Network Intrusion Detection System with Machine Learning

Evangelou Sotirios, Kalais Konstantinos, Chatziefremidis Eleftherios

[1]Department of Electrical and Computer Engineering, University of Thessaly, Greece

Summary: A Network IDS system, that analyzes network packets and their metadata, in order to decide on the normality of the traffic, and the existence of an attack as well as its type.

Motivation

- Billions of Personal computers, Servers, Mobile phones, IOT devices and many more are connected each day on the Internet, and generally external or internal networks.
- The attack surface on network security widens by the day, and there is great need for well-performing Network Security Systems such as Firewalls, IDS, IPS etc.
- Handling billions of packets flowing through the network, and setting security rules to avoid attacks cannot be done effectively by individual security professionals.
- Smart systems that leverage the power of Machine Learning can perform well on analysing this kind of “Big Data”, and pave the road to autonomous security.

Introduction

- The first step in delving into the field of Security and Intrusion Detection with Machine Learning was to study papers[1][2] that focus on this particular matter.
- The primary methods for analysing network packets for security, referenced in these papers were mainly Support Vector Machines and Neural Networks.
- Based on the dataset, we face a supervised problem of multiclass classification (on attack type) or binary classification (on normal/abnormal network traffic).

Method

- Preprocessing the data was done by turning the nominal characteristics to numeric firstly, and dropping highly correlated features (85%) from the dataset.
- We are using FeatureSelector[3] library in order to find and drop highly correlated features.

- We approached three methods to analyse the problem, namely *SVM*, *Random Forest* and *Multilayer Perceptron*, leveraging the Keras framework and Scikit learn library.

- The **MLP** we use in the project is a deep neural network with three Fully Connected layers of 250-100-10 neurons.
- We found that the best approach was to classify by attack type and aggregate the attacks to a class of abnormal behaviour to decide on the normality of the traffic.
- We are using Adam optimizer, a categorical crossentropy loss function, and fit the model on scaled data to a range of (0,1).

- The **SVM** we use in the project uses a C value of 100 and a gamma value of 0.00001. This method proved to be very slow for the size of the dataset and problem.

- The **Random Forest** uses 1000 subtrees-estimators and a max-depth of 5. The results it produces are slightly better than the SVM’s performance but still not as good as MLP’s.

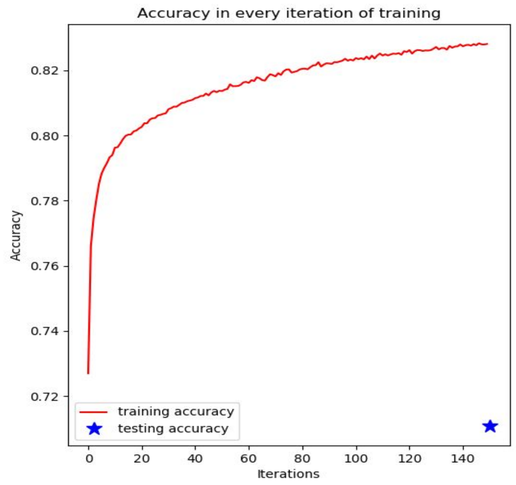
Data

- The dataset we used belongs to the Cyber Range Lab of the Australian Centre for Cyber Security (ACCS), and was created artificially with the IXIA PerfectStorm tool.
- The dataset is pre-split into a training set containing 175341 instances and a testing set containing 82332 instances.
- The dataset contains 43 features from which we remove 14 highly correlated features during the preprocessing phase.
- The dataset also provides 2 target columns, one according to the normality of the traffic (2 classes) and one according to the attack type (10 classes).

Experimental Results

- After training and evaluating with all three methods, we found MLP to be the most accurate choice for this dataset and problem.
- Namely, the results from SVM and Random Forest are:

SVM:
Normal/Abnormal Classification accuracy: 76.101%
Attack Type Classification accuracy: 68.087%
Random Forest:
Normal/Abnormal Classification accuracy: 81.336%
Attack Type Classification accuracy: 68.087%
MultiLayer Perceptron:
Training Normal/Abnormal classification accuracy: 82.81%
Normal/Abnormal Classification accuracy: 83.960%
Attack Type Classification accuracy: 71.075%



Conclusions

- A suitable solution for developing a trusted Network IDS can be Machine Learning.
- Specifically MLP produces high accuracy results with low number of false positive results, and can definitely be added to the production of any security oriented company or individual.

References:
[1]Mahdi Zamani, Machine Learning Techniques for Intrusion Detection, University of New Mexico, 2013.
[2]Srinivas Mukkamala, Guadalupe Janoski, Andrew Sung, Intrusion Detection: Support Vector Machines and Neural Networks , New Mexico Institute of Mining and Technology, 2002
[3] https://github.com/WillKoehrsen/feature-selector
[4] https://www.unsw.adfa.edu.au/unsw-canberra-cyber/cybersecurity/ADFA-NB15-Datasets/

Spring 2019, ECE 417 Poster Session