Configuration Manual

MSc Internship
Cyber Security

Yash Shukla
Student ID: x18175104

School of Computing
National College of Ireland

Supervisor: Mr.Vikas Sahni

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Yash Shukla
Student ID:	x18175104
Programme:	Cyber Security
Year:	2019
Module:	MSc Internship
Supervisor:	Mr.Vikas Sahni
Submission Due Date:	17/08/2020
Project Title:	Configuration Manual
Word Count:	610
Page Count:	8

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

I agree to an electronic copy of my thesis being made publicly available on NORMA the National College of Ireland's Institutional Repository for consultation.

Signature:	Yash Shukla
Date:	14th August 2020

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

Yash Shukla
x18175104

Threat Hunting Using a Machine Learning Approach

1 Intended Audience

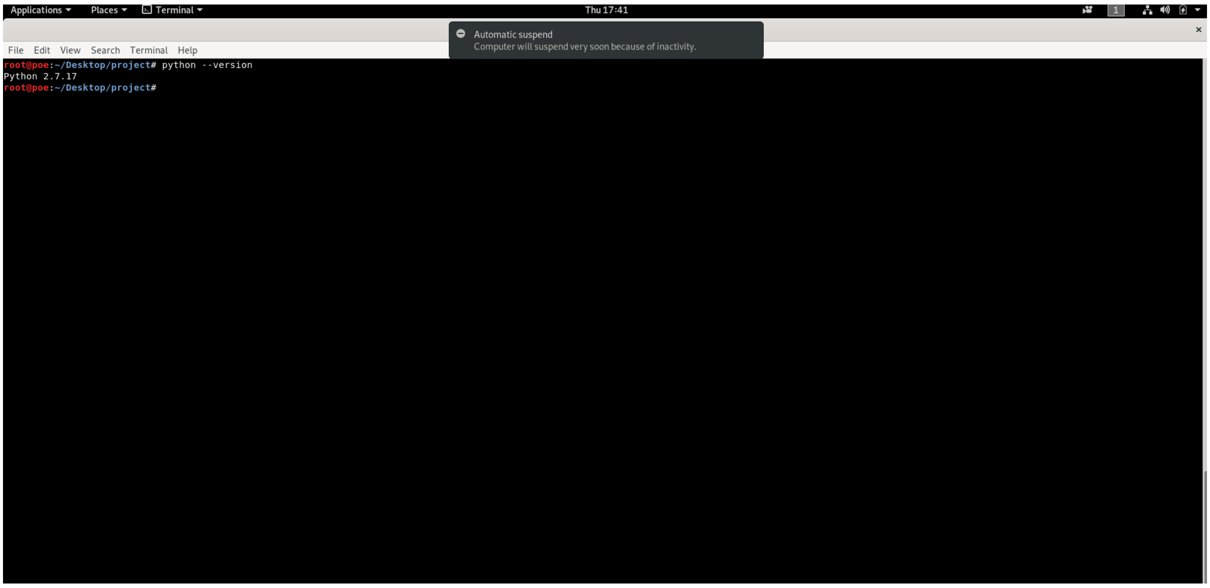
This document would be appropriate for researchers programmers or system administrators with a moderate degree of technical understanding of python. It explains how to configure a Linux system for executing python code and write adequate hypothesis required for threat hunting. [1].

Abbreviations:

1. IP= Internet Protocol
2. Pd= Pandas variable
3. Np= Numpy variable
4. plt= matplotlib.pyplot variable
5. cm= matplotlib.cm variable
6. NB= Naive Bayes
7. SVC= Support Vector Classification

2 Installing Python

Python is installed by default on some Linux distributions. (Note:This may not be the latest version or support all the functionality needed by the application)



Automatic suspend
Computer will suspend very soon because of inactivity.

```
root@poe:~/Desktop/project# python --version
Python 2.7.17
root@poe:~/Desktop/project#
```

Figure 1: Python Version Check

2.1 Steps for Ubuntu

- At first check the current installed version of python

  Command :python —version

- Get an update on your system

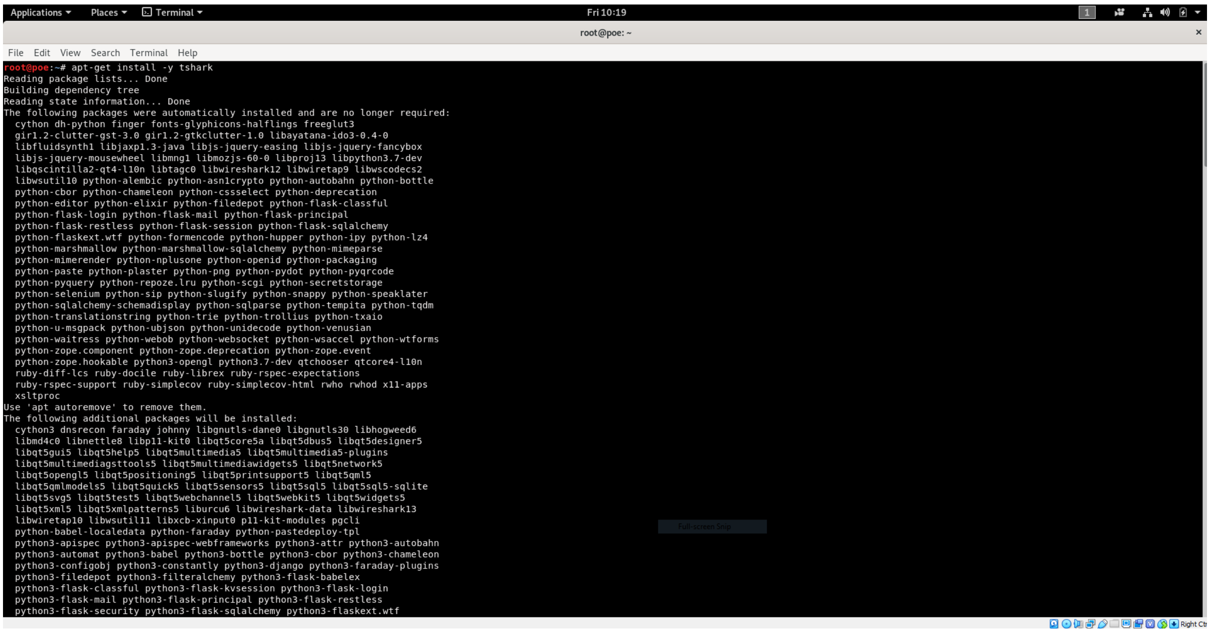
  command : sudo apt update

- Install python

  command : sudo apt install python(req version)

