Machine learning cybersecurity

**pcap file feature extraction**

## Lesson 1 lab Solutions

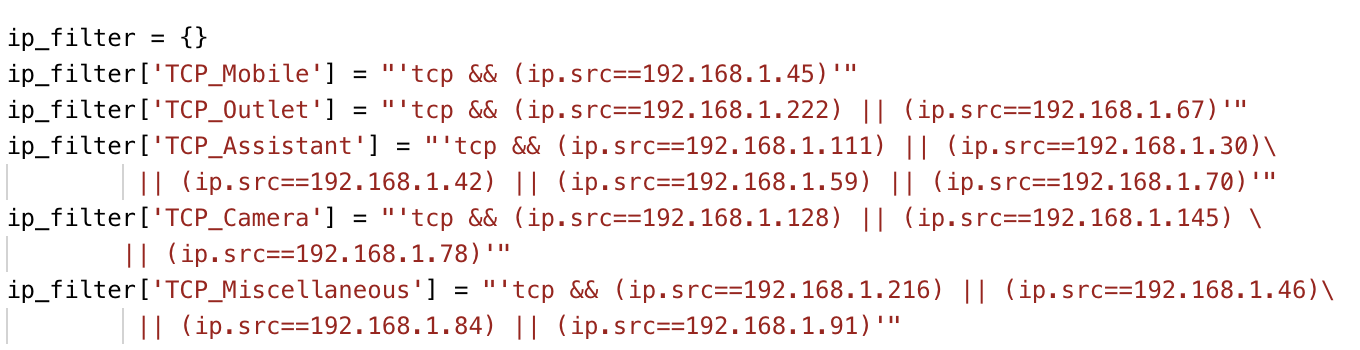
### Lab Exercise 1

This lab is to extract features from pcap files in order to represent the raw data in the vector space model.

1. Import required libraries

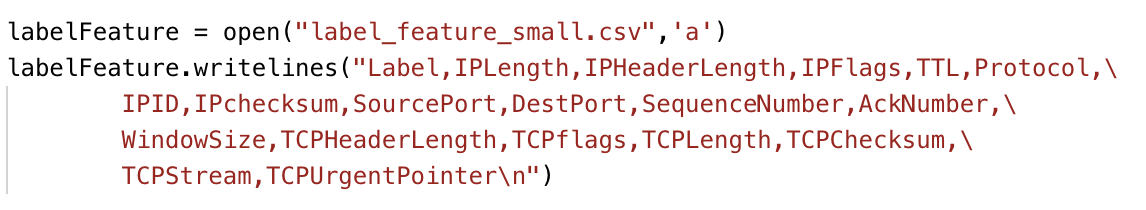
A. Import required libraries
A. Import required libraries
A. Import required libraries
Import os and glob

1. Define a dictionary to store the string for each device to filtering.



* ip\_filter is the name of the dictinoary
* 'TCP\_Outlet' is the key of the filtering string for a small power outlet device
* ' tcp && (ip.src==192.168.1.222) || (ip.src==192.168.1.67)' means to filter the tcp package from either 192.168.1.222 or 192.168.1.67, which are the ip addresses of outlet.

1. Open a csv file to store the labels and features. The header should be the name of features and the first column should be the label.



* label\_feature\_small.csv is the name of the file
* 'a' means append data to an existing file
* You need to write string 'label' and 18 features names into the file

1. Filter out all the packets from the 5 different devices in the original pcap files and save the result into 5 new pcap files.



* glob.glob will process all the pcap files in the original\_pcap folder
* -r means to read the local pcap file
* oriPcapFile is one of the pcap files in the original folder
* -w- and -Y means to write the packets matching the filter to the specified file
* ip\_filter[k] is the filtering command for the device k
* >> will append the filtering results to the pcap files in the filtered\_pcap folder
* The pcap files are named with k

1. Create the labels and extract the features for them with the newly generated pcap files.



* glob.glob will get all the pcap file names in the filtered\_pcap folder
* The names of the pcap files will be used to create the label
* Use tshark command to extract features from the currently processed file
* -r means to read the local pcap file
* ftdPcapFile is the name of the currently processed file
* Use several -e options for each feature (field) that you want to extract from the pcap file, such as -e ip.len
* -T option should be selected if you want to use -e option
* os.popen(tsharkCommand).read() will excute the command and save the result to the allFeatures parameter
* Before writing the label and features to the csv file, you need to convert the tab-separated results to the comma-separated results and breaking the results at line boundaries.

1. Reuslts

Reuslts


* + First row shows the label name (class) and feature names. Other rows show the label values and the feature values