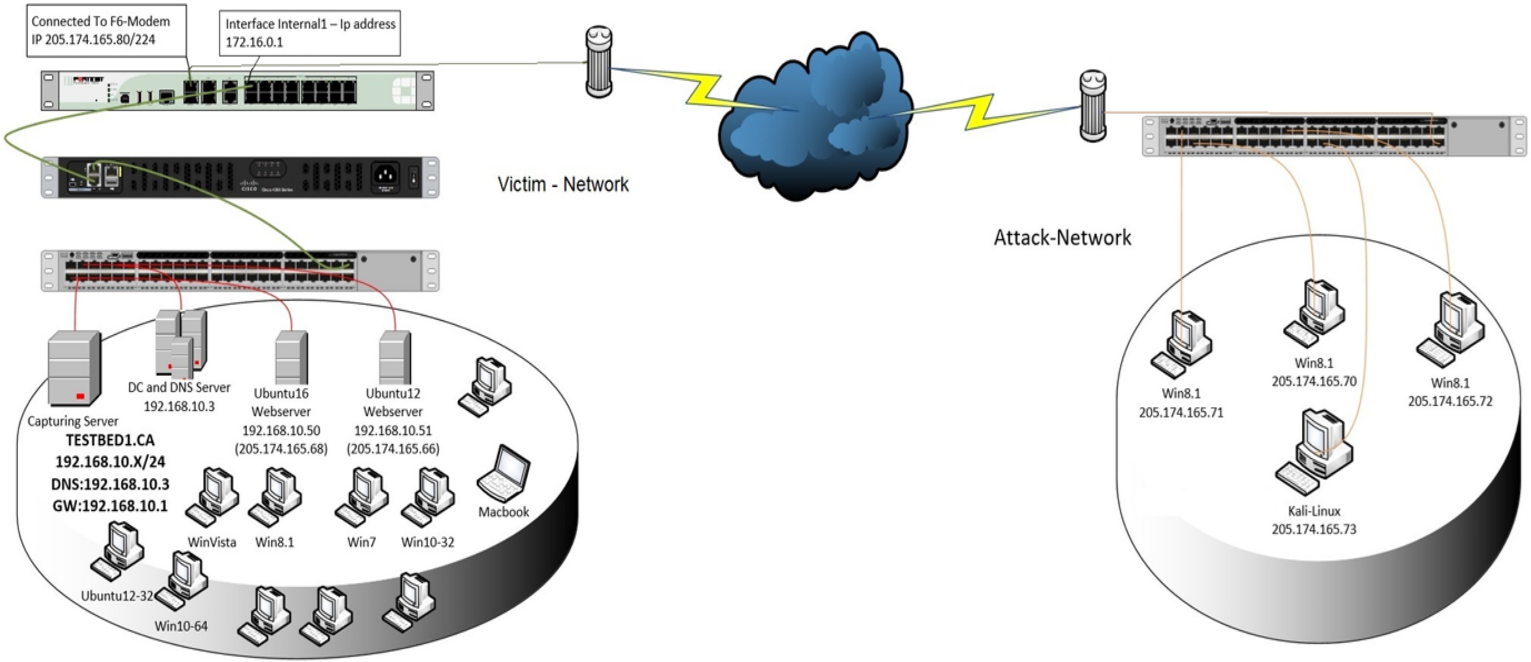
- The core part of methodology is to examine and compare the performance of ML models in identifying DDoS attacks accurately with the help of CICIDS2017 dataset.
- The methodology starts with data preprocessing on the dataset which is then divided into training (80%) and testing (20%) sets to ensure sufficient data for training the model and its evaluation.
- Further, ML models - Artificial Neural Network, Convolutional Neural Network, Decision Tree, Logistic Regression, Naive Bayes, Random Forest, Support Vector Machine, and k-Nearest Neighbor are implemented to classify network traffic as either Benign or DDoS.
- Model performance is evaluated using accuracy, precision, recall, F1 scores, and confusion matrices.