2.2 Data conversion steps

- The pcap file is processed by a program called Tshark. It is a network protocol analyzer, that has the ability to read saved network files. The program is responsible for the conversion of pcap files to text.



```
root@poe:~# apt-get install -y tshark
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages were automatically installed and are no longer required:
  cython dh-python finger fonts-glyphicons-halflings freeglut3
  gir1.2-clutter-gst-3.0 gir1.2-gtkclutter-1.0 libayatana-ido3-0.4-0
  libfluidsynth1 libjaxp1.3-java libjs-jquery-easing libjs-jquery-fancybox
  libjs-jquery-mousewheel libmng1 libmozjs-60-0 libproj13 libpython3.7-dev
  libqscintilla2-qt4-l10n libtagc0 libwireshark12 libwiretap9 libwscodecs2
  libwsutil10 python-alembic python-asn1crypto python-autobahn python-bottle
  python-cbor python-chameleon python-cssselect python-deprecation
  python-editor python-elixir python-filedepot python-flask-classful
  python-flask-login python-flask-mail python-flask-principal
  python-flask-restless python-flask-session python-flask-sqlalchemy
  python-flaskext.wtf python-formencode python-hupper python-ipy python-lz4
  python-marshmallow python-marshmallow-sqlalchemy python-mimeparse
  python-mimerender python-nplusone python-openid python-packaging
  python-paste python-plaster python-png python-pydot python-pyqrcode
  python-pyquery python-repoze.lru python-scgi python-secretstorage
  python-selenium python-sip python-slugify python-snappy python-speaklater
  python-sqlalchemy-schemadisplay python-sqlparse python-tempita python-tqdm
  python-translationstring python-trie python-trollius python-txaio
  python-u-msgpack python-ubjson python-unidecode python-venusian
  python-waitress python-webob python-websocket python-wsaccel python-wtforms
  python-zope.component python-zope.deprecation python-zope.event
  python-zope.hookable python3-opengl python3.7-dev qtchooser qtcore4-l10n
  ruby-diff-lcs ruby-docile ruby-librex ruby-rspec-expectations
  ruby-rspec-support ruby-simplecov ruby-simplecov-html rwho rwhod x11-apps
  xsltproc
Use 'apt autoremove' to remove them.
The following additional packages will be installed:
  cython3 dnsrecon faraday johnny libgnutls-dane0 libgnutls30 libhogweed6
  libmd4c0 libnettle8 libp11-kit0 libqt5core5a libqt5dbus5 libqt5designer5
  libqt5gui5 libqt5help5 libqt5multimedia5 libqt5multimedia5-plugins
  libqt5multimediagsttools5 libqt5multimediawidgets5 libqt5network5
  libqt5opengl5 libqt5positioning5 libqt5printsupport5 libqt5qml5
  libqt5qmlmodels5 libqt5quick5 libqt5sensors5 libqt5sql5 libqt5sql5-sqlite
  libqt5svg5 libqt5test5 libqt5webchannel5 libqt5webkit5 libqt5widgets5
  libqt5xml5 libqt5xmlpatterns5 liburcu6 libwireshark-data libwireshark13
  libwiretap10 libwsutil11 libxcb-xinput0 p11-kit-modules pgcli
  python-babel-localedata python-faraday python-pastedeploy-tpl
  python3-apispec python3-apispec-webframeworks python3-attr python3-autobahn
  python3-automat python3-babel python3-bottle python3-cbor python3-chameleon
  python3-configobj python3-constantly python3-django python3-faraday-plugins
  python3-filedepot python3-filteralchemy python3-flask-babelex
  python3-flask-classful python3-flask-kvsession python3-flask-login
  python3-flask-mail python3-flask-principal python3-flask-restless
  python3-flask-security python3-flask-sqlalchemy python3-flaskext.wtf
```

Figure 2: Tshark Installation

- Install Tshark using `sudo apt-get install -y tshark`
- This step is essential for taking the pcap data into a python program using dataframe. A dataframe is a type of data structure in python which helps access data in the form of rows and columns.

- At this phase data cleaning is necessary, all the columns with missing values are checked and the missing values are either replaced or their mean is inserted.
- Dataset to be downloaded from https://www.netresec.com/?page=PcapFiles The pcap file has the following features available for selection, which are inserted into the dataframe.

