Explore Watson Studio - Mini Project for DDoS attack

Group - 11 (Members : Jaspreet Singh, Amit Sharma, Deepa Vyasabhat, Ambika Na)

DDoS Attack Problem :

Distributed denial-of-service (DDoS) attack is a specific type of Denial of Service (DoS) attack where there is incoming traffic originating from multiple or distributed sources which results in flooding the victim which could be a targeted server, network or a service. This effectively makes it impossible to stop the attack simply by blocking a single source. Generally, the main targets for these attacks are public web services like banks, payment gateways and shopping sites to block users from accessing them.

Purpose of the DoS attack to make a machine or network resource unavailable to its intended users by temporarily or indefinitely disrupting services of a host. DDoS makes this attack more effective by using multiple compromised systems and resources including IOT devices.

Dataset for this Project:

Our Goal in this project is to predict possible DDoS attacks by using NIST vulnerability datasets (available at https://nvd.nist.gov/download.cfm). We started exploring the national vulnerability dataset of NIST but the available data was of very little help for predicting DDoS attack, So we searched for alternative datasets and used another application layer dataset from Kaggle.
https://www.kaggle.com/wardac/applicationlayer-ddos-dataset

The dataset has 809361 rows and 78 columns/features, and through using network analysis this dataset has been labeled into 3 categories:

1. **Benign** : Dataset of this category is normal data, without any attack.
2. **DoS slowloris** : This data is related to DoS attacks.
3. **DoS Hulk** : This data belongs to DDoS attacks.

Implementation Details :

As the dataset was having a very large amount of data to be processed, we used the pyspark module of Apache Spark with Python. We began by creating a spark session in Jupyter Notebook and loaded the complete dataset in pyspark.sql. As already

discussed before, there are three target labels but they had to be converted before we could apply any machine learning algorithm on the dataset.

So after doing some processing on the dataset, we then divided the dataset into training and testing dataset (70:30) so that we can train a Linear Regression Algorithm model on it. We then trained our regression model and tested the other dataset for evaluating the performance of the model. All of this was first done locally, and then the same thing was repeated on the Watson Studio available at IBM cloud and found similar results.

Conclusion :

The model results were then analysed and we found that the model was able to predict the DDoS attack with accuracy of about 93.04% , which is fairly good for the available dataset. In process, we also learned about the IBM Cloud , Watson Studio and Spark service for general purpose cluster computing. So, in conclusion it is safe to say that there a fixed pattern through which the vulnerabilities like DDoS attempt to attack and sabotage the network's or services and these can be predicted using the available data on these vulnerabilities, training a model and predict the possible attack and to either stop them from happening or take some other appropriate measures to minimize any loss.