**IBM PROJECT-MALWARE DETECTION**

GROUP 22

Sandeep Rajakrishnan

Sudhay Senthilkumar

Karthik Shriram.G.S

**AIM**

Detection and Prevention of Advanced Persistent Threat (APT) activities in heterogeneous networks using SIEM and Deep Learning.

**COMMON FEATURES**

* From the 3 datasets obtained namely UNSW\_NB15, IoT Botnet and NSL KDD, we were able to find out the influencing features and common features across these datasets.
* Our observation also includes a multiclass classified output with various types of attacks like DoS, Backdoor, Reconnaissance, etc…
* Hence, using these common features, we understand that when a new data point is provided and asked to classify under a type of attack with these respective columns’ data, we can find it’s type of attack.

\*Protocol

\*Service \* spkts

\*State \* src\_bytes

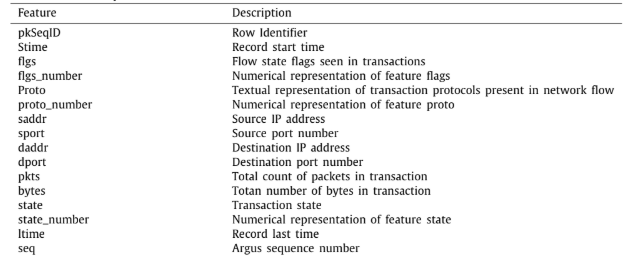
\*Duration \* dpkts

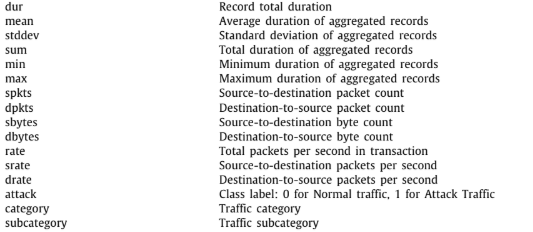
\*category / attack\_cat / labels \*dest\_bytes

**DATASET DESCRIPTION**

The dataset we have considered is UNSW\_2018\_IoT\_Botnet\_Full5pc\_4

The meaning of each column has been described below





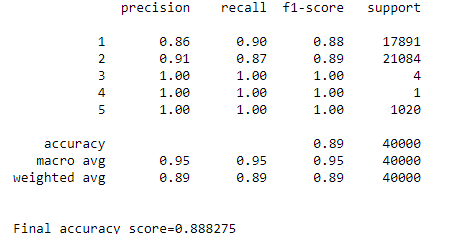
**ALGORITHM USED AND APPROACH**

* We have used the Support Vector Machine Algorithm on the dataset to obtain the confusion matrix and a classification report.
* To begin, we have performed :
  + Encoding - To give integer representations to strings
  + Scaling - Used StandardScaler() - to normalize all values
  + Obtained correlation between columns to find influential attributes
  + Split dataset into Test and Train datasets
  + Applied SVM algorithm and obtained a predicted dataset with test set
  + Compared with actual test dataset so as to obtain performance metrics

**PERFORMANCE MATRIX**

Accuracy Score = 0.888275

Algorithm Used : SVM



**REFERENCES:**

* Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics: Bot-IoT dataset

**LINK TO DATASET:**

* <https://raw.githubusercontent.com/defcom17/NSL_KDD/master/KDDTrain%2B.csv>
* <https://cloudstor.aarnet.edu.au/plus/s/umT99TnxvbpkkoE?path=%2FCSV%2FTraning%20and%20Testing%20Tets%20(5%25%20of%20the%20entier%20dataset)%2FAll%20features>

**GITHUB REPOSITORY :**

[**https://github.com/IBM-ML-PROJECT/GROUP-22-ML-project**](https://github.com/IBM-ML-PROJECT/GROUP-22-ML-project)