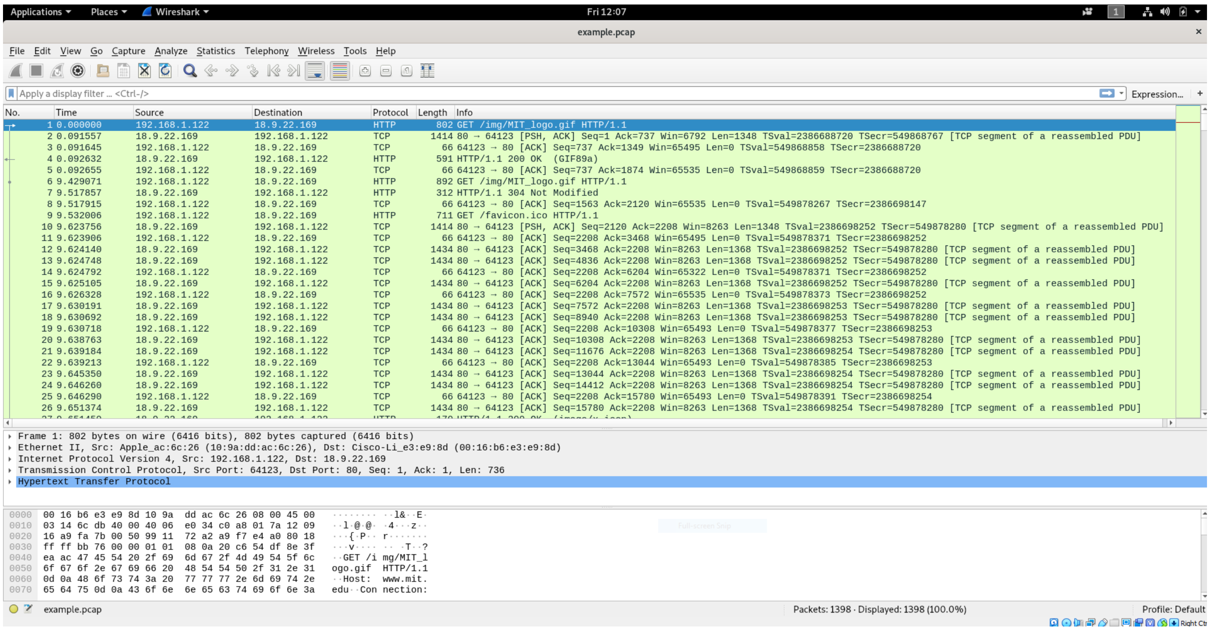
The following are the columns that are derived from the pcap conversion:

1. 'Sr.No'
2. 'Date'
3. 'Time'
4. 'Source_IP'
5. 'Arrrow'
6. 'Dest_IP'
7. 'ProtocolUsed'
8. 'Length','Method'

9. ‘link’

10. ‘Version’

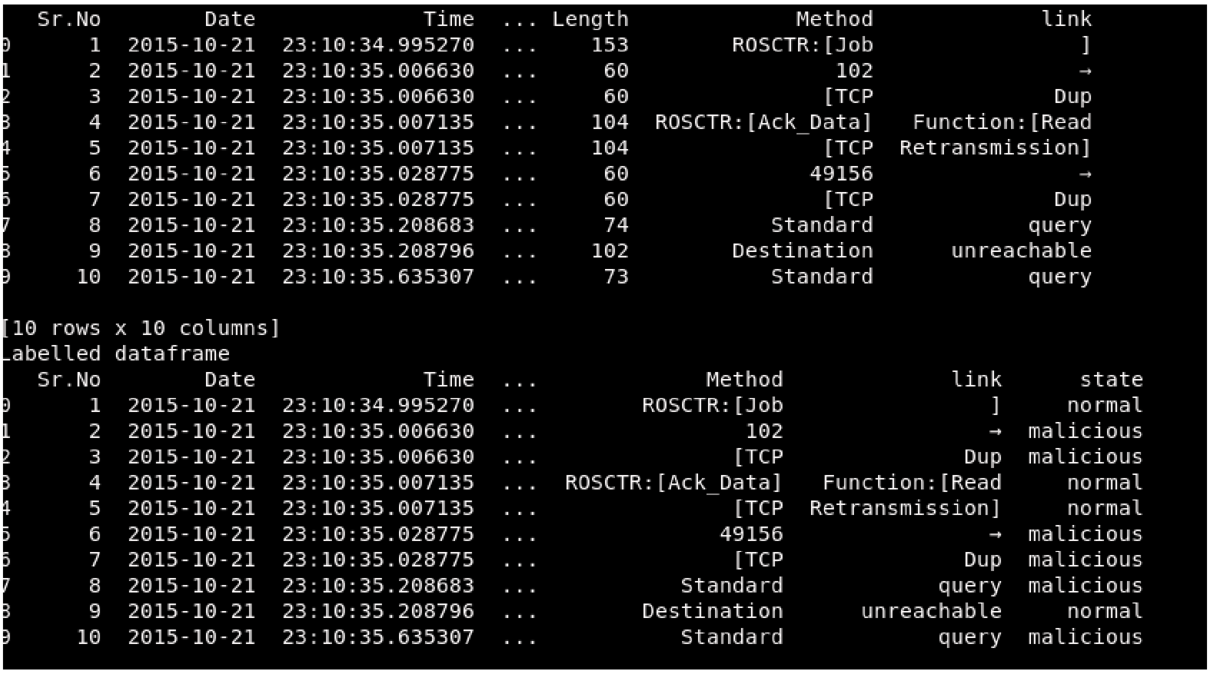
11. ‘Extra’