Implementation

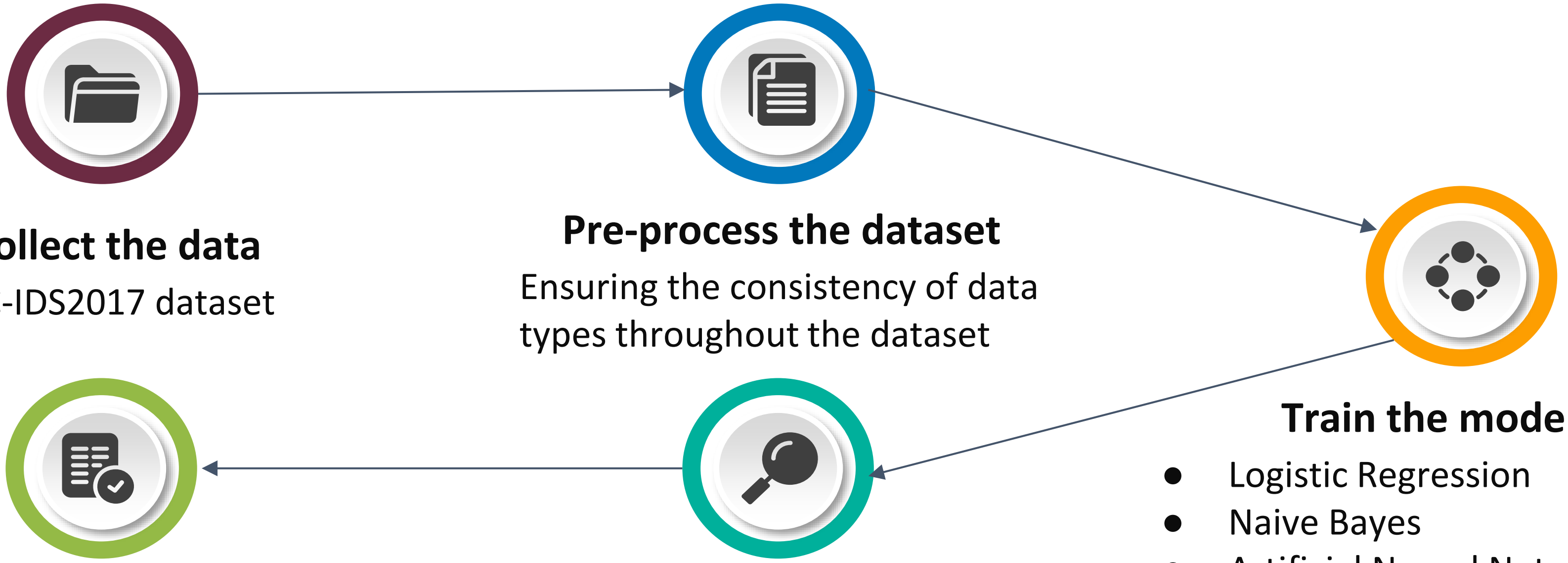
Published by the Canadian Institute for Cybersecurity, CICIDS2017, the dataset, consists of labeled network flows designed to mimic real network data. The data contains both benign traffic and various cybersecurity attacks, covering a diverse set of attack scenarios including **Brute Force FTP**, **Brute Force SSH**, **DoS**, **Heartbleed**, **Web Attack**, **Infiltration**, **Botnet**, and **DDoS.**

The dataset includes around 80 traffic features, with the ones most pertinent to our study being **backward packet length**, **average packet size**, and ones related to **inter-arrival time**.

Testbed Architecture



Workflow



Collect the data

CIC-IDS2017 dataset

Pre-process the dataset

Ensuring the consistency of data types throughout the dataset

Train the models

- Logistic Regression
- Naive Bayes
- Artificial Neural Network
- Convolutional Neural Network
- Random Forest
- Decision Tree
- Support Vector Machines

Evaluate the models

Accuracy, Precision,
Recall, F1 Score
Confusion Matrix

Model comparison

Performance of the different models is compared using ROC curves and AUC scores to identify the best-performing model for DDoS attack classification

Conclusion

- This study rigorously evaluates the effectiveness of eight machine learning algorithms for DDoS detection on the CICIDS2017 dataset. Each model demonstrated moderate to high accuracy: Random Forest and Decision Tree both achieved the highest accuracy at 99.98%, proved most effective for DDoS detection, providing high sensitivity with minimal false positives.
- Decision tree-based algorithms excelled, underscoring the value of ensemble learning for complex network data.
- Further optimization, including hyperparameter tuning and dimensionality reduction, is vital for real-time applications with high traffic, ensuring robust, efficient DDoS detection.

Deficiencies Summary

Total number of deficiencies submitted	37
Number of merged deficiencies	22
Number of valid merged deficiencies	12
Number of invalid merged deficiencies	10

Peer Summary

Peer Name	Number of Deficiencies	IDs
Saud Alsaif	V 5 I 3	V: 1-2, 2-1, 2-2, 3-2, 3-3 I: 1-1, 3-1, 5-3
Melanie Hendricks	V 1 I 2	V: ?-1 I: ?-2, ?-3
Shrajal Kelkar	V 2 I 1	V: 1-1, 5-1 I: 2-1
Kumar Hasti	V 3 I 1	V: 1-1, 2-1, 3-1 I: 5-1
Pravalikka Putha	V 2 I 2	V: 1-1, 2-2b I: 2-1, 2-2a
Sai Tharun Venkatramanan	V 2 I 5	V: 3-1, 5-1b I: 1-1, 1-2, 2-1, 3-2, 5-1a
Kashyap Laxmikant Patel	V 2 I 5	V: 1-1, 2-3 I: 1-2, 2-1, 2-2, 5-1, 5-2
Spoorthi Uday Karakaraddi	V 0 I 1	I: 5-1