BOTNET ATTACK DETECTION

Lab Project 21.11.2018

Information Security Lab 15B17CI576

Naima Farooqi
16803005

Aditya Gupta
16803004

Kanishka Garg
16803012

Harsh Vishnoi
16803030

Batch : B13

Submitted To : Mrs. Gagandeep Kaur

INTRODUCTION

In October 2016, the Mirai botnet took down domain name system provider Dyn, waking much of the world up to the fact that Internet of Things devices could be weaponized in a massive distributed denial of service (DDoS) attack. Although DDoS attacks have been around since the early days of the modern internet, IT communities around the globe came to realize that IoT devices could be **leveraged in botnet attacks to go after all kinds of targets**.

In the case of Dyn, the cyberattack took huge chunks of the web offline, since Dyn served as a hub and routing service for internet traffic. The attack temporarily shut off access to Twitter, Netflix, Spotify, Box, GitHub, Airbnb, reddit, Etsy, SoundCloud and other sites.

What Is a DDoS Attack?

A DDoS attack is a cyberattack in which **multiple compromised systems attack a given target**, such as a server or website, to deny users access to that target.

Attackers often use compromised devices — desktops, laptops, smartphones or IoT devices — to command them to generate traffic to a website in order to disable it, in ways that the user does not even detect.

The malware then visits or sends special network packets (OSI Layer 7 and Layer 3, respectively) to the website or DNS provider. The attack then generates what looks like, to most cybersecurity tools, normal traffic or unsuccessful connection attempts.

What Is a Botnet Attack?

Botnet attacks are related to DDoS attacks. Not all botnets are malicious; a botnet is a simply **a group of connected computers working together to execute repetitive tasks**, and can keep websites up and running. However, malicious botnets use malware to take control of internet-connected devices and then use them as a group to attack.

“More often than not, what botnets are looking to do is to add your computer to their web,” a blog post from anti-virus firm Norton notes. “That usually happens through a drive-by download or fooling you into installing a Trojan horse on your computer. Once the software is downloaded, the botnet will now contact its master computer and let it know that everything is ready to go. Now your computer, phone or tablet is entirely under the control of the person who created the botnet.”

Malicious botnets are often used to amplify DDoS attacks, as well as sending out spam, generating traffic for financial gain and scamming victims.

The rise of the IoT makes botnets more dangerous and potentially virulent. The IoT means there are simply **many more (usually unsecured) connected devices for attackers to target**. As a result, the DHS/Commerce report notes, “DDoS attacks have grown in size to more than one terabit per second, far outstripping expected size and excess capacity. As a result, recovery time from these types of attacks may be too slow, particularly when mission-critical services are involved.”

Further, the report adds, traditional DDoS mitigation techniques, such as network providers building in excess capacity to absorb the effects of botnets, “were not designed to remedy other classes of malicious activities facilitated by botnets, such as ransomware or computational propaganda.”

METHODOLOGY

1. Introduction to Dataset

This was the dataset we scraped from Internet.

1	Serial Number
2	Botnet ID
3	Source IP Addresses
4	Source Port Addresses
5	Source IP ASN Number
6	Target IP Address
7	Target Port Address
8	Source IP Region
9	Source IP City
10	Source IP Latitude
11	Source IP Longitude
12	Threat Confidence

Table 1 : Dataset Columns

2. Encoding of Categorical Data

In python using **LabelEncoder** and **OneHotEncoder** from **sklearn's preprocessing Library** we encoded the "Threat Confidence Column [12]" in 0 and 1 for Low and High

Threat.Confidence	Threat_Classify
High	1
High	1
High	1
High	1
High	1

Table 2 : Threat Confidence Encoding

3. Extracting the Host Address from the Target IP Address

We noticed that from the feature of Target IP Address, the part which had any effect On the threat was just the Host Address. So we extracted it and made it into a separate column.

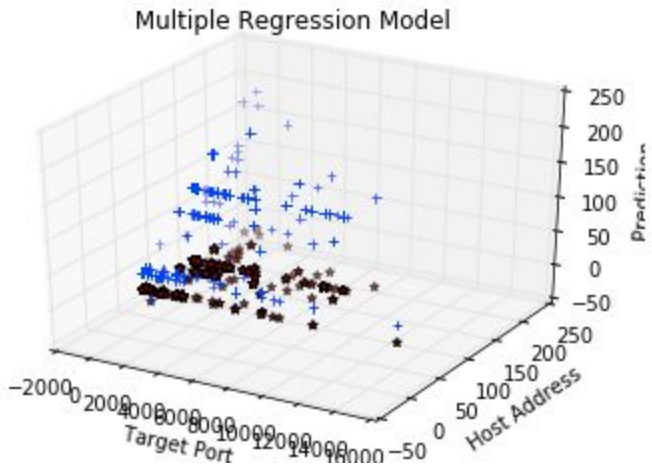
TargetIp
204.95.99.31
204.95.99.86
204.95.99.109
204.95.99.109
204.95.99.26
204.95.99.86
204.95.99.109
204.95.99.86

Host Address
31
86
109
109
26
86
109
86

Table 3 and 4 : Target IP and Host Address Column

4. Applying Multiple Regression To our Model

We applied Multiple Regression to our data the most relevant columns i.e. Target Port Address and Target Host Address as independent variables. We applied regression on Our threat classification and considered value greater than 0.9 as 1 or otherwise 0. The graph plotted from it is:



Graph 1: Regression Results

Here, X-axis: Target Port Address
Y-axis: Target Host Address
Z-axis: Threat variable

Blue points : Actual results
Red points : Predicted results
Purple points : Exact Intersection

5. Applying various Classification Techniques

The Classification techniques we applied are:

K - Nearest Neighbour Classification
Support Vector Machine Classification
Kernel Support Vector Machine Classification
Decision Tree Classification
Random Forest Classification

And we achieved different accuracy for each of these algorithms which we will discuss in results.

RESULTS

We achieved the best answer by **Decision Tree Classification Technique.**

Accuracy:

Technique	Wrong Predictions	Accuracy
K - Nearest Neighbour Classification	12	99.971%
Support Vector Machine Classification	20	95.215%
Kernel Support Vector Machine Classification	16	96.2085
Decision Tree Classification	0	100%
Random Forest Classification	5	98.845%

Table 5: Accuracy of classification predictions