Applications
Places
Wireshark
Fri 12:07
example.pcap
File Edit View Go Capture Analyze Statistics Telephony Wireless Tools Help
Apply a display filter ... <Ctrl-/>
Expression...

```
No.   Time       Source          Destination     Protocol Length Info
    1 0.000000   192.168.1.122   18.9.22.169     HTTP      802 GET /img/MIT_logo.gif HTTP/1.1
    2 0.091557   18.9.22.169     192.168.1.122   TCP      1414 80 → 64123 [PSH, ACK] Seq=1 Ack=737 Win=6792 Len=1348 TSval=2386688720 TSecr=549868767 [TCP segment of a reassembled PDU]
    3 0.091645   192.168.1.122   18.9.22.169     TCP        66 64123 → 80 [ACK] Seq=737 Ack=1349 Win=65495 Len=0 TSval=549868858 TSecr=2386688720
    4 0.092632   18.9.22.169     192.168.1.122   HTTP      591 HTTP/1.1 200 OK  (GIF89a)
    5 0.092655   192.168.1.122   18.9.22.169     TCP        66 64123 → 80 [ACK] Seq=737 Ack=1874 Win=65535 Len=0 TSval=549868859 TSecr=2386688720
    6 9.429071   192.168.1.122   18.9.22.169     HTTP      892 GET /img/MIT_logo.gif HTTP/1.1
    7 9.517857   18.9.22.169     192.168.1.122   HTTP      312 HTTP/1.1 304 Not Modified
    8 9.517915   192.168.1.122   18.9.22.169     TCP        66 64123 → 80 [ACK] Seq=1563 Ack=2120 Win=65535 Len=0 TSval=549878267 TSecr=2386698147
    9 9.532006   192.168.1.122   18.9.22.169     HTTP      711 GET /favicon.ico HTTP/1.1
   10 9.623756   18.9.22.169     192.168.1.122   TCP      1414 80 → 64123 [PSH, ACK] Seq=2120 Ack=2208 Win=8263 Len=1348 TSval=2386698252 TSecr=549878280 [TCP segment of a reassembled PDU]
   11 9.623906   192.168.1.122   18.9.22.169     TCP        66 64123 → 80 [ACK] Seq=2208 Ack=3468 Win=65495 Len=0 TSval=549878371 TSecr=2386698252
   12 9.624140   18.9.22.169     192.168.1.122   TCP      1434 80 → 64123 [ACK] Seq=3468 Ack=2208 Win=8263 Len=1368 TSval=2386698252 TSecr=549878280 [TCP segment of a reassembled PDU]
   13 9.624748   18.9.22.169     192.168.1.122   TCP      1434 80 → 64123 [ACK] Seq=4836 Ack=2208 Win=8263 Len=1368 TSval=2386698252 TSecr=549878280 [TCP segment of a reassembled PDU]
   14 9.624792   192.168.1.122   18.9.22.169     TCP        66 64123 → 80 [ACK] Seq=2208 Ack=6204 Win=65322 Len=0 TSval=549878371 TSecr=2386698252
   15 9.625105   18.9.22.169     192.168.1.122   TCP      1434 80 → 64123 [ACK] Seq=6204 Ack=2208 Win=8263 Len=1368 TSval=2386698252 TSecr=549878280 [TCP segment of a reassembled PDU]
   16 9.626328   192.168.1.122   18.9.22.169     TCP        66 64123 → 80 [ACK] Seq=2208 Ack=7572 Win=65535 Len=0 TSval=549878373 TSecr=2386698252
   17 9.630191   18.9.22.169     192.168.1.122   TCP      1434 80 → 64123 [ACK] Seq=7572 Ack=2208 Win=8263 Len=1368 TSval=2386698253 TSecr=549878280 [TCP segment of a reassembled PDU]
   18 9.630692   18.9.22.169     192.168.1.122   TCP      1434 80 → 64123 [ACK] Seq=8940 Ack=2208 Win=8263 Len=1368 TSval=2386698253 TSecr=549878280 [TCP segment of a reassembled PDU]
   19 9.630718   192.168.1.122   18.9.22.169     TCP        66 64123 → 80 [ACK] Seq=2208 Ack=10308 Win=65493 Len=0 TSval=549878377 TSecr=2386698253
   20 9.638763   18.9.22.169     192.168.1.122   TCP      1434 80 → 64123 [ACK] Seq=10308 Ack=2208 Win=8263 Len=1368 TSval=2386698253 TSecr=549878280 [TCP segment of a reassembled PDU]
   21 9.639184   18.9.22.169     192.168.1.122   TCP      1434 80 → 64123 [ACK] Seq=11676 Ack=2208 Win=8263 Len=1368 TSval=2386698254 TSecr=549878280 [TCP segment of a reassembled PDU]
   22 9.639213   192.168.1.122   18.9.22.169     TCP        66 64123 → 80 [ACK] Seq=2208 Ack=13044 Win=65493 Len=0 TSval=549878385 TSecr=2386698253
   23 9.645350   18.9.22.169     192.168.1.122   TCP      1434 80 → 64123 [ACK] Seq=13044 Ack=2208 Win=8263 Len=1368 TSval=2386698254 TSecr=549878280 [TCP segment of a reassembled PDU]
   24 9.646260   18.9.22.169     192.168.1.122   TCP      1434 80 → 64123 [ACK] Seq=14412 Ack=2208 Win=8263 Len=1368 TSval=2386698254 TSecr=549878280 [TCP segment of a reassembled PDU]
   25 9.646290   192.168.1.122   18.9.22.169     TCP        66 64123 → 80 [ACK] Seq=2208 Ack=15780 Win=65493 Len=0 TSval=549878391 TSecr=2386698254
   26 9.651374   18.9.22.169     192.168.1.122   TCP      1434 80 → 64123 [ACK] Seq=15780 Ack=2208 Win=8263 Len=1368 TSval=2386698254 TSecr=549878280 [TCP segment of a reassembled PDU]

▸ Frame 1: 802 bytes on wire (6416 bits), 802 bytes captured (6416 bits)
▸ Ethernet II, Src: Apple_ac:6c:26 (10:9a:dd:ac:6c:26), Dst: Cisco-Li_e3:e9:8d (00:16:b6:e3:e9:8d)
▸ Internet Protocol Version 4, Src: 192.168.1.122, Dst: 18.9.22.169
▸ Transmission Control Protocol, Src Port: 64123, Dst Port: 80, Seq: 1, Ack: 1, Len: 736
▸ Hypertext Transfer Protocol

0000  00 16 b6 e3 e9 8d 10 9a  dd ac 6c 26 08 00 45 00   ········ ··l&··E·
0010  03 14 6c db 40 00 40 06  e0 34 c0 a8 01 7a 12 09   ··l·@·@· ·4···z··
0020  16 a9 fa 7b 00 50 99 11  72 a2 a9 f7 e4 a0 80 18   ···{·P·· r·······
0030  ff ff bb 76 00 00 01 01  08 0a 20 c6 54 df 8e 3f   ···v···· ·· ·T··?
0040  ea ac 47 45 54 20 2f 69  6d 67 2f 4d 49 54 5f 6c   ··GET /i mg/MIT_l
0050  6f 67 6f 2e 67 69 66 20  48 54 54 50 2f 31 2e 31   ogo.gif  HTTP/1.1
0060  0d 0a 48 6f 73 74 3a 20  77 77 77 2e 6d 69 74 2e   ··Host:  www.mit.
0070  65 64 75 0d 0a 43 6f 6e  6e 65 63 74 69 6f 6e 3a   edu··Con nection:

example.pcap          Packets: 1398 · Displayed: 1398 (100.0%)          Profile: Default
```

