Project Team Forming and Ideation

CS-584 Machine Learning

Group members

- Gomez Valverde, Ignacio
- De Lucas García, Mario

List of project descriptions

Project 1 - Detect DDoS with AI

[An approach to detect DDoS]

[Paper que describe el dataset]

[Link para descargar PCAPs]

Qué vamos a hacer?

Preprocessing del dataset nosotros
Regresión logística propia
Research para aplicar alguna otra técnica para ver los paquetes

Resumen paper

We will focus on SYN flood, which is a specific type of TCP flood that targets the initial handshake of the TCP connection. The SYN flood sends a large number of SYN (synchronize) packets to the targeted server, but it never completes the handshake by sending the final ACK (acknowledge) packet.

Topology:

- The IoT topology deployed to produce the CICIoT2023 is illustrated in Figure 2 and comprises 105 IoT devices. A total of 67 IoT devices were directly involved in the attacks and other 38 Zigbee and Z-Wave devices were connected to five hubs. (Figure 2)
- Every packet sent through the network is stored in separate computers. In fact, the network has two different interfaces, which are associated with two other monitoring

ports that send incoming packets to these computers. Hence, the network traffic is monitored using Wireshark [47] and stored in pcap format. Since two data streams are stored, mergecap [48] is used to unify pcap files for each experiment. (es decir, un pcap es una ristra de paquetes que se envían por dos interfaces y tiene paquetes benignos y malignos)

- Hping3 is used for creating SYN flood attacks (see Table 2)
- 4,059,190 de rows en el dataset

Técnicas de ML que se han utilizado:

- Multiple features were extracted from the network traffic and used by a deep-learning autoencoder for attack detection.
- Multiple machine learning methods were used in the evaluation process (e.g., SVM, G-NB, LDA, and LR)focusing on attack detection and classification.

Lo que hacen ellos es clasificar el tipo de problema (combinan todos los datasets y predicen el tipo de ataque)

Table 6. Results obtained in the classification process conducted using different machine learning models (illustrated in Figure 10).

	Metric	Logistic Regression	Perceptron	Adaboost	Random Forest (RF)	Deep Neural Network (DNN)
34 classes	Accuracy	0.80231507	0.8195961	0.607888	0.99164365	0.986118011
	Recall	0.59520185	0.507506	0.607675	0.831586401	0.731868794
	Precision	0.486752461	0.454634	0.479621	0.704492066	0.665295126
	F1-score	0.49388408	0.4472933	0.473498	0.714021981	0.672346883
8 classes	Accuracy	0.831674188	0.8663152	0.351357	0.994368173	0.991147043
	Recall	0.696055597	0.6591315	0.487789	0.91001105	0.906642708
	Precision	0.512409686	0.5239188	0.464924	0.705407564	0.679434746
	F1-score	0.539424048	0.5551339	0.368663	0.71928904	0.69726491
2 classes	Accuracy	0.989023188	0.9817525	0.995899	0.99680798	0.994422814
	Recall	0.890400624	0.7970288	0.947303	0.965163906	0.933277496
	Precision	0.863157959	0.825432	0.965631	0.965395244	0.947579486
	F1-score	0.876258983	0.8105374	0.956273	0.965279544	0.940305998

-

Project 2