Figure 3: Orignal Pcap file



```
   Sr.No        Date             Time  ...  Length           Method           link
0      1  2015-10-21  23:10:34.995270  ...     153      ROSCTR:[Job              ]
1      2  2015-10-21  23:10:35.006630  ...      60              102              →
2      3  2015-10-21  23:10:35.006630  ...      60             [TCP            Dup
3      4  2015-10-21  23:10:35.007135  ...     104  ROSCTR:[Ack_Data]  Function:[Read
4      5  2015-10-21  23:10:35.007135  ...     104             [TCP Retransmission]
5      6  2015-10-21  23:10:35.028775  ...      60            49156              →
6      7  2015-10-21  23:10:35.028775  ...      60             [TCP            Dup
7      8  2015-10-21  23:10:35.208683  ...      74         Standard          query
8      9  2015-10-21  23:10:35.208796  ...     102      Destination    unreachable
9     10  2015-10-21  23:10:35.635307  ...      73         Standard          query

[10 rows x 10 columns]
abelled dataframe
   Sr.No        Date             Time  ...           Method           link      state
0      1  2015-10-21  23:10:34.995270  ...      ROSCTR:[Job              ]     normal
1      2  2015-10-21  23:10:35.006630  ...              102              →  malicious
2      3  2015-10-21  23:10:35.006630  ...             [TCP            Dup  malicious
3      4  2015-10-21  23:10:35.007135  ...  ROSCTR:[Ack_Data]  Function:[Read     normal
4      5  2015-10-21  23:10:35.007135  ...             [TCP Retransmission]     normal
5      6  2015-10-21  23:10:35.028775  ...            49156              →  malicious
6      7  2015-10-21  23:10:35.028775  ...             [TCP            Dup  malicious
7      8  2015-10-21  23:10:35.208683  ...         Standard          query  malicious
8      9  2015-10-21  23:10:35.208796  ...      Destination    unreachable     normal
9     10  2015-10-21  23:10:35.635307  ...         Standard          query  malicious
```

Figure 4: Labelled data Frame

- Use the following command to generate text file from pcap file

```
tshark -r /PATH/Filename. Pcap -t ad > /PATH/ Filename.txt
```

- The final step is to label the text file and use it for machine learning.

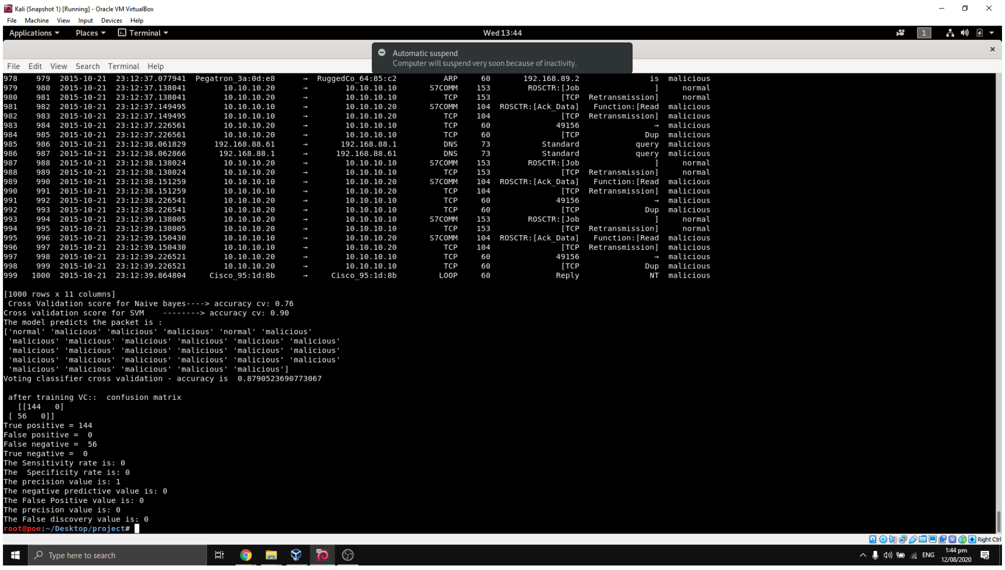
2.3 Libraries that have been imported for machine learning

1. from pandas import DataFrame [2]
2. import pandas as pd
3. import pandas
4. import numpy [3]
5. import numpy as np
   SKlearn [4]
6. from sklearn.externals import joblib
7. from sklearn.naive_bayes import MultinomialNB
8. from sklearn.feature_extraction.text import CountVectorizer
9. from sklearn.svm import LinearSVC
10. from sklearn.ensemble import VotingClassifier
11. from sklearn.linear_model import LogisticRegression
12. from optparse import OptionParser
13. from sklearn.cluster import KMeans
14. from sklearn import metrics
15. from sklearn.model_selection import train_test_split
16. from sklearn.model_selection import *
17. from sklearn import model_selection
18. from sklearn.linear_model import LogisticRegression
19. from sklearn import EnsembleVoteClassifier
20. import matplotlib [5]
21. import matplotlib.pyplot as plt
22. import matplotlib.cm as cm

3 Execution steps

1. python file.py

   Execute the file using `python -w file.py` to convert the pcap text file to a dataframe.



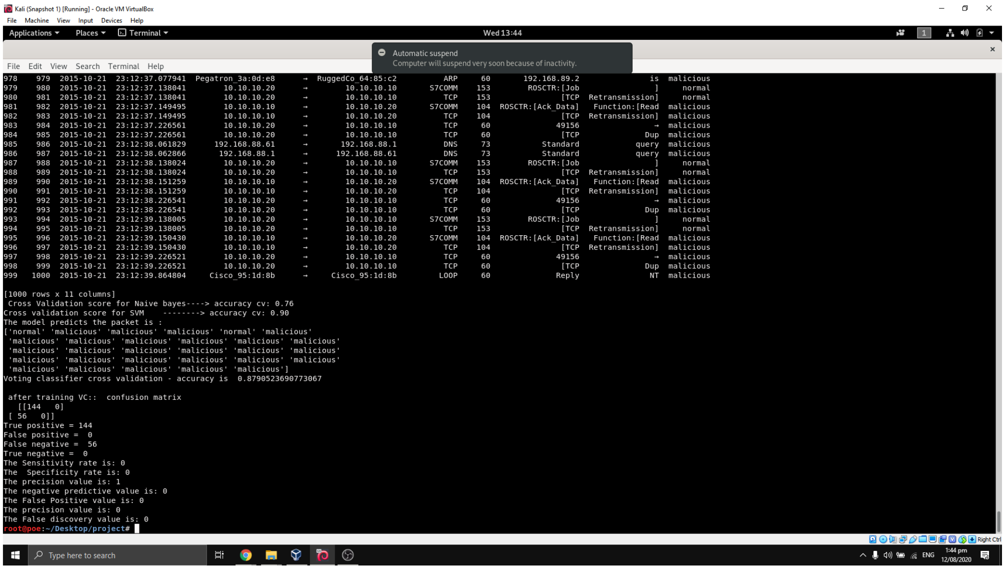
```
978     979  2015-10-21  23:12:37.077941  Pegatron_3a:0d:e8    →  RuggedCo_64:85:c2    ARP     60       192.168.89.2               is  malicious
979     980  2015-10-21  23:12:37.138041        10.10.10.20    →        10.10.10.10 S7COMM    153        ROSCTR:[Job                 ]     normal
980     981  2015-10-21  23:12:37.138041        10.10.10.20    →        10.10.10.10    TCP    153               [TCP  Retransmission]     normal
981     982  2015-10-21  23:12:37.149495        10.10.10.10    →        10.10.10.20 S7COMM    104  ROSCTR:[Ack_Data]   Function:[Read  malicious
982     983  2015-10-21  23:12:37.149495        10.10.10.10    →        10.10.10.20    TCP    104               [TCP  Retransmission]  malicious
983     984  2015-10-21  23:12:37.226561        10.10.10.20    →        10.10.10.10    TCP     60              49156                →  malicious
984     985  2015-10-21  23:12:37.226561        10.10.10.20    →        10.10.10.10    TCP     60               [TCP              Dup  malicious
985     986  2015-10-21  23:12:38.061829       192.168.88.61   →       192.168.88.1    DNS     73           Standard            query  malicious
986     987  2015-10-21  23:12:38.062866       192.168.88.1    →      192.168.88.61    DNS     73           Standard            query  malicious
987     988  2015-10-21  23:12:38.138024        10.10.10.20    →        10.10.10.10 S7COMM    153        ROSCTR:[Job                 ]     normal
988     989  2015-10-21  23:12:38.138024        10.10.10.20    →        10.10.10.10    TCP    153               [TCP  Retransmission]     normal
989     990  2015-10-21  23:12:38.151259        10.10.10.10    →        10.10.10.20 S7COMM    104  ROSCTR:[Ack_Data]   Function:[Read  malicious
990     991  2015-10-21  23:12:38.151259        10.10.10.10    →        10.10.10.20    TCP    104               [TCP  Retransmission]  malicious
991     992  2015-10-21  23:12:38.226541        10.10.10.20    →        10.10.10.10    TCP     60              49156                →  malicious
992     993  2015-10-21  23:12:38.226541        10.10.10.20    →        10.10.10.10    TCP     60               [TCP              Dup  malicious
993     994  2015-10-21  23:12:39.138005        10.10.10.20    →        10.10.10.10 S7COMM    153        ROSCTR:[Job                 ]     normal
994     995  2015-10-21  23:12:39.138005        10.10.10.20    →        10.10.10.10    TCP    153               [TCP  Retransmission]     normal
995     996  2015-10-21  23:12:39.150430        10.10.10.10    →        10.10.10.20 S7COMM    104  ROSCTR:[Ack_Data]   Function:[Read  malicious
996     997  2015-10-21  23:12:39.150430        10.10.10.10    →        10.10.10.20    TCP    104               [TCP  Retransmission]  malicious
997     998  2015-10-21  23:12:39.226521        10.10.10.20    →        10.10.10.10    TCP     60              49156                →  malicious
998     999  2015-10-21  23:12:39.226521        10.10.10.20    →        10.10.10.10    TCP     60               [TCP              Dup  malicious
999    1000  2015-10-21  23:12:39.864804     Cisco_95:1d:8b    →     Cisco_95:1d:8b   LOOP     60              Reply               NT  malicious

[1000 rows x 11 columns]
 Cross Validation score for Naive bayes----> accuracy cv: 0.76
Cross validation score for SVM    --------> accuracy cv: 0.90
The model predicts the packet is :
['normal' 'malicious' 'malicious' 'malicious' 'normal' 'malicious'
 'malicious' 'malicious' 'malicious' 'malicious' 'malicious' 'malicious'
 'malicious' 'malicious' 'malicious' 'malicious' 'malicious' 'malicious'
 'malicious' 'malicious' 'malicious' 'malicious' 'malicious' 'malicious'
 'malicious' 'malicious' 'malicious' 'malicious' 'malicious']
Voting classifier cross validation - accuracy is  0.8790523690773067

 after training VC::  confusion matrix
   [[144   0]
 [ 56   0]]
True positive = 144
False positive =  0
False negative =  56
True negative =  0
The Sensitivity rate is: 0
The  Specificity rate is: 0
The precision value is: 1
The negative predictive value is: 0
The False Positive value is: 0
The precision value is: 0
The False discovery value is: 0
root@poe:~/Desktop/project# 
```

Kali (Snapshot 1) [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places Terminal
Wed 13:44
Automatic suspend
Computer will suspend very soon because of inactivity.
File Edit View Search Terminal Help
Type here to search

Figure 5: Python Code for machine learning models

2. python check.py

   Execute the file using `python -w check.py` to convert the pcap text file to a dataframe.

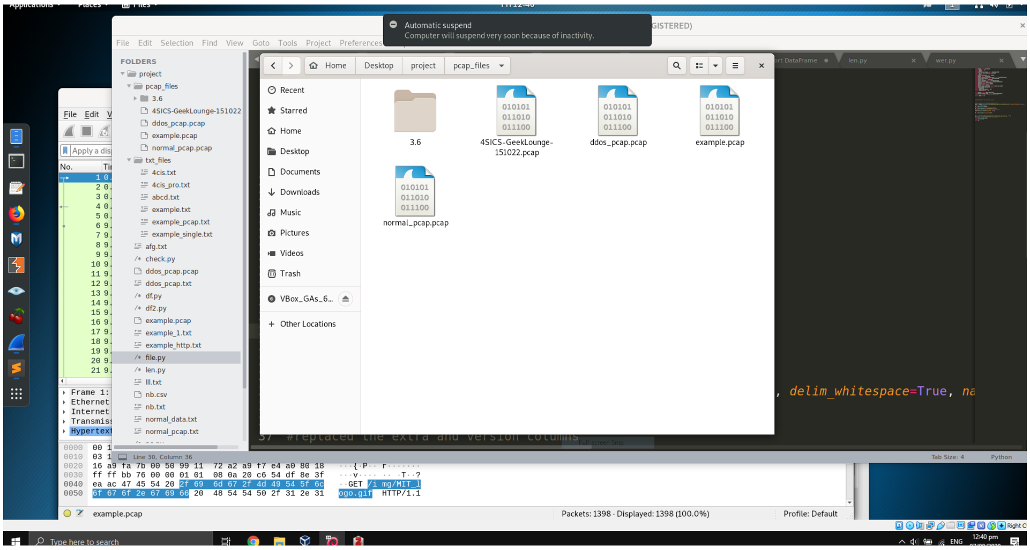


```
978     979  2015-10-21  23:12:37.077941  Pegatron_3a:0d:e8    →  RuggedCo_64:85:c2    ARP     60       192.168.89.2               is  malicious
979     980  2015-10-21  23:12:37.138041        10.10.10.20    →        10.10.10.10 S7COMM    153        ROSCTR:[Job                 ]     normal
980     981  2015-10-21  23:12:37.138041        10.10.10.20    →        10.10.10.10    TCP    153               [TCP  Retransmission]     normal
981     982  2015-10-21  23:12:37.149495        10.10.10.10    →        10.10.10.20 S7COMM    104  ROSCTR:[Ack_Data]   Function:[Read  malicious
982     983  2015-10-21  23:12:37.149495        10.10.10.10    →        10.10.10.20    TCP    104               [TCP  Retransmission]  malicious
983     984  2015-10-21  23:12:37.226561        10.10.10.20    →        10.10.10.10    TCP     60              49156                →  malicious
984     985  2015-10-21  23:12:37.226561        10.10.10.20    →        10.10.10.10    TCP     60               [TCP              Dup  malicious
985     986  2015-10-21  23:12:38.061829       192.168.88.61   →       192.168.88.1    DNS     73           Standard            query  malicious
986     987  2015-10-21  23:12:38.062866       192.168.88.1    →      192.168.88.61    DNS     73           Standard            query  malicious
987     988  2015-10-21  23:12:38.138024        10.10.10.20    →        10.10.10.10 S7COMM    153        ROSCTR:[Job                 ]     normal
988     989  2015-10-21  23:12:38.138024        10.10.10.20    →        10.10.10.10    TCP    153               [TCP  Retransmission]     normal
989     990  2015-10-21  23:12:38.151259        10.10.10.10    →        10.10.10.20 S7COMM    104  ROSCTR:[Ack_Data]   Function:[Read  malicious
990     991  2015-10-21  23:12:38.151259        10.10.10.10    →        10.10.10.20    TCP    104               [TCP  Retransmission]  malicious
991     992  2015-10-21  23:12:38.226541        10.10.10.20    →        10.10.10.10    TCP     60              49156                →  malicious
992     993  2015-10-21  23:12:38.226541        10.10.10.20    →        10.10.10.10    TCP     60               [TCP              Dup  malicious
993     994  2015-10-21  23:12:39.138005        10.10.10.20    →        10.10.10.10 S7COMM    153        ROSCTR:[Job                 ]     normal
994     995  2015-10-21  23:12:39.138005        10.10.10.20    →        10.10.10.10    TCP    153               [TCP  Retransmission]     normal
995     996  2015-10-21  23:12:39.150430        10.10.10.10    →        10.10.10.20 S7COMM    104  ROSCTR:[Ack_Data]   Function:[Read  malicious
996     997  2015-10-21  23:12:39.150430        10.10.10.10    →        10.10.10.20    TCP    104               [TCP  Retransmission]  malicious
997     998  2015-10-21  23:12:39.226521        10.10.10.20    →        10.10.10.10    TCP     60              49156                →  malicious
998     999  2015-10-21  23:12:39.226521        10.10.10.20    →        10.10.10.10    TCP     60               [TCP              Dup  malicious
999    1000  2015-10-21  23:12:39.864804     Cisco_95:1d:8b    →     Cisco_95:1d:8b   LOOP     60              Reply               NT  malicious

[1000 rows x 11 columns]
 Cross Validation score for Naive bayes----> accuracy cv: 0.76
Cross validation score for SVM    --------> accuracy cv: 0.90
The model predicts the packet is :
['normal' 'malicious' 'malicious' 'malicious' 'normal' 'malicious'
 'malicious' 'malicious' 'malicious' 'malicious' 'malicious' 'malicious'
 'malicious' 'malicious' 'malicious' 'malicious' 'malicious' 'malicious'
 'malicious' 'malicious' 'malicious' 'malicious' 'malicious' 'malicious'
 'malicious' 'malicious' 'malicious' 'malicious' 'malicious']
Voting classifier cross validation - accuracy is  0.8790523690773067

 after training VC::  confusion matrix
   [[144   0]
 [ 56   0]]
True positive = 144
False positive =  0
False negative =  56
True negative =  0
The Sensitivity rate is: 0
The  Specificity rate is: 0
The precision value is: 1
The negative predictive value is: 0
The False Positive value is: 0
The precision value is: 0
The False discovery value is: 0
root@poe:~/Desktop/project# 
```

Kali (Snapshot 1) [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places Terminal
Wed 13:44
Automatic suspend
Computer will suspend very soon because of inactivity.
File Edit View Search Terminal Help
Type here to search

Figure 6: Python Code for machine learning models

3. file structure for pcap files



Automatic suspend
Computer will suspend very soon because of inactivity.

```
, delim_whitespace=True, na
57  #replaced the extra and version columns
```

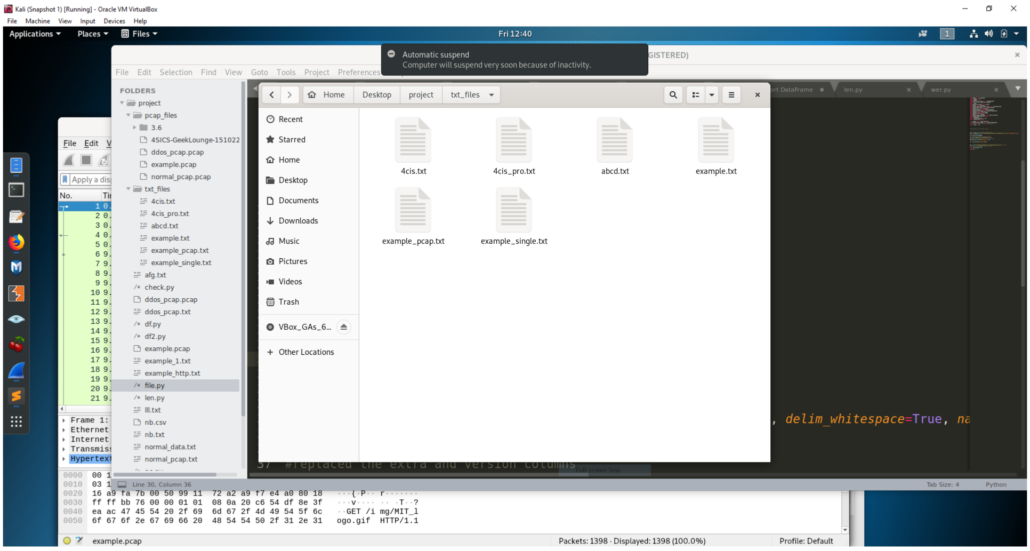
```
0000  00 1
0010  03 1
0020  16 a9 fa 7b 00 50 99 11  72 a2 a9 f7 e4 a0 80 18   ···{·P·· r·······
0030  ff ff bb 76 00 00 01 01  08 0a 20 c6 54 df 8e 3f   ···v···· ·· ·T··?
0040  ea ac 47 45 54 20 2f 69  6d 67 2f 4d 49 54 5f 6c   ··GET /i mg/MIT_l
0050  6f 67 6f 2e 67 69 66 20  48 54 54 50 2f 31 2e 31   ogo.gif  HTTP/1.1
```

Line 30, Column 36

example.pcap   Packets: 1398 · Displayed: 1398 (100.0%)   Profile: Default

Figure 7: Pcap file Repository

4. file structure for text files



Automatic suspend
Computer will suspend very soon because of inactivity.

```
, delim_whitespace=True, na
57  #replaced the extra and version columns
```

```
0000  00 1
0010  03 1
0020  16 a9 fa 7b 00 50 99 11  72 a2 a9 f7 e4 a0 80 18   ···{·P·· r·······
0030  ff ff bb 76 00 00 01 01  08 0a 20 c6 54 df 8e 3f   ···v···· ·· ·T··?
0040  ea ac 47 45 54 20 2f 69  6d 67 2f 4d 49 54 5f 6c   ··GET /i mg/MIT_l
0050  6f 67 6f 2e 67 69 66 20  48 54 54 50 2f 31 2e 31   ogo.gif  HTTP/1.1
```

Line 30, Column 36

example.pcap   Packets: 1398 · Displayed: 1398 (100.0%)   Profile: Default

Figure 8: Text file Repository

References

[1] “Welcome to Python.org,” library Catalog: www.python.org. [Online]. Available: https://www.python.org/

[2] “pandas - Python Data Analysis Library.” [Online]. Available: https://pandas.pydata.org/

[3] “NumPy.” [Online]. Available: https://numpy.org/

[4] “scikit-learn: machine learning in Python — scikit-learn 0.23.2 documentation.” [Online]. Available: https://scikit-learn.org/stable/

[5] “Matplotlib: Python plotting — Matplotlib 3.3.0 documentation.” [Online]. Available: https://matplotlib